Assignment 9b

Satvik Saha

2024-11-11

Answer 1

```r
df <- data.frame(
    i = c("Audrey", "Anna", "Bob", "Bill", "Caitlin", "Cara", "Dave", "Doug"),
    x = c(40, 45, 50, 55, 60, 65, 70, 75),
    z = c(0, 1, 1, 0, 0, 0, 1, 1),
    y0 = c(140, 140, 150, 150, 160, 160, 170, 170),
    y1 = c(135, 135, 140, 140, 155, 155, 160, 160)
)
```

(a)
```r
audrey <- df$i == "Audrey"
df$y1[audrey] - df$y0[audrey]
```

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

Thus, the treatment effect for Audrey is -5.

(b)
```r
df$tau <- df$y1 - df$y0
mean(df$tau)
```

```
## [1] -7.5
```

Thus, the sample average treatment effect is -7.5.

(c)
```r
df$y <- ifelse(df$z == 0, df$y0, df$y1)
lm(y ~ z, data = df)
```

```
##
## Call:
## lm(formula = y ~ z, data = df)
##
## Coefficients:
## (Intercept)            z
##      152.50        -3.75
```

Thus, the estimated treatement effect is -3.75.

For the sake of curiosity, consider

```r
lm(y ~ z + x, data = df)
```

```
##
## Call:
## lm(formula = y ~ z + x, data = df)
##
```

```
## Coefficients:
## (Intercept)           z           x
##     104.375      -8.125       0.875
```

Answer 2

It would seem that (b) is true; the prediction $X\beta$ has a smoothing effect leading to the absence of more extreme values for income. In particular, the imputed data will be 'missing' extreme values produced by the error terms.