PS200C: Quantitative Methods

In-class Activity: Selection on Observables

Name: _____

Question 1: The identification argument

Suppose you assume conditional ignorability: $\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp D_i \mid X_i$.

(a) In one sentence, what does this assumption say in plain English?

(b) What are some other names for this assumption?

Question 2: Evaluating SOO claims

For each study, state what the SOO/CI assumption would require, and whether you find it plausible.

(a) A researcher estimates the effect of college attendance on earnings by comparing college-goers to non-goers, "controlling for" high school GPA, parental income, and state of residence.

(b) Blattman & Annan (2009) estimate the effect of abduction by the LRA on education, conditioning on village location and age, arguing that abduction was quasi-random conditional on these variables.

Question 3: Post-treatment variables

A randomized job training program (D) aims to increase earnings (Y). A researcher notices that some treated individuals dropped out of the program early, and restricts the analysis to "completers" vs. controls.

(a) Is "program completion" a pre-treatment or post-treatment variable?

(b) Why could restricting to completers bias the estimate, even though D was randomly assigned? (Think about what kinds of people complete vs. drop out.)

Question 4: Subclassification

You have two strata defined by a covariate $X \in \{\text{low}, \text{high}\}$:

Stratum	$\bar{Y}$ (treated)	$\bar{Y}$ (control)	N_j (total)
$X = \text{low}$	12	8	60
$X = \text{high}$	20	18	40
Total			100

(a) Compute the subclassification estimate of the ATE.

(b) Suppose the treated group is disproportionately in the "high" stratum. Would you expect the unadjusted DIM to overestimate or underestimate the ATE? Why?

(c) Under what assumption is your answer to (a) a valid causal effect?