PHS 2000B Problem Set 1

Counterfactuals and Causal Effects

Due: Thursday, February 6, 2025

1 Counterfactuals (10 points)

In December 2019, the New York Times published an article[1] titled *"Another Benefit to Going to Museums? You May Live Longer"* that briefly went viral on Twitter. The article recounted findings from a longitudinal study[2] published in the British Medical Journal: participants who reported engaging in arts activities, like going to the museum or attending a concert at least once or twice a year, had rates of death that were 14 percentage points lower than those who did not (HR = 0.86, 95% CI: 0.77 to 0.96). The study provoked considerable outrage both because the title seemed shamelessly designed to stroke the ego of a certain segment of the Times readership and because the results seemed likely to be subject to residual confounding.

For simplicity, let's consider the relationship between death (represented by binary random variable Y) and exposure to arts (represented by binary random variable A), setting aside the complexity of the study's time-to-event outcomes. We will jointly denote the set of observed confounding factors controlled for in the study by the vector L.

1. Define (in words **and** symbols) the relevant counterfactual/potential outcomes necessary to make the causal contrast between arts exposure and mortality that the study is interested in. (2 points)
2. The study is observational, but the authors *adjusted* their estimates for a number of possible confounding factors. What assumptions are necessary to infer that the reported estimates are causal? Describe these assumptions in words and symbols. (2 points)
3. For each assumption you mentioned above, what is one example of how the assumption could be violated in this setting? (2 points)
4. Which assumption would be violated by "residual confounding"? Draw a DAG showing how this would lead to bias. (2 points)
5. Think back to Issa's comments on the philosophy of causal inference. Do you think the exposure in this study is defined clearly enough to draw causal conclusions from the data? Justify your answer. (2 points)

[1] https://www.nytimes.com/2019/12/22/us/arts-health-effects-ucl-study.html?smtyp=cur&smid=tw-nythealth

[2] https://www.bmj.com/content/367/bmj.l6377

2 Conditional and Marginal Treatment Effects (14 points)

Let's say we are interested in studying the causal effect of a housing subsidy on educational outcomes. For a small cohort of Boston teenagers, we obtain the following data on whether they received the subsidy ($A = 1$ subsidy; $A = 0$ no subsidy) and whether they graduated from high school ($Y = 1$ graduated; $Y = 0$ didn't graduate). Suppose you have a time machine and can go back in time to observe both counterfactual/potential outcomes for each person.

You suspect that those with parents who work evening shifts may not have been able to attend the meeting announcing the subsidy and therefore may have been less likely to participate. You also think that there may be meaningful differences in the impact of housing subsidies in rural and urban areas. Thus, you collect information on the following covariates: parental employment status L_1 ($L_1 = 1$ if parents work the night shift; $L_1 = 0$ if parents do not work the night shift), and rural/urban environment L_2 ($L_2 = 1$ if the teen lives in a rural environment; $L_2 = 0$ if the teen lives in an urban environment).

You record the following data:

ID	L_1	L_2	A	Y
1	0	0	0	0
2	0	1	0	1
3	1	0	0	0
4	1	1	0	0
5	1	0	0	1
6	1	1	0	1
7	1	0	0	1
8	1	1	0	0
9	0	0	1	0
10	0	1	1	1
11	0	0	1	1
12	0	1	1	1
13	0	0	1	0
14	0	1	1	0
15	1	0	1	1
16	1	1	1	1

1. In plain language, define $P(Y = 1|A = a, L_1 = l_1, L_2 = l_2)$ in the context of this study. (2 points)
2. Compute the **conditional average treatment effects** (CATE: $E[Y^1|L_1 = l_1, L_2 = l_2] - E[Y^0|L_1 = l_1, L_2 = l_2]$) for each combination of covariates using the following formulas. You may assume that the three identifiability conditions hold. Please show your work. Interpretations are not required. (2 points)
   - $E[Y = 1|A = 1, L_1 = 0, L_2 = 0] - E[Y = 1|A = 0, L_1 = 0, L_2 = 0]$
   - $E[Y = 1|A = 1, L_1 = 0, L_2 = 1] - E[Y = 1|A = 0, L_1 = 0, L_2 = 1]$
   - $E[Y = 1|A = 1, L_1 = 1, L_2 = 0] - E[Y = 1|A = 0, L_1 = 1, L_2 = 0]$
   - $E[Y = 1|A = 1, L_1 = 1, L_2 = 1] - E[Y = 1|A = 0, L_1 = 1, L_2 = 1]$
3. You are not satisfied with having only the conditional average treatment effects. Now, we want to understand how marginalization works. First, let's do this the long way using the g-formula.

   Recall the g-formula as Issa presented it to us:

   $$E(Y^a) = E\left[E(Y|L, A = a)\right]$$

   which, for categorical L, we can write as:

$$\sum_l E(Y|A=a, L=l)P(L=l)$$

Let's start by computing the probability of each covariate pattern, $P(L=l) = P(L_1 = l_1, L_2 = l_2)$ from the table above. (2 points)

- $P(L_1 = 0, L_2 = 0) =$
- $P(L_1 = 0, L_2 = 1) =$
- $P(L_1 = 1, L_2 = 0) =$
- $P(L_1 = 1, L_2 = 1) =$

4. By hand, using the g-formula (below), compute the **marginal average treatment effect** (ATE). (2 points)

$$\sum_l \Big[E[Y|A=1, L_1 = l_1, L_2 = l_2] - E[Y|A=0, L_1 = l_1, L_2 = l_2]\Big] P(L_1 = l_1, L_2 = l_2)$$

5. Now we will approach the problem using inverse probability weighting (IPW). Recall that in lecture, Issa showed that the IPW estimator of the ATE is equivalent to the g-formula estimator under certain conditions:

$$ATE = E\big[E(Y|L, A=1) - E(Y|L, A=0)\big] = E\left[\frac{I(A=1)Y}{P(A=1|L)}\right] - E\left[\frac{I(A=0)Y}{P(A=0|L)}\right]$$

Now let's calculate the inverse probability weights,

$$\frac{1}{P(A=1|L)} \quad \text{and} \quad \frac{1}{P(A=0|L)}$$

Fill in the missing IPWs in the table below. Make sure you pay attention to $A=1$ and $A=0$ to write in the appropriate IPW. (4 points)

ID	L1	L2	A	Y	IPW
1	0	0	0	0	
2	0	1	0	1	
3	1	0	0	0	
4	1	1	0	0	
5	1	0	0	1	
6	1	1	0	1	
7	1	0	0	1	
8	1	1	0	0	
9	0	0	1	0	
10	0	1	1	1	
11	0	0	1	1	
12	0	1	1	1	
13	0	0	1	0	
14	0	1	1	0	
15	1	0	1	1	
16	1	1	1	1	

6. Now calculate the marginal average treatment effect using these weights and the formula,

$$ATE = E\left[\frac{I(A=1)Y}{P(A=1|L)}\right] - E\left[\frac{I(A=0)Y}{P(A=0|L)}\right]$$

How do your results from using IPW compare to those obtained using the g-formula? (2 points)