Chapter 16 Solutions

Andrew Wu
Wasserman: All of Statistics

August 12, 2025

Problem 16.1. Create an example in which $\alpha > 0$ and $\theta < 0$. Here, α is the association, and θ is the average causal effect.

Solution. Here's a table in the form of Example 16.2. As usual, the asterisks denote unobserved values.

X	Y	C_0	C_1
0	1	1	1*
0	0	0	0*
0	0	0	0*
0	0	0	0*
1	1	1*	1
1	1	1*	1
1	1	1*	1
1	0	1*	0

We have

$$\begin{aligned}\theta &= \mathbb{E}(C_1) - \mathbb{E}(C_0)\\ &= \frac{1}{2} - \frac{5}{8} = -\frac{1}{4}\end{aligned}$$

and

$$\begin{aligned}\alpha &= \mathbb{E}(Y|X=1) - \mathbb{E}(Y|X=0)\\ &= \frac{3}{4} - \frac{1}{4} = \frac{1}{2}\end{aligned}$$

so indeed, $\alpha > 0$ and $\theta < 0$. □

Problem 16.2. Let's generalize beyond the binary case. Suppose that X is some random variable. For example, X could be the dose of a drug, in which case $X \in \mathbb{R}$. The counterfactual function $C(x)$ returns the outcome a subject would have if they received dose x. The observed response is given by $Y \equiv C(X)$.

The causal regression function is $\theta(x) = \mathbb{E}(C(x))$, and the regression function, which measures association, is $r(x) = \mathbb{E}(Y|X=x)$.

Prove that when X is randomly assigned, then $\theta(x) = r(x)$, though in general $\theta(x) \neq r(x)$.

Solution. First, we'll present an example where $\theta(x) \neq r(x)$.

Suppose that $X \in [0,1]$. Take $X_1 = 0, X_2 = 0.5, X_3 = 1, X_4 = 0.5$. Let the counterfactual function for X_i be denoted by $C_i(x)$.

Then $\theta(x) = \frac{1}{4}(C_1(x)+C_2(x)+C_3(x)+C_4(x))$, and $r(x) = \mathbb{E}(C(X)|X=x)$. Suppose that the functions C_i are all constant over all values of x, and that $C_1(x) = 2$, $C_2(x) = 1$, $C_3(x) = 0$, and $C_4(x) = 5$; that is, there is no causal effect.

Then $\theta(x) = 2$ for all x. But $r(0.5) = \mathbb{E}(C(X)|X=0.5) = \frac{1}{2}(1+5) = 3$, so $\theta(x) \neq r(x)$. Here $\theta(x)$ is the average potential outcome under treatment x in the entire population X_1 through X_4; $r(x)$ is the average observed outcome under treatment x for the population observed to have undergone treatment x, which in the case of $x = 0.5$ is just X_2 and X_4.

Now we'll show that when X is randomly assigned, $\theta(x) = r(x)$.

We have

$$\begin{aligned}
r(x) &= \mathbb{E}(Y|X = x) \\
&= \sum_y y f_{Y|X}(y|x) \\
&= \sum_y y \frac{f_{X,Y}(x, y)}{f_X(x)} \\
&= \sum_y y f_Y(y)
\end{aligned}$$

where the last step follows from $f_{X,Y}(x, y) = f_X(x)f_Y(y)$, a consequence of assigning X randomly. But $\mathbb{E}(Y) = \sum_y y f_Y(y)$, so we are done. □

Problem 16.3. Suppose you are given data $(X_1, Y_1), \ldots, (X_n, Y_n)$ from an observational study, where $X_i \in \{0, 1\}$ and $Y_i \in \{0, 1\}$. Although it is not possible to estimate the causal effect θ, it is possible to put bounds on θ. Find upper and lower bounds on θ that can be consistently estimated from the data. Show that the bounds have width 1.

Solution. We have $\theta = \mathbb{E}(C_1) - \mathbb{E}(C_0)$, where $Y = C_0$ if $X = 0$ and $Y = C_1$ if $X = 1$. As usual, we don't observe C_1 when $X = 0$ and we don't observe C_0 when $X = 1$, giving us unobserved values.

Let's first try to maximize θ. We do this by maximizing $\mathbb{E}(C_1)$ and minimizing $\mathbb{E}(C_0)$. Therefore, we ought to set $C_1 = 1$ for all the X_is with $X_i = 0$, and we ought to set $C_0 = 0$ for all the X_is with $X_i = 1$.

Suppose that there are m X_is with $X_i = 0$ and $n - m$ with $X_i = 1$. Suppose furthermore that there are z_0 C_0s with $X_i = 0$ and $C_0 = 1$ and z_1 C_1s with $X_i = 1$ and $C_1 = 1$.

Then $\mathbb{E}(C_1) = \frac{1}{n}(m + z_1)$ and $\mathbb{E}(C_0) = \frac{z_0}{n}$. That yields a maximum $\theta = \frac{m+z_1-z_0}{n}$.

Similarly, to minimize θ, we want to minimize $\mathbb{E}(C_1)$ and maximize $\mathbb{E}(C_0)$. We set $C_1 = 0$ for all the X_is with $X_i = 0$ and $C_0 = 1$ for all the X_is with $X_i = 1$. That yields a minimum $\theta = \frac{z_1+m-z_0-n}{n}$.

The width of the bound is thus $\frac{m+z_1-z_0}{n} - \frac{z_1+m-z_0-n}{n} = \frac{n}{n} = 1$. □

Problem 16.4. Suppose that $X \in \mathbb{R}$, and that, for each subject i, $C_i(x) = \beta_{1i}x$. Each subject has their own slope β_{1i}. Construct a joint distribution on (β_1, X) such that $\mathbb{P}(\beta_1 > 0) = 1$ but $\mathbb{E}(Y|X = x)$ is a decreasing function of x, where $Y = C(X)$. Interpret.

Solution. Suppose that $\beta_1 \sim \text{Uniform}(0, 1)$. Let $X = \frac{1}{\sqrt{\beta_1}}$. Evidently, $\mathbb{P}(\beta_1 > 0) = 1$. Moreover, $Y = C(X) = \beta_1 X$, and using $\beta_1 = \frac{1}{X^2}$, we obtain $Y = \frac{1}{X}$; thus $\mathbb{E}(Y|X = x) = \frac{1}{x}$ is a decreasing function of x.

Here, the interpretation is that the association is negative ($\mathbb{E}(Y|X = x)$ is a decreasing function of x) but the causal effect is positive. □

Problem 16.5. Let $X \in \{0, 1\}$ be a binary treatment variable, and let (C_0, C_1) denote the corresponding potential outcomes. Let $Y = C_X$ denote the observed response. Let F_0 and F_1 be the cumulative distribution functions for C_0 and C_1. Assume that F_0 and F_1 are both continuous and strictly increasing. Let $\theta = m_1 - m_0$ where $m_0 = F_0^{-1}(1/2)$ is the median of C_0 and $m_1 = F_1^{-1}(1/2)$ is the median of C_1. Suppose that the treatment X is assigned randomly. Find an expression for θ involving only the joint distribution of X and Y.

Solution. We have

$$\begin{aligned}
F_0(t) &= \mathbb{P}(C_0 \le t) \\
&= \mathbb{P}(X = 0)\mathbb{P}(Y \le t|X = 0) + \mathbb{P}(X = 1)\mathbb{P}(C_0 \le t|X = 1) \\
&= \mathbb{P}(X = 0)\mathbb{P}(Y \le t|X = 0) + \mathbb{P}(X = 1)\mathbb{P}(C_0 \le t|X = 0) \\
&= \mathbb{P}(Y \le t|X = 0)
\end{aligned}$$

and similarly $F_1(t) = \mathbb{P}(Y \leq t|X = 1)$. The derivation above follows from

$$\mathbb{P}(C_0 \leq t|X = 0) = \mathbb{P}(C_0 \leq t|X = 1),$$

which is due to X being assigned randomly. (Essentially, the equation says that the underlying distribution of the outcomes for treatment 0 for both the groups—the group that in reality got treatment 0 and the group that in reality got treatment 1—is the same.)

Then m_0, the median of C_0, is the unique number such that $F_0(m_0) = \frac{1}{2}$. (We can say this rather than citing infimum conditions due to knowing that F_0 and F_1 are strictly increasing.) So it is also the unique number such that $F_{Y|X}(m_0|0) = \frac{1}{2}$. Using similar methods for m_1, it follows that

$$\begin{aligned}\theta &= m_1 - m_0 \\ &= F_{Y|X}^{-1}(0.5|1) - F_{Y|X}^{-1}(0.5|0).\end{aligned}$$

□