

Statistics 256: Causality

In-class Activity 4: Experiments

Name(s): _____

Question 1: The identification result

Suppose treatment D_i is randomly assigned so that $(Y_{1i}, Y_{0i}) \perp\!\!\!\perp D_i$.

- (a) Starting from the ATE, $\tau = \mathbb{E}[Y_{1i} - Y_{0i}]$, show step by step that τ equals the difference-in-means estimand $\mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0]$. At each step, name which assumption you are using (ignorability or consistency).
- (b) Does this identification result also hold for the ATT? That is, does $\mathbb{E}[Y_{1i} - Y_{0i} | D_i = 1] = \mathbb{E}[Y_i | D_i = 1] - \mathbb{E}[Y_i | D_i = 0]$ under the same assumption? Try it out.

Question 2: Fisher's exact test

Six patients are randomly assigned to drug ($D = 1$) or placebo ($D = 0$), with exactly 3 treated. The observed outcomes are:

i	D_i	Y_i
1	1	7
2	1	4
3	1	5
4	0	2
5	0	3
6	0	1

- (a) Compute $\hat{\tau}_{\text{obs}}$ (the observed difference in means).
- (b) Under the sharp null $H_0 : Y_{1i} = Y_{0i}$ for all i , fill in the full science table (both potential outcomes for every unit).

- (c) How many distinct assignments of exactly 3 out of 6 to treatment are there?
- (d) Pick two alternative assignments and compute $\hat{\tau}(\omega)$ for each under the sharp null.
- (e) Without enumerating all of them: would you expect the two-sided p -value to be small or large? Why?

Question 3: Design and analysis choices

- (a) You run an experiment where treatment is assigned at the village level but outcomes are measured per individual. Your HC2 standard errors treat each individual as independent. What goes wrong, and what should you use instead?
- (b) You have a strong pre-treatment predictor of your outcome (e.g., last year's test score). Name two ways to exploit it to improve precision, and say which is a *design* choice vs. an *analysis* choice.
- (c) A colleague runs $\text{lm}(Y \sim D + X)$ in a randomized experiment. Under what condition on treatment effects is this fine? What would you recommend instead, and why?