

Does Social Pressure Affect Turnout?

Part II: Estimating an Average Causal Effect

Let's continue working with the data from the randomized experiment conducted in Michigan and estimate the average causal effect of receiving the social pressure message on the probability of voting.

The dataset we will use is in a file called "voting.csv". Table 1 shows the names and descriptions of the variables in this dataset, where the unit of observation is registered voters.

variable	description
<i>birth</i>	year of birth of registered voter
<i>message</i>	whether registered voter was assigned to receive the social pressure message: "yes", "no"
<i>voted</i>	whether registered voter voted in the August 2006 election: 1=voted, 0=didn't vote

Table 1: Variables in "voting.csv"

In this problem set, we practice how to create new variables, how to estimate an average treatment effect using data from a randomized experiment, and how to write a conclusion statement.

As always, we start by loading and looking at the data:

```
## load and look at the data
voting <- read.csv("voting.csv") # reads and stores data
head(voting) # shows first observations
##   birth message voted
## 1 1981      no     0
## 2 1959      no     1
## 3 1956      no     1
## 4 1939     yes     1
## 5 1968      no     0
## 6 1967      no     0
```

1. To estimate the average causal effect of receiving the social pressure message on the probability of voting:
 - a. What should be our outcome variable (Y) and what type of variable is it? Please just provide the name of the variable and indicate whether the variable is character, numeric binary, or numeric non-binary. (2.5 points)
 - b. Do we already have in the dataframe our treatment variable (X), that is, a numeric binary variable that identifies the observations that received the treatment? A yes/no answer will suffice. (2.5 points)
 - c. Let's create our treatment variable by using the function `ifelse()` and the existing variable `message`. Name this new variable `pressure` and make sure to store it in the dataframe `voting`. (R code only.) (2.5 points)

- d. Look at the first few observations of *voting* to confirm that the new variable was created correctly. (R code only.) (2.5 points)
 - e. Considering that the dataset comes from a randomized experiment, what can we compute to estimate the average causal effect of receiving the social pressure message on the probability of voting? Please provide the name of the estimator and its definition. (2.5 points)
 - f. What are the individuals or entities in the treatment group and what are the individual or entities in the control group? Please be precise. (2.5 points)
2. Let's start by computing the average outcome for the treatment group and the average outcome for the control group separately. Please interpret the output of both with a full sentence and make sure to include the unit of measurement. (10 points)
 3. Now, let's answer the question: What is the estimated average causal effect of receiving the social pressure message on the probability of voting? Please provide a full substantive answer (make sure to include the assumption, why the assumption is reasonable, the treatment, the outcome, as well as the direction, size, and unit of measurement of the average treatment effect) (25 points)