Political Science 207
Problem Set 1

Professor: Daniel Masterson

Due Monday April 12th at 5:00pm

Submit your completed assignment as a single PDF or HTML file (including your write-up and all R code) on GauchoSpace under the assignment for Week 2. I strongly recommend that you write your problem sets in R Markdown. (I recommend this tutorial.)

Make sure you follow good coding style (see tutorial in link), and show your work for problems that require calculations.

You are encouraged to work in groups, but you should write up the problem set alone, and you should note at the top of your problem set who you worked with.

Problem 1

This problem provides a quick review of the basic potential outcomes notation we discussed in class.

Suppose the government of Pakistan institutes a new reform that provides Rs. 1,000 per month to households in the country whose monthly income is below Rs. 5,000. The objective of the reform is to improve child enrollment in schools. It is believed that reducing household budget constraints (through the income supplement) will allow households to put more children in school.

a) Suppose we label this reform as a binary treatment $D \in \{0, 1\}$, for all households in Pakistan indexed by i. We call the share of children in school in every household Y. What is the meaning of Y_{1i} and Y_{0i}?

   The quantity Y_{1i} designates the outcome that household i would have in terms of the share of children in school Y *were it to receive the subsidy* (regardless of whether it actually receives the subsidy). Similarly, Y_{0i} is the non-treatment outcome for household i, regardless of whether or not it receives the subsidy.

b) Define the average treatment effect (ATE). Describe in words what the ATE means for the program. Let's say that analysts collected some survey data of a sub-sample of potential beneficiaries and estimated the average treatment effect. Let's call this $\widehat{ATE}$. What is the difference between the ATE and $\widehat{ATE}$?

   The ATE is $E[Y_1 - Y_0]$. It compares the hypothetical outcome under treatment and under non-treatment for *each household i*, and computes an expectation across all units. $\widehat{ATE}$ represents the specific estimate of the ATE based on a particular sample of households.

c) What is the Average Treatment on the Treated (ATT)? Describe in words what the ATT means for the program.

The ATT is defined as $E[Y_1 - Y_0|D = 1]$. In the scope of this program, the ATT represents the *average* effect of the subsidy on the share of children in school among households that actually receive the subsidy.

d) When will the ATT and the ATE be equal to each other? Prove it.

Let's see when the ATE and ATT are the same:

$$\begin{aligned} ATE &= ATT \\ E[Y_1 - Y_0] &= E[Y_1 - Y_0|D = 1] \\ E[Y_1] - E[Y_0] &= E[Y_1|D = 1] - E[Y_0|D = 1] \end{aligned}$$

This is guaranteed to be true when $E[Y_1] = E[Y_1|D = 1]$ and $E[Y_0] = E[Y_0|D = 1]$. This happens when $Y_1 \perp D$ and $Y_0 \perp D$. This is the case under randomized experiments, where we ensure that both potential outcomes Y_1 and Y_0 are independent of the treatment status D.

e) Show formally that $\tau_{ATE(x)} = E[Y_1 - Y_0|X = x]$ (i.e. a subgroup average treatment effect) is identifiable given random assignment.

We can estimate $\tau_{ATE(x)}$ by using the difference-in-means estimator within subgroup $X = x$. Since the treatment is randomly assigned among all units, it follows that it is also randomly assigned among any given subgroup of units.

$$\begin{aligned} ATE_X &= E[Y_1 - Y_0|X = x] \\ &= E[Y_1|X = x] - E[Y_0|X = x] \\ &= E[Y_1|X = x, D = 1] - E[Y_0|X = x, D = 0] \text{ by random assignment of D} \end{aligned}$$

Since the last line can be estimated from the data, identification has been achieved.

$$E[Y_1|X = x, D = 1] - E[Y_0|X = x, D = 0] = E[Y|X = x, D = 1] - E[Y|X = x, D = 0]$$

This shows that you can estimate the ATE within various sub-groups, and it will still be identified. In addition, while you may obtain a different quantity for every choice of X, you could also take the expectation of ATE_X over these different quantities and obtain a single number. We will return to this issue in blocking. Being in a particular block is like choosing $X = x$. You get a treatment effect for each block and then compute the average over all the blocks.

Problem 2

To reinforce the intuition behind the potential outcomes framework, consider the fictional data set "PO-data.csv." In these fictional data, we observe an outcome for each unit both under treatment and under control (which, again, is usually impossible in the real world).

(a) Write down the formula for individual level treatment effects and explain the fundamental problem of causal inference.

Individual-level treatment effects are $\tau_i = Y_{1i} - Y_{0i}$. The fundamental problem of causal inference is that we cannot observe both Y_{1i} and Y_{0i}, and therefore we cannot usually calculate τ_i (this fictional data set is an exception).

(b) Define the Average Treatment Effect (ATE) and calculate the ATE in these data.

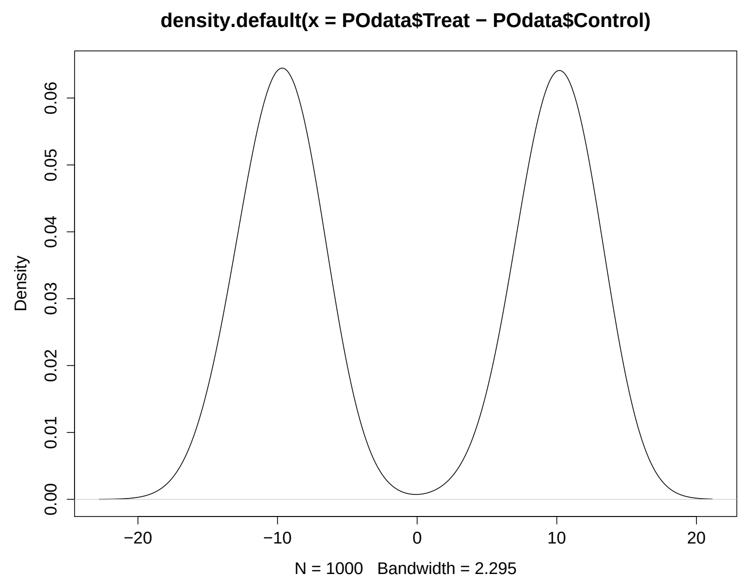
The ATE is defined as $\tau_{ATE} = E[Y_{1i} - Y_{0i}] = E[\tau_i]$.

```
> data <- read.csv("POdata.csv")
> ATE <- mean(data$Treat) - mean(data$Control)
[1] 0.4376972
```

(c) Plot the distribution of the individual treatment effects. Does the treatment seem to have an effect? How well is it captured by the ATE?

```
> plot(density(POdata$Treat-POdata$Control))
```

The distribution of the treatment effect is bimodal, with a peak at -10 and 10. The ATE does not capture this pattern, since it only computes an average across all units.


density.default(x = POdata$Treat − POdata$Control)
Density
0.00
0.01
0.02
0.03
0.04
0.05
0.06
−20
−10
0
10
20
N = 1000   Bandwidth = 2.295