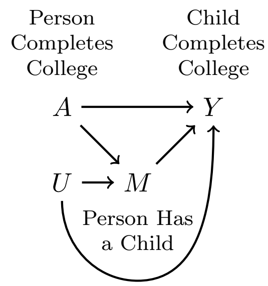
Info 6751. Fall 2022. 1 Nov. Principal Stratification Exercise

Education promotes financial well-being and the ability to pass on opportunities to one's children. Meanwhile, college also takes time and may reduce the probability of having children at all.



This creates an interesting causal structure: Y is meaningless if $M = 0$.

To focus on that problem, we will assume throughout that A is unconfounded: pretend that college is randomly assigned. Having a child M, however, may be confounded.

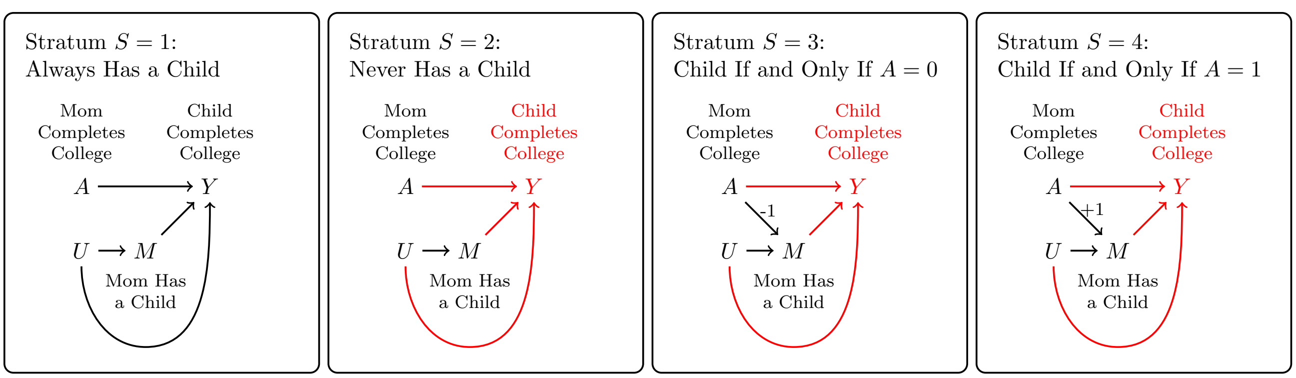
To solve this conundrum, suppose the population is comprised of four principal strata, which are not observed.

- $S = 1$: Those who would have a child regardless of whether they finished college
- $S = 2$: Those who would not have a child regardless of whether they finished college
- $S = 3$: Those who would have a child if and only if they did not complete college
- $S = 4$: Those who would have a child if and only if they completed college

These strata are population subgroups that exist before treatment is assigned.

Question 1. What is the causal effect of college on fertility in each of the four strata?

Note that within strata 1 and 2, M is not post-treatment because it is not caused by A. If we could restrict to these strata, we could condition on M without inducing collider bias. This is why principal stratificaton is a helpful idea: it reconceptualizes a post-treatment problem into pre-treatment groups.



(I am using red in the above to denote components that are undefined when $M = 0$)

Notation and a table to fill in over the exercise

Let $\pi_s = P(S = s)$ denote the proportion of the population in stratum s.

- For instance, μ_1^0 is the proportion of mothers in stratum 1 (subscript $s = 1$) whose child would complete college if the mom did not complete college (superscript $a = 0$).

Let $\mu_s^a = E(Y^a \mid S = s)$ denote the mean outcome under treatment a in stratum s.

- For instance, μ_1^0 is the proportion of mothers in stratum 1 (subscript $s = 1$) whose child would complete college if the mom did not complete college (superscript $a = 0$).

Your goal will be to fill in the following table

	Population Proportion π_S	Outcome Under Mom College μ_S^1	Outcome Under Mom No College μ_S^0
Stratum $S = 1$ (Always Has Child)			
Stratum $S = 2$ (Never Has Child)			
Stratum $S = 3$ (Child Only if No College)			
Stratum $S = 4$ (Child Only if College)			

Question 2. Start by putting NA in the 4 cells where μ_s^a is undefined.

Assumptions

We will make two key assumptions:

- Exchangeability of A: Think of college as randomly assigned
- Monotonicity: College (if anything) reduces the probability of having a child. It never causes people to have a child.

Question 3. Under monotonicity, which stratum is empty? Fill in $\pi_s = 0$ in that stratum.

Observable data

We observe the following data.[1]

	Proportion who have a child	Among those who have a child, proportion whose child completes college
Among women who completed college	71%	40%
Among women who did not complete college	83%	18%

Question 4. One group we observe are mothers with a college degree. From which strat(um/a) do these women come?

Question 5. One group we observe are mothers without a college degree. From which strat(um/a) do these women come?

[1] To calculate these data, I analyzed the National Longitudinal Survey of Youth 1979 cohort merged with children born to mothers in that cohort, who are surveyed in the Child and Young Adult Supplement. These numbers represent women ages 14–22 in 1979 who form the mother generation, and each of these mothers is equally weighted in the child generation regardless of how many children she has.

Identifying stratum sizes

Question 6. Under the assumptions of (1) exchangeability of college degrees and (2) monotonicity, we can identify π_1: the proportion of the population who would have a child regardless of college. Fill in an estimate for $\hat{\pi}_1$.

Question 7. Under those same assumptions, we can identify another number which is $\pi_1 + \pi_3$. What is that number? Given the number and your estimate $\hat{\pi}_1$, estimate $\hat{\pi}_3$.

Identifying stratum outcomes

Question 8. Can you identify μ_1^1 with observable data? Fill in an estimate $\hat{\mu}_1^1$.

Question 9. Identifying μ_1^0 is harder because the moms observed without a college degree are a mix of two strata, and we only want one of those strata. Let $\bar{y}^0$ be the proportion of moms without a college degree whose children completed college. Write a formula for $\bar{y}^0$ as a function of $\mu_1^0, \mu_3^0, \pi_1, \pi_3$.

Rearrange your formula so that the target quantity μ_1^0 is on the left. Which term on the right remains unknown?

Set identifying the stratum average causal effect

Question 10. Plug in extreme values (1 and 0) for the unknown term. Produce an interval estimate for $E(Y^1 - Y^0 \mid S = 1) = \mu_1^1 - \mu_1^0$.