

Homework Assignment 2

DPI 610

Assignment Details

This homework assignment is due *March 5* before midnight. Everyone must complete their own code and assignment, though you may discuss the homework with classmates.

To access the code and data for the assignment, go to Canvas or the FASOnDemand Cluster. To turn in the assignment, upload two files to Canvas: (1) your code file (.Rmd), and (2) a PDF version (click Knit to create PDF).

Technical Note: For Questions 1 and 2, use the R object `f1` – this is the data from the experiment described in the prompt below. In Question 3, use the data from the CCES named `f1_cces` as well.

Setup

The following code loads the dataset and packages necessary to complete the problem set. Please make sure to run the following code before starting the problem set.

```
## load packages
library("car")
library("pROC")
#library("estimatr")

## load data
dat <- readRDS("fl_data.rds")
f1 <- dat$f1
f1_cces <- dat$f1_cces
```

Introduction

You are employed at a political consulting firm that will be working on several races in Florida in 2020. Your boss has assigned you to look back over data gathered from an experiment conducted by your firm preceding the 2014 general election that (for some strange reason) was never analyzed (the data set is named `f1` and pre-loaded).

In the experiment, the firm mailed postcards to a randomly selected set of registered voters, showing the recipient their past turnout history and comparing it to the level of participation of the typical person in their state.¹ Your task is to determine what lessons, if any, can be drawn from this experiment and applied to future efforts at boosting turnout among target voters.

¹The data comes from the replication archive of the paper “The Generalizability of Social Pressure Effects on Turnout Across High-Salience Electoral Contexts: Field Experimental Evidence From 1.96 Million Citizens in 17 States” by Alan Gerber, Greg Huber, Albert Fang, and Andrew Gooch.

Question 1: Sample Population and Randomization

Question 1(a)

The 2014 American Community Survey (ACS) estimates that the demographic breakdown (in terms of Race) in Florida is the following:

Race	Pct
White	56.6
Black	15.4
Hispanic	23.3
Other	4.7

How does the demographic breakdown of the sample of voters in our `fl` data compare to the ACS estimates in terms of race? Use the `race` variable to compute the proportion of racial categories. Briefly interpret the result.

Answer 1(a)

#Insert Answer Here

Question 1(b)

In terms of gender, the 2014 ACS estimates were that 51.1 percent of the population in Florida is female. How does our sample compare?

Answer 1(b)

#Insert Answer Here

Question 1(c)

In terms of turnout, the breakdown between 2006 and 2014 in Florida² was as follows:

Year	Pct
2006	47
2008	75
2010	49
2012	72
2014	51

How does our sample compare?

Answer 1(c)

#Insert Answer Here

²Available at <https://dos.myflorida.com/elections/data-statistics/elections-data/voter-turnout>

Question 1(d)

One reason why randomized experiments allow us to estimate the causal impact of an intervention is because, through randomization, the observable and unobservable characteristics of the subjects in the experiment are independent of, or at least uncorrelated with, treatment assignment. One consequence of this is that covariates will be “balanced” between the treatment and control group. You can check for balance by calculating the mean of a variable for the treatment group and for the control group.

Is the turnout history of subjects in the experiment between 2006 and 2012 balanced between treatment and control?

Answer 1(d)

#Insert Answer Here

Question 1(e)

Are gender, marriage, age and race balanced between treatment and control?

Answer 1(e)

#Insert Answer Here

Question 1(f)

What is the probability that the observed difference in 2012 turnout rate between treatment and control would occur due to random chance? Use permutation inference to determine this (use `set.seed(1234)` so your results are replicable).

Specifically, first compute the difference-in-means on `voted12` between the treatment and control group. Then, permute the treatment assignment 1000 times, and compute the difference-in-means based on each permuted treatment assignment. Finally, plot the histogram of difference-in-means over the 500 permutations. Show the observed difference-in-means (i.e., the actual value that we observed in the data) as the vertical line.

Compute the two-sided p-value based on the observed and permuted difference-in-means. Can we reject the null hypothesis of no effect on the past outcome?

Answer 1(f)

Question 2: Estimating Treatment Effects

In Question 2, we estimate the treatment effect using the `f1` data.

Question 2(a)

What is the overall 2014 turnout rate for people in the sample? What is the 2014 turnout rate for people in the treatment group? In the control group? What is the estimate of the average treatment effect (ATE) from the experiment? Using permutation inference, determine the statistical significance of the difference in turnout rates for people in the treatment group versus the control group (use 500 draws and `set.seed(02138)` so your results are replicable).

Answer 2(a)

#Insert answer here

Question 2(b)

The intervention involved sending a postcard with someone's past voting history on it and telling the recipient if he or she voted more than the typical person in the state. Would you hypothesize that the ATE is: (1) larger for people with a history of regular voting, (2) larger for people with a history of not voting, or (3) the same for both types? Why?

Answer 2(b)

Insert answer here.

Question 2(c)

Create a new variable called `vote_history` and store the individual-level propensity to vote by computing the proportion of turnout for each person between 2006 and 2012 (use `voted06`, `voted08`, `voted10`, and `voted12`).

Then, Determine the average past turnout in the state according to this measure by computing the mean of `vote_history`. Create a second new variable that indicates people with lower than average turnout in the Florida data.

Answer 2(c)

#Insert answer here

Question 2(d)

Estimate the average treatment effects for (1) those with lower past turnout than average in the state and (2) those with past turnout greater than or equal to the average in the state. Which effect is larger? Does this support your hypothesis from Question 2(b)?

Answer 2(d)

#Insert answer here

Question 2(e)

What do these results suggest about the ability of social-pressure based GOTV efforts to turn out voters who have not voted in the past?

Answer 2(e)

Insert answer here.

Question 2(f)

Estimate the difference in treatment effects for (1) men versus women and (2) people whose age is young versus middle versus older (18–29, 30–64 and 65+). Briefly interpret the result.

Answer 2(f)

#Insert answer here

Question 3: Optional Challenge Question

In the experiment conducted in 2014, the subjects' party registration statuses were not available. Suppose you would like to learn whether untargeted efforts at boosting turnout in Florida through the social pressure mailing helped the prospects of the Democratic Party or the Republican Party overall. Use the survey results from the CCES conducted in Florida to predict who leans towards the Republican party based on age, gender, race, and marital status. You may use stated party leanings (based on the variable `pid3`) as the outcome. Then, make predictions about who is likely to vote Republican (versus Democrat or Independent) in the experimental sample and calculate treatment effects for these groups. What do you find?

Answer 3

#Insert answer here