PSC 794 Spring 2022 Problem Set 1

Please show your work via code to get complete credit.

Question 1 (1 points)

The following is a mini dataset of 10 people. The variable Y_i is whether they voted in the 2018 Mexican general election and the variable X_i is whether they voted in the 2012 Mexican general election. For both variables, 1 means the person voted and 0 means the person did not vote. Unfortunately, there are 3 missing observations for Y_i.

Person	Voted in 2018 (Y_i)	Voted in 2012 (X_i)
1	1	1
2	0	1
3	1	1
4	-99	0
5	1	0
6	1	1
7	0	0
8	-99	1
9	-99	0
10	1	1

Our goal is to estimate $E[Y_i]$ (the proportion of people who voted).

- Estimate the upper and lower Manski bounds for $E[Y_i]$
- Estimate $E[Y_i]$ assuming missing completely at random
- Estimate $E[Y_i]$ assuming missing at random conditional on X_i using the following imputation methods:
    - post-stratification
    - linear regression (OLS)

Question 2 (3 points)

A researcher collected information about 500 households in a town in New York State. These households were randomly sampled with equal probability from the town. Please see the following dataset:

Here are the variables in the dataset:

- `household_id`: household ID
- `number_of_people`: number of people in the houehold

- `head_has_job` : head of household is employed; TRUE = employed, FALSE = not unemployed
- `rents_housing` : the houehold rents their housing; TRUE = rents, FALSE = does not rent
- `household_income` : household income in USD
- `household_income_observed` : TRUE = observed; FALSE = missing

A) Estimate the town's mean household income

Please estimate the town's household income using the following imputation method:

- linear regression (OLS)
- inverse probability weighting (IPW), where you estimate the propensity score using logistic regression

B) Estimate the 95% confidence interval of the mean household income using the bootstrap method

For each of the imputation methods used, please estimate the 95% confidence interval of the mean using the bootstrap method. Make sure you generate at least 500 with-replacement samples.

Here is some helpful code:

https://cdsamii.github.io/cds-demos/quant2/bootstrap.html

C) Data visualization

Output your estimated means and 95% confidence intervals as a table and a coefficient plot. For the coefficient plot, please make it visually appealing and easy to understand.

Question 3 (2 points)

Suppose you want to do a field experiment in the town from Question 2 in which you randomly assign some households to receive a guaranteed income of 1000 USD each month for an entire year. Suppose you have 2400000 USD budgeted as a guaranteed income for this field experiment. How would you randomly assign the treatment (i.e., guaranteed income) if you were to do:

- **Complete randomization**
- **Block randomization** based on need: this is open to interpretation and you should explain your decision

Please explain your randomization procedure and implement your randomization using code. Give a summary table showing your random assignment results.

Question 4 (1 point)

The following is a mini dataset from a field experiment in which a researcher randomly assigned 5 students out of 10 to watch a documentary about the women's suffrage movement. The randomization procedure is complete randomization. All the students assigned to watch the documentary watched it, and all the students in the control group did not watch the documentary.

The variable D_i is whether the student watched the documentary (1 = yes, 0 = no). The variable Y_i is the students' self-reported political efficacy using a 5-point Likert scale (1 = very low, 5 = very high).

Student	Watched documentary (D_i)	Political efficacy (Y_i)
1	1	4
2	0	5
3	1	2
4	1	3
5	1	1
6	0	4
7	0	5
8	1	2
9	0	4
10	0	1

What is the average treatment effect (ATE) of watching the documentary?