

## ECON 104

## Truth in Numbers

## Homework 1

For this homework we will see if we can recover the actual effect of a get out the vote effort on voter turnout. We will do this using selection on observable designs and contrast the estimates with an experimental benchmark. The data is from the following paper. Arce-neaux, Kevin, Alan Gerber, and Donald Green. "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment." *Political Analysis*, 2006, 14, 37-62. This study was intended to assess the effectiveness of get out the vote efforts. The authors attempted to deliver the following get out the vote message

*Hello, may I speak with (name of person) please? Hi. This is (caller's name) calling from Vote 2002, a nonpartisan effort working to encourage citizens to vote. We just wanted to remind you that elections are being held this Tuesday. The success of our democracy depends on whether we exercise our right to vote or not, so we hope you'll come out and vote this Tuesday. Can I count on you to vote next Tuesday?*

For this homework use the dataset on voting from the course web page that corresponds with the first letter of your last name. The dataset includes the following variables

- **vote02:** Which is the outcome of interest- vote in 2002
- **treat\_real:** Assignment to treatment group
- **contact:** Received phone call and responded to question "*Can I count on you to vote next Tuesday?*" regardless of answer
- **newreg** - Newly registered voter
- **busy** - Whether the phone line was busy when call was made
- **age** - Age of the individual in years
- **female** - Gender of the individual

- **vote00** - Voted in 2000 (prior to treatment)
- **state**: State of residence 1 for Iowa and 0 for Michigan
- **comp\_mi**: Value of 1 for - competitive district in Michigan (Michigan A in the paper)
- **comp\_ia**: Value of 1 for - competitive district in Iowa (Iowa A in the paper)
- **vote98**: Voted in 98 (prior to treatment)

The rest of the covariates are reasonably well labeled.

1. Provide evidence that the randomization worked by comparing the means of the sample characteristics in the treatment and control groups. Please create a clean table that includes columns with the means of each group, the difference between the two groups and the p-value of the difference. The table should be comprehensible on its own.
2. Is the table you produced in answer to question 1 consistent with the randomization being correctly implemented? Why or why not?
3. Estimate the difference in the voting rates for the treatment and control group. How big an effect did getting assigned to get a phone call have on the probability of voting. Is it statistically significant? Is it large in a practical sense?
4. Create a carefully labeled table where each column corresponds to a regression. The first column contains the parameters of a the regression  $vote02 = B_0 + B_1 treat\_real + u$ . In each following column you add one more covariate to the regression.
5. What effect does adding covariates have on your estimate of the treatment effect? What does this tell you about the relationship between the covariates and the outcome?
6. Will comparing the voting rate of the group that got assigned to get the call with the voting rate of the group that didn't get assigned to get a call get us an unbiased estimate of the causal effect of getting assigned to get encouraging call on the probability of voting? Why or why not?

7. Please attach all your code.