

Statistics 256: Causality

In-class Activity 5: Selection on Observables — Identification

Name(s): _____

Question 1: Positivity in the wild

A researcher estimating the effect of a job-training program on earnings conditions on $X =$ (age, gender, prior-earnings quartile).

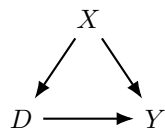
- (i) Would we believe CI here? Why or why not?

- (ii) In the cell {age 55+, male, top quartile} there are 0 treated units (though several control units). What problem does this cause?

- (iii) List one or two things you could do to address the last problem.

Question 2: Seeing CI on the graph

Consider the following DAG. X is observed; D is treatment; Y is the outcome.



- (a) Draw $G_{\underline{D}}$ (the test graph: delete all arrows out of D ; relabel the outcome as Y_d). How can you use this to argue for conditional ignorability on the graph? (What test do you do; what is the CI statement you are looking to support?).

- (b) Draw the SWIG under $do(D = d)$ (split the D node as $D \mid d$; descendants of D become subscripted). Using d-separation in the SWIG, can you support CI?

- (c) Now suppose there is *also* an unobserved variable U that causes both D and Y . Using either the test-graph or SWIG, see if conditioning on X alone still gives CI?

Question 3: A job-training study

A program (D) may affect earnings (Y). You have pre-treatment data on age, education, and average earnings over the prior 5 years.

- (a) Assuming $X = (\text{age, education, prior earnings})$ blocks all backdoor paths from D to Y , write conditional ignorability in PO notation using this X .

- (b) Give an example of an *unobserved* variable that would plausibly violate CI. (Hint: what might affect both program participation and earnings but is not in X ?)

- (c) Suppose a colleague adds “whether the participant found a job within 3 months of program start” to X . Is this going to help or hurt, and why?