PPHA 421: Problem Set 1

Professor: Koichiro Ito

March 26, 2022

Causality and RCT

Question 1

1. The Stable Unit Treatment Value Assumption (SUTVA) is required for a causal interpretation for an RCT, but it is not required for the difference-in-differences method. Is this true/false/uncertain? Explain why.

2. Consider a binary treatment D_i and outcome data (Y_i) for treated and untreated individuals. Suppose that the potential outcome for the treated status (Y_{1i}) is independent of D_i, but the potential outcome for the untreated status (Y_{0i}) is not independent of D_i. In this case, does the difference in $E[Y_i]$ between the two groups provide the average treatment effects (ATE), the average treatment effects on the treated (ATET), the average treatment effects on the untreated (ATEU), or none of them? Explain why.

3. Consider an RCT, where people in the treatment group are forced to take treatment. People in the control group are allowed to choose to stay untreated or take treatment. We estimate the local average treatment effects (LATE) by using the initial treatment assignment as an instrument for the actual treatment status. All assumptions required for the LATE are satisfied. Is the LATE in this case equivalent to the ATE, the ATET, the ATEU, or none of them? Explain why.

Question 2

Consider a RCT with a single treatment $D_i = \{0, 1\}$ with sample size of households N. A proportion P of the sample is treated, and ϵ_i is i.i.d., with variance σ^2. Consider estimating the following equation by OLS:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i.$$

If you need the general formula of the variance-covariance matrix of the OLS estimator, it is $\sigma^2(X'X)^{-1}$.

1. To minimize the variance of $\hat{\beta}_0$ (intercept), we should set $P = 0.5$. Is this true/false/uncertain? Explain why.

2. To minimize the variance of $\hat{\beta}_1$, we should set $P = 0.5$. Is this true/false/uncertain? Explain why.

3. Suppose that you currently have $P = 0.5$ with sample size N. There is another household available at no additional cost. So, you can make the sample size $N + 1$. Show whether including the additional household lowers or increases the variance of $\hat{\beta}_1$ (the coefficient for D_i).

4. Suppose that the cost of an untreated household is 1, and the cost of a treated household is c. Denote the number of treated households by T, the number of untreated households by U, and the total budget by B. Find the optimal ratio $\frac{T}{U}$ that minimizes the variance of $\hat{\beta}_1$ given the budget B.

5. You expect that some households in the treatment group are unlikely to comply with the treatment. To minimize the variance of the intention-to-treat estimate, the optimal P should be larger than the case with perfect compliance. Is this true/false/uncertain? Explain why.

6. You found that the mean of a covariate, which was collected before the experiment, is statistically different between households who were randomly assigned to the treatment group and those who were randomly assigned to the control group. This implies that the randomization was not correctly done. Is this true/false/uncertain? Explain why.

Question 3

You are estimating the effect of binary treatment T on outcome Y and using binary random assignment Z as an instrument. You allow for heterogeneous impact of the treatment.

1. Describe the monotonicity assumption of the instrument Z on T.

2. Provide an example of an empirical setting where the monotonicity assumption may not hold.

3. Will the first stage be likely to be stronger or weaker without monotonicity? Why?

Suppose that the monotonicity assumption fails. Then, there are both compliers (those who get treatment only when $Z_i = 1$) and definers (those who get treatment only when $Z_i = 0$). In this case, the LATE can be expressed as follows:

$$\beta_1 = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[T_i|Z_i = 1] - E[T_i|Z_i = 0]} = \frac{E[Y_{1i} - Y_{0i}|\text{complier}]P_C - E[Y_{1i} - Y_{0i}|\text{defier}]P_D}{P_C - P_D}$$

where:

$$\begin{aligned}
P(T_i = 1|Z_i = 1) &= P_A + P_C \text{ (Always Takers and Compliers)} \\
P(T_i = 1|Z_i = 0) &= P_A + P_D \text{ (Always Takers and Defiers)} \\
P(T_i = 0|Z_i = 1) &= P_N + P_D \text{ (Never Takers and Defiers)} \\
P(T_i = 0|Z_i = 0) &= P_A + P_D \text{ (Never Takers and Compliers)}
\end{aligned}$$

4. If we assume that the ATE is the same for compliers and the defiers, show how this expression for β_1 can be changed. Provide the interpretation of the LATE for this special case.

5. Show the proof of the denominator. That is, show that $E[T_i|Z_i = 1] - E[T_i|Z_i = 0] = P_C - P_D$.

6. Show the proof of the numerator. That is, show that $E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_{1i} - Y_{0i}|\text{complier}]P_C - E[Y_{1i} - Y_{0i}|\text{defier}]P_D$.