Causal Inference - Homework 1

Introduction to Causal Inference 097400
Winter 2020/2021

Submission date: 11 November, 2020

Question 1

Using potential outcomes notation, give examples of data generating processes (a joint distribution) which include a binary treatment T, and two potential outcomes Y_0 and Y_1, such that:

1. $\mathbb{E}[Y_1 - Y_0] = \mathbb{E}[Y|T = 1] - \mathbb{E}[Y|T = 0]$
2. $\mathbb{E}[Y_1 - Y_0] \neq \mathbb{E}[Y|T = 1] - \mathbb{E}[Y|T = 0]$,

where $Y = T \cdot Y_1 + (1 - T) \cdot Y_0$.

Question 2

Let $(t_1, y_1), \ldots, (t_n, y_n)$ be a sample from a randomized controlled trial (RCT), where for each $i = 1, \ldots, n$, $t_i \in \{0, 1\}$ is a binary treatment, and $y_i \in \mathbb{R}$ is an outcome measured after the treatment. We define $ATT := \mathbb{E}[Y_1 - Y_0 \mid T = 1]$ to be the Average Treatment Effect on the Treated population (ATT). Using potential outcome notation, prove that ATT can be estimated from data gathered as the sample above. Please be clear about which assumptions you have used.

Hint: Think about the relationship between ATT and ATE in this settings.

Question 3

Give an example of a real-world dataset with features X and one or more observed outcome variables Y. For this dataset give:

1. Two examples of interesting causal questions relating one of the features and one of the outcomes. Explain what would be the treatment and what would be the *potential* outcomes in this case.

2. Two examples of interesting prediction questions which do not require causal reasoning.

Examples can come from the fields of politics, biology, sports, economics, entertainment, medicine, transportation and so on - use your imagination. You might find the article " *A Second Chance to Get Causal Inference Right: A Classification of Data Science Task*" by Miguel Hernán (available on the course Moodle) to be helpful here, though please do not use the exact same examples.