20-3-11

Nikhil Gopal

4/1/2022

20.3

Understanding the difference between average treatment effect in the treated and control groups: Create a hypothetical dataset in which the average treatment effect on the treated and controls (ATT and ATC) are clearly different. What are the distinguishing characteristics of this dataset?

```
df <- data.frame(
  pre_treatment_value <- rep(50, 500),
  is_treatment_group <- c(rep(0,250), rep(1,250)),
  post_treatment <- c(rep(50, 250), rep(100, 250)),
  difference <- post_treatment - pre_treatment_value
  )

colnames(df) <- c("pre_treatment_value", "is_treatment_group", "post_treatment", "difference")

df %>% group_by(is_treatment_group) %>%
  summarise(mean_difference = mean(difference))
```

```
## # A tibble: 2 x 2
##   is_treatment_group mean_difference
##                <dbl>           <dbl>
## 1                  0               0
## 2                  1              50
```

Above is a sample dataset I created with a mean difference of 0 in the control group, and a mean difference of 50 in the treatment group.I purposefully made the treatment effect size bigger to demonstrate a difference between the effect size in treatment/control groups.

20.11

Observational studies and hypothetical experiments: Consider an applied problem of interest to you with a causal effect that has been estimated using an observational study. Think about possible hypothetical experiments that could be performed to estimate different aspects of this causal effect. Consider how the effect might vary across the population and across different implementations of the treatment.

I would like to investigate if religion increases the odds of someone voting for a Republican candidate in presidential elections. The variable of interest would be if someone self identifies as religious and the outcome variable would be a binary indicator of weather they voted Republican.

Special attention should be paid to the sampling, as people should not just be sampled from a specific location. People from different income levels, locations and education levels should all be included in the sample. Additionally, I would like to collect data on the frequency that one attends religious services, and include this in the model if possible. Some people might self identify as religious but not actually practice.