Problem Set 6

BUSF-SHU 210: Business Analytics (Spring 2019)

1. Potential Outcomes Model

Consider the WeChat promotion problem discussed in class. Assume that we have 12 customers as shown in the following table:

Individual	W_i	$Y_i(1)$	$Y_i(0)$
1	0	1	0 (∗)
2	0	0	0 (∗)
3	0	1	0 (∗)
4	0	0	1 (∗)
5	0	1	1 (∗)
6	0	0	1 (∗)
7	1	1 (∗)	0
8	1	1 (∗)	1
9	1	1 (∗)	0
10	1	0 (∗)	0
11	1	0 (∗)	1
12	1	0 (∗)	0

Here, $W_i = 1$ (resp. $W_i = 0$) means promotion is received (resp. not received) for customer i. We call the individuals with $W = 1$ as in the treatment group, and the individuals with $W = 0$ as in the control group. $Y_i(0)$ is the outcome of customer i if promotion is not received, and $Y_i(1)$ is the outcome of customer i if promotion is received. $Y_i = 1$ means individual i is converted, whereas $Y_i = 0$ means the individual is not converted. The stared entries (∗) mean that we can observe them. Please briefly answer the following questions:

(a) What is the value of the estimated average treatment effect, $\widehat{ATE}$?

(b) What is the **true** average treatment effect ATE? You can use the sample average to estimate the average treatment effect.

(c) From parts (a) and (b), do you think $\widehat{ATE}$ is not a good estimator for ATE? Please use your own words to explain why or why not.

(d) We define $ATT = \mathbb{E}[Y(1) - Y(0)|W = 1]$ as the Average Treatment Effect for the treatment group, i.e., conditioned on $W_i = 1$, the average causal effect for the individuals in the treatment group. We also define $ATC = \mathbb{E}[Y(1) - Y(0)|W = 0]$ as the Average Treatment Effect for the control group, i.e., conditioned on $W_i = 0$, the average causal effect for the individuals in the control group. What are the values of ATT and ATC? You can use the sample average to estimate them.

(e) We define the selection bias for the controlled outcome as

$$SB(0) = \mathbb{E}[Y(0)|W = 1] - \mathbb{E}[Y(0)|W = 0]$$

and the selection bias for the treated outcome as:

$$SB(1) = \mathbb{E}[Y(1)|W = 1] - \mathbb{E}[Y(1)|W = 0]$$

What are the values of $SB(1)$ and $SB(0)$? You can use the sample average to estimate $SB(0)$ and $SB(1)$.

2. Balance Check for Sopcial Pressure Experiment

Use $t-$test and regression to finish the balance check of all treatments (Civic Duty, Hawthorne, Self, and Neighbors) associated with all covariates:

- *hh_size*

- *g2000*

- *g2002*

- *g2004*

- *p2000*

- *p2002*

- *p2004*

- *gender*

- *yob*

Is the sample balanced?

3. Monitoring Corruption Experiment

A randomized experiment was conducted in Indonesia to evaluate an effort to reduce corruption in road building projects. One of the treatments, which we focus on in this problem, sought to enhance participation at accountability meetings, the village-level meetings in which project officials account for how they spent project funds. Before construction began, residents in treated villages were encouraged to attend these meetings. The villages are randomly assigned to treatment or control groups.

The data set is `CorruptionExperiment.csv`, which has the following variables:

- *id*: The ID of the village.

- *pct.missing*: The outcome of interest to us, which measures the difference between what officials claimed they spent on road construction and an independent measure of expenditures.

- *treat.invite*: Treatment assignment mechanism, which takes a value of 1 if the village received the intervention and 0 if it did not.

- *head.edu*: The education of the village head.

- *mosques*: Mosques per 1,000 residents.

- *pct.poor*: The percentage of households below the poverty line.

- *total.budget*: The budget for each project.

Please answer the following questions.

(a) Is the sample balanced? Please apply balance check to address this question. Based on your results, do you think the experiment is well randomized?

(b) Estimate ATE and its standard deviation for the treatment considered in this experiment. Does the treatment have any causal effect on the quantity of interest? Please use both the average treatment estimator and linear regression to address this question.

(c) Re-estimate ATE using a regression specification that includes pre-treatment covariates (additively and linearly). Report your estimate of the treatment effect. Does this estimate differ significantly from the difference-in-means estimate? Explain why or why not.