

## Homework1: Introduction to Causal Inference

1 (20'). Supposed that you have a random sample of size  $n$  from the population of interest. Answer the following questions that are designed to help you get familiar with potential outcomes. Try to keep your answers brief and your language precise. Throughout the problem, assume that the Stable Unit Treatment Value Assumption (SUTVA) holds.

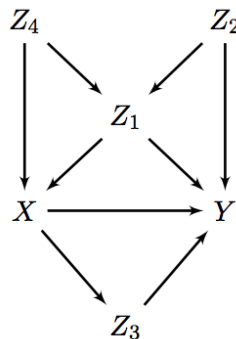
- a) Contrast the meaning of  $Y^0$  with the meaning of  $Y$
- b) Contrast the meaning of  $E(Y^0)$  with the meaning of  $E(Y|A = 0)$
- c) Contrast the meaning of  $E(Y^0|A = 1)$  with the meaning of  $E(Y^0|A = 0)$
- d) Which of the following quantities can be identified from observed data, assuming SUTVA?
  - a.  $E(Y^0|A = 1)$
  - b.  $E(Y^0)$
  - c.  $E(Y|A = 0)$
  - d.  $E(Y^0|A = 0)$
- e) Now, further assume that the units in this sample are randomly assigned to treatments, which means the assumption of exchangeability holds. Which of the above quantities can be identified from the observed data?

2 (10'). Under the complete randomization of treatment assignment, the average treatment effect is identified because the randomization guarantees statistical independence between the treatment indicator  $A$  and the observed outcome  $Y$ .

- a) True
- b) False

Why? Justify your choice.

3 (20'). Consider the following DGA



- a) Enumerate all paths from  $X$  to  $Y$
- b) In the path  $X \leftarrow Z_4 \rightarrow Z_1 \leftarrow Z_2 \rightarrow Y$ , what type of node is  $Z_1$ ? Does conditioning on  $z_1$  block or unblock this path from  $X$  to  $Y$ ?
- c) In the path  $X \leftarrow Z_1 \rightarrow Y$ . In this path, what type of node is  $Z_1$ ? Does conditioning on  $Z_1$  here block or unblock this path?

- d) Pearl's back-door criterion is a way to rule out confounding via conditioning, thus identifying the effect of one variable on another. For identifying the effect of any  $X$  on any  $Y$ , the backdoor criterion has two parts. First, the conditioning set  $C$  may not include any descendent of  $X$ . That means, anything that  $X$  affects cannot be conditioned on (c.f. post-treatment bias). Second, the conditioning set must block all back-door paths from  $X$  to  $Y$ . That is, by conditioning on our set  $C$ , we should break any back-door paths that may simultaneously generate confounded covariance in  $X$  and  $Y$ .
- Is  $X \rightarrow Z_3 \rightarrow Y$  a back door path from  $X$  to  $Y$ ? Why?
  - Based on your DAG, enumerate the minimum conditioning sets that satisfy the back door criterion for identifying the effect of  $X$  on  $Y$ ?

4. Consider a randomized experiment with four observations, of which two units were randomly assigned to treatment via complete randomization. We use  $A_i \in \{0, 1\}$  and  $Y_i$  to denote the treatment (1 for treatment and 0 for control) and the observed outcome for unit  $i$ , respectively.

a) (4') The table below shows the data observed from this experiment, augmented with columns for potential outcomes and the treatment effect for each unit. Fill in all the empty cells in the table based on the observed information, denoting unknown information with “?”

$i$	$Y$	$A$	$Y^1$	$Y^0$	$\tau$
1	2	1			
2	0	0			
3	1	0			
4	3	1			

b) (6') Define the population average treatment effect for the treated (ATT) using the above notation and propose an unbiased estimator for this estimand. Then using the data in the table estimate this quantity.

5. (10') Researchers are studying the causal effect of education level on income. They believe that job type is a mediator (i.e., higher education leads to better job opportunities, which in turn increases income). Additionally, they suspect that family background influences both education and income.
- Draw a DAG that includes the above variables.
  - Is Job Type a mediator or confounder in this DAG?
  - How would you use the DAG to assess the causal effect of Education Level on Income?
6. (10') You are tasked with investigating the causal effect of air pollution on asthma in children. Consider the following:
- Parental Smoking increases the likelihood of both air pollution exposure and asthma in children.
  - Living in an urban area increases exposure to air pollution but is not directly related to asthma.

- a) Draw a DAG including the above variables.
- b) What are the confounders you should adjust for?
- c) Should you adjust for Urban Living? Explain why or why not based on your DAG.