This material was produced for instructors using Llaudet, Elena and Kosuke Imai.

Data Analysis for Social Science: A Friendly and Practical Introduction. (Princeton University Press) and should not be shared beyond those who are enrolled in this class.

Does Social Pressure Affect Turnout? Part III: Estimate an Average Causal Effect, Determine Statistical Significance and Discuss Internal and External Validity

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
birth	year of birth of registered voter
message	whether registered voter was assigned to receive the social pressure message: "yes", "no"
voted	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 (1) estimate an average treatment effect using data from a randomized experiment, (2) determine whether the estimated average treatment effect is statistically significant at the 5% level, and (3) determine the internal and external validity of the study.

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</pre>
head(voting) # shows first six observations
##
      birth message voted
## 1 1981
                no
                       \cap
## 2 1959
                       1
                no
## 3 1956
                       1
                no
## 4 1939
                       1
               yes
## 5 1968
                       0
                no
                       0
## 6 1967
```

Then, we run the following code to create our treatment variable and confirm that it is created correctly:

```
## create variable pressure inside dataframe voting

voting$pressure <- # stores return values in new variable

ifelse (voting$message=="yes", # logical test

1, # return value if logical test is true

0) # return value if logical test is false
```

look at the first observations again to ensure pressure was created correctly head(voting, n=4) # shows first four observations

This material was produced for instructors using Llaudet, Elena and Kosuke Imai.

Data Analysis for Social Science: A Friendly and Practical Introduction. (Princeton University Press) and should not be shared beyond those who are enrolled in this class.

##	birth	message vo	oted	pressure
## 1	1981	no	0	0
## 2	1959	no	1	0
## 3	1956	no	1	0
## 4	1939	yes	1	1

- 1. What is the estimated average causal effect of receiving the social pressure message on the probability of voting?
 - a. Fit a linear model to the data in such a way that the estimated slope coefficient is equivalent to the difference-in-means estimator you are interested in and store the fitted model in an object called *fit* (R code only). (5 points)
 - b. What is the estimated slope coefficient, $\widehat{\beta}$? (5 points)
 - c. Now, let's answer the question: What is the estimated average treatment effect? 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) (10 points)
- 2. Is the effect statistically significant at the 5% level?
 - a. Let's start by specifying the null and alternative hypotheses. Please provide both the mathematical notations and their meaning. (5 points)
 - b. What is the value of the observed test statistic, z^{obs} ? (Hint: the code summary()\$coeff might be helpful here.) (5 points)
 - c. What is the associated p-value? (5 points)
 - d. Now, let's answer the question: Is the effect statistically significant at the 5% level? Please provide your reasoning. (5 points)
- 3. Can we interpret the estimated effect as causal? In other words, how strong is the internal validity of this study? Have the researchers accurately measured the average causal effect on the sample of registered voters who were part of the study? Explain your reasoning. (5 points)
- 4. Can we generalize the results? In other words, how strong is the external validity of this study? Please explain your reasoning and be specific about what population you think the findings can or cannot be generalized to. (5 points)