PS1

12265092

19/01/2022

```
rm(list=ls())

library(haven)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

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

```
data <- rep(c(1,0),times = c(600,400))

#data

data <- as.data.frame(data)
colnames(data) <- c("treatment_status")
set.seed(3132022)

##Create a dataset containing 1000 observations.
##Set a seed of 3132022 and construct a variable treated that takes a value of 1 for the  first  600  o
##Simulate  two  random variables X1 and X2, where X1 and X2 are each uniformly distributed on [0,1]. L

X1 <- runif(1000,min = 0 , max = 1)
X2 <- runif(1000,min = 0, max = 1)

ui <- X1
Ti <- (-2*X2) + 1

Yic <- 5 + ui
data$control <- Yic

Yit <- 5 + Ti + ui
data$treatment <- Yit
```

```r
## a) Define a new variable Yobs equal to the length of hospitalization that would be observed for each

Yobs = ifelse(data == 1, Yit, Yic)

#Yobs

#data <- data %>% mutate(Yobs = ifelse(treatment_status == 1, Yit , Yic))
```

```r
## b) Use  Yobs  to  compute  an  estimated  average  treatment  effect  on  length  of  hospitalizatio

b_1 <- data %>% group_by(treatment_status) %>% summarise(avg = mean(Yobs))

b_2 <- data %>% group_by(treatment_status) %>% summary(Yobs)

b_1
```

```
## # A tibble: 2 x 2
##   treatment_status   avg
##              <dbl> <dbl>
## 1                0  5.49
## 2                1  5.49
```

```r
b_2
```

```
##  treatment_status    control         treatment
##  Min.   :0.0      Min.   :5.002   Min.   :4.047
##  1st Qu.:0.0      1st Qu.:5.249   1st Qu.:4.992
##  Median :1.0      Median :5.480   Median :5.484
##  Mean   :0.6      Mean   :5.492   Mean   :5.480
##  3rd Qu.:1.0      3rd Qu.:5.750   3rd Qu.:5.993
##  Max.   :1.0      Max.   :5.999   Max.   :6.978
```

```r
#calculate from treated group and control group
Ti_estimate  <- mean(Yobs[1:600]) - mean(Yobs[601:1000])

Ti_estimate
```

```
## [1] -0.02165013
```

```r
print(paste("Eestimated  average  treatment  effect  = " , Ti_estimate))
```

```
## [1] "Eestimated  average  treatment  effect  =  -0.0216501270141451"
```

```r
## c) What is the population mean of Ti? Is your answer to Part (b) exactly equal to this population me

data$Ti  <- Ti

c <- mean(data$Ti)

c
```

```
## [1] -0.01267211
```

```r
print(paste("Population mean of Ti = " , c))
```

```
## [1] "Population mean of Ti =  -0.0126721069491468"
```

Response for c) The estimated treatment effect calculated in part b is of value -0.0216501270141451 where as the population mean of treatment calculated with the sample i.e average treatment effect calculated with the samples is -0.0126721069491468. Both are not equal.This is due to the reason that we CANNOT observe what happens to a person in both treatment and control states. This is a fundamental problem that we face while analysing the data. If only we could observe the outcome in both states, we could simply compare the two observed outcomes and determine the effect of treatment on a person.

##Using the notations in Rubins model. Suppose, we have a person (i) and intervention (t) whose effects we want to estimate as compared to a control (c) which is a lack of intervention. The person i can be in either groups i.e treated group denoted by state of world S = (t) or control group denoted by state of world S = (c). The outcomes before the treatment is assigned , we imagine measure of interest Yit (associated with State t) and measure of interest Yic (associated with state c). The effect of treatment (or) the causal effect (or) difference in potential outcomes can be determined using Yit - Yic i.e difference of the outcomes in the two states. But however, once the state of world S for a person is assigned, we cannot observe both the outcomes. Thus without the counterfactual, we cannot determine the Average treatement effect Yt - Yc.

```r
## d) Calculate the mean of Ti in the sample (using all 1000 observations). Is the mean of Ti exactly th

Ti_sample <- mean(Yit - Yic)

print(paste( "Mean of Ti in the sample = " , Ti_sample))
```

```
## [1] "Mean of Ti in the sample =  -0.0126721069491468"
```

Response for d) The estimated treatment effect calculated in part b is of value -0.0216501270141451 where as the sample mean of treatment calculated with the sample is -0.0126721069491468. Both are not equal.This is due to the reason that we CANNOT observe what happens to a person in both treatment and control states. This is a fundamental problem that we face while analysing the data. If only we could observe the outcome in both states, we could simply compare the two observed outcomes and determine the effect of treatment on a person.

##Using the notations in Rubins model. Suppose, we have a person (i) and intervention (t) whose effects we want to estimate as compared to a control (c) which is a lack of intervention. The person i can be in either groups i.e treated group denoted by state of world S = (t) or control group denoted by state of world S = (c). The outcomes before the treatment is assigned , we imagine measure of interest Yit (associated with State t) and measure of interest Yic (associated with state c). The effect of treatment (or) the causal effect (or) difference in potential outcomes can be determined using Yit - Yic i.e difference of the outcomes in the two states. But however, once the state of world S for a person is assigned, we cannot observe both the outcomes. Thus without the counterfactual, we cannot determine the Average treatement effect Yt - Yc.

```r
## e) Suppose you are a researcher that is reviewing the outcomes of this trial. What could you conclude

Ti_estimate_2 = mean(Yobs[1:600]) - mean(Yobs[601:1000])
```

```r
Ti_estimate_2
```

```
## [1] -0.02165013
```

```r
print(paste("Estimated Average treatment effect Yit - Yic = ", Ti_estimate_2))
```

```
## [1] "Estimated Average treatment effect Yit - Yic =  -0.0216501270141451"
```

```r
###The estimated average treatment effect is -0.0216501270141451
```

Response for e) The estimated average treatment effect is -0.0216501270141451. The estimate firs of all cannot confirm if a particular unit would gain or lose from the treatment. But this estimate can tell what happened in the sample. As the value is negative, it can be said that the effect of treatment reduced the lenght of hospitalization. It might vary in groups among the sample where some of the population might act against the mean effect.

```r
### f) In this simulation, we know the true distribution of Ti. (Note that this is seldom if ever possi

##Range

range <- max(Ti) - min(Ti)

print(paste("Range of treatment effects in population is = ", range ))
```

```
## [1] "Range of treatment effects in population is =  1.9957196386531"
```

```r
##Share beneffitted

beneffitted <- ifelse(Yit < Yic, 1, 0)

total_population <- length(beneffitted)

share_beneffitted <- (sum(beneffitted)/total_population)

percentage_benefitted <- share_beneffitted * 100

print(paste("Share of population benefitted from treatment = ", share_beneffitted))
```

```
## [1] "Share of population benefitted from treatment =  0.519"
```

```r
print(paste("Percentage of population benefitted from treatment = ", percentage_benefitted))
```

```
## [1] "Percentage of population benefitted from treatment =  51.9"
```

```r
##Harmed population

## where Yit > Yic, similar to above

harmed_population <- ifelse(Yit >Yic, 1, 0)


share_harmed <- sum(harmed_population)/total_population

percentage_harmed <- share_harmed * 100

print(paste("Share of population harmed from treatment = ", share_harmed))
```

```
## [1] "Share of population harmed from treatment =  0.481"
```

```r
print(paste("Percentage of population harmed from treatment = ", percentage_harmed))
```

```
## [1] "Percentage of population harmed from treatment =  48.1"
```