18-4-18-5

Nikhil Gopal

3/6/2022

```
rm(list = ls())
library(tidyverse)
```

18.4

Average treatment effects: The table below describes a hypothetical experiment on 8 people. Each row of the table gives a participant and her pre-treatment predictor x, treatment indicator z, and potential outcomes y0 and y1.

- a) Give the average treatment effect in the population, the average treatment effect among the treated, and the estimated treatment effect based on a simple comparison of treatment and control.

```
df <- data.frame(
  patient <- c("Anna", "Beth", "Cari", "Dora", "Edna", "Fala", "Geri", "Hana"),
  x <- c(3,5,2,8,5,10,2,11),
  z <- c(0,0,1,0,0,1,1,1),
  y1 <- c(5,8,5,12,4,8,4,9),
  y2 <- c(5,10,3,13,2,9,1,13)
  )

colnames(df) <- c("patient", "pre_treatment", "treatment", "outcome0", "outcome1")
df$treatment <- ifelse(df$treatment == 0, "Control", "Treatment")
```

Average Treatment effect in the population:

```
df$treat_effect <- df$outcome1 - df$outcome0

mean(df$treat_effect)
```

```
## [1] 0.125
```

In this case, y0 represents the pre-treatment value, and y1 represents the value observed post treatment. Thus, the average treatment effect is the average of y1 - y0, which is 0.125.

Average treatment effect among treated:

```
df %>% group_by(treatment) %>%
  summarise(mean_treat_effect = mean(treat_effect))
```

```
## # A tibble: 2 x 2
##   treatment mean_treat_effect
##   <chr>                 <dbl>
## 1 Control                0.25
## 2 Treatment              0
```

Estimated treatment effect based on a simple comparison of treatment and control:

```
mean(df$treat_effect[df$treatment=="Treatment"]) - mean(df$treat_effect[df$treatment=="Control"])
```

```
## [1] -0.25
```

The estimated treatment effect based on a comparison of treatment and control groups is -0.25.

- b) Simulate a new completely randomized experiment on these 8 people; that is, resample z at random with the constraint that equal numbers get the treatment and the control. Report your new randomization and give the corresponding answers for (a).

```
set.seed(1234)
random_df <- data.frame(
  patient <- c("Anna", "Beth", "Cari", "Dora", "Edna", "Fala", "Geri", "Hana"),
  x <- c(3,5,2,8,5,10,2,11),
  z <- sample(c(0,0,0,0,1,1,1,1), 8, replace = FALSE),
  y1 <- c(5,8,5,12,4,8,4,9),
  y2 <- c(5,10,3,13,2,9,1,13)
  )

colnames(random_df) <- c("patient", "pre_treatment", "treatment", "outcome0", "outcome1")
random_df$treatment <- ifelse(random_df$treatment == 0, "Control", "Treatment")
```

Average Treatment effect in the population:

```
random_df$treat_effect <- random_df$outcome1 - random_df$outcome0

mean(random_df$treat_effect)
```

```
## [1] 0.125
```

The average treatment effect in the population remains the same, 0.125.

Average treatment effect among treated:

```
random_df %>% group_by(treatment) %>%
  summarise(mean_treat_effect = mean(treat_effect))
```

```
## # A tibble: 2 x 2
##   treatment mean_treat_effect
##   <chr>                 <dbl>
## 1 Control                0
## 2 Treatment              0.25
```

Estimated treatment effect based on a simple comparison of treatment and control:

```
mean(random_df$treat_effect[random_df$treatment=="Treatment"]) - mean(random_df$treat_effect[random_df$
```

```
## [1] 0.25
```

This time, the treatment effect size was 0.25.

18.5

Potential outcomes: The table below describes a hypothetical experiment on 2400 people. Each row of the table specifies a category of person, as defined by his or her pre-treatment predictor x, treatment indicator z, and potential outcomes y0, y1. For simplicity, we assume unrealistically that all the people in this experiment fit into these eight categories.

```
df2 <- data.frame(
  category <- c(1,2,3,4,5,6,7,8),
  number_in_cat <- c(300,300,500,500,200,200,200,200),
  x <- c(0,1,0,1,0,1,0,1),
  z <- c(0,0,1,1,0,0,1,1),
  y0 <- c(4,4,4,4,10,10,10,10),
  y1 <- c(6,6,6,6,12,12,12,12)
)
```

```
colnames(df2) <- c("category", "number_in_cat", "pre_treat_predictor", "treatment", "outcome0", "outcome
```

In making the table we are assuming omniscience, so that we know both y0 and y1 for all observations. But the (non-omniscient) investigator would only observe x, z, and yz for each unit. For example, a person in category 1 would have x =0, z=0, y =4, and a person in category 3 would have x=0, z=1, y=6.

- a) What is the average treatment effect in this population of 2400 people?

```
df2$treat_effect <- df2$outcome1 - df2$outcome0
```

```
mean(df2$treat_effect)
```

```
## [1] 2
```

The average treatment effect is 2.

- b) Another summary is the mean of y for those who received the treatment minus the mean of y for those who did not. What is the relation between this summary and the average treatment effect (ATE)?

```
mean(df2$treat_effect[df2$treatment == 1]) - mean(df2$treat_effect[df2$treatment == 0])
```

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

The mean difference in treatment in effect between the groups was 0. For both groups, the difference was 2 in all cases, meaning the mean difference would be 0. This explains why the ATE observed above was 2.

- c) Is it plausible to believe that these data came from a completely randomized experiment? Defend your answer.

It is possible that this data came from a randomized experiement, but extremely unlikely. It could be that the treatment variable did not have a causal relationship on the outcome, and as we noticed, there was an even distribution of people in the $z = 0$ and $z = 1$ for $y = 4/6$ and $10/12$, meaning there were no major flaws in survey design.

However, it is extremely suspicious that the mean difference between treatment and control groups was exactly 0. This likely suggests doctored data.

- d) For these data, is it plausible to believe that treatments were assigned using randomized blocks conditional on given x? Defend your answer.

Again, while it is plausible it is extremely unlikely. It is extremely unlikely that the first four values of y0 and y1 would be 4/6 and then the next four would also be 10/12. This likely suggests that a human assigned people to groups.