

## Does Social Pressure Affect Turnout?

### Part I: Loading and Making Sense of Data, and Computing and Interpreting Means

(Based on Alan Gerber, Donald Green, and Christopher Larimer. 2008. "Social Pressure and Voter Turnout: Evidence from a Large-Scale Field Experiment." *American Political Science Review*, 102(1): 33-48.)

In part II, we will estimate the average causal effect of receiving a message designed to induce social pressure to vote on the probability of voting. For this purpose, we will analyze data from an experiment conducted in Michigan in 2006, where registered voters were randomly assigned to receive a postcard with the following message:

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED? ... We're sending this mailing to you and your neighbors to publicize who does and does not vote. The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not. DO YOUR CIVIC DUTY—VOTE!

MAPLE DRIVE	Aug 2004	Nov 2004	Aug 2006
9993 [YOU]	Didn't vote	Voted	???
9995 JOSEPH JAMES SMITH	Voted	Voted	???
9997 RICHARD B JACKSON	Didn't vote	Voted	???
9999 KATHY MARIE JACKSON	Didn't vote	Voted	???

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 (i) how to load and make sense of data and (ii) how to compute and interpret means.

1. Use the function `read.csv()` to read the CSV file "voting.csv" and use the assignment operator `<-` to store the data in an object called *voting*. (Do not forget to set the working directory first.) Provide the R code you used (without the output). (5 points)

2. Use the function `head()` to view the first few observations of the dataset. Provide the R code you used (without the output). (5 points)
3. What does each observation in this dataset represent? (5 points)
4. Please substantively interpret the first observation in the dataset. (5 points)
5. For each variable in the dataset, please identify the type of variable (character vs. numeric binary vs. numeric non-binary) (5 points)
6. How many observations are in the dataset? In other words, how many registered voters were part of this experiment? (Hint: the function `dim()` might be helpful here.) Provide the R code you used (without the output) and provide the substantive answer. (5 points)
7. Now, let's use the function `mean()` to calculate the average of the variable *birth*. Please provide a full substantive interpretation of what this average means. Make sure to provide the unit of measurement. (10 points)
8. Finally, use the function `mean()` to calculate the average of the variable *voted*. Please provide a full substantive interpretation of what this average means. Make sure to provide the unit of measurement. (10 points)