PHS2000B: Causal Effects and Regression

01/30/2023

Contents

1 Introduction

The purpose of this handout is to delve deeper into the identification of causal effects from observed data, and connect identification of causal effects to regression methods. We will explore a toy example using observational data to understand the effect of statin, a cholesterol-lowering medication, on blood cholesterol levels. We will posit that biological sex is the only confounder on the statin - cholesterol level relationship, such that controlling for biological sex will allow us to identify causal effects.

Variable	Description	Coding
Y	Blood cholesterol level	Continuous
A	Statin therapy	0 Untreated; 1 Treated
C	Biological sex	0 Male; 1 Female

Based on the below DAG, we can see that we must condition on C to block the back-door, non-causal path from $C \rightarrow Y$.

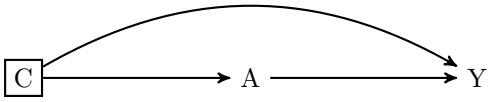

Figure 1: DAG

We can transform the DAG into a Single World Intervention Graph (SWIG) by splitting the treatment node into the observed data, A, and the counterfactual a, which represents what would happen if we forced all individuals to take a certain level of the treatment (a). For more information on SWIGs, please refer to the *What If* textbook, Chapter 7. Below we see that conditioning on C gives us the exchangeability, $Y^a \perp\!\!\!\perp A|C$. This conditional exchangeability will be key in identifying our causal effects.

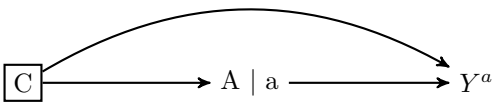

Figure 2: SWIG

2 Identifying Causal Effects

Our research question, what is the effect of statin therapy on blood cholesterol level, can be expressed in counterfactual notation as $E[Y^{a=1} - Y^{a=0}]$. Note that we are interested in the marginal causal effect, i.e., the effect of statins on cholesterol in the total population, had everyone been treated compared to had no one been treated. When we talk about *identification* of causal effects, what we mean is re-writing counterfactuals as observed data. Recall that we need three causal assumptions - *consistency*, *exchangeability*, and *positivity* - to allow us to *identify* causal effects.

2.1 Assumptions

2.1.1 Consistency

Under consistency, an individual with observed treatment $A = a$ has the observed outcome Y equal to their counterfactual outcome Y^a when $A = a$ for that individual. Considering a specific level $a = 1$, consistency allows us to equate an individual's counterfactual outcome when we force them to have treatment, to an individual's observed outcome given that individual actually received treatment, i.e., $(Y^{a=1}|A = 1) = (Y|A = 1)$. We can extend our use of consistency to the population to say that $E[Y^{a=1}|A = 1] = E[Y|A = 1]$, i.e., the expected value of the counterfactual outcome had we forced everyone to have treatment is equal to the expected value of the observed outcomes among those who actually received treatment (i.e., conditional or given individuals for whom $A = 1$).

2.1.2 Exchangeability

Groups are exchangeable if their counterfactual outcomes observed under any treatment level $A = a$ is independent of the actual treatment level that group received in our observed data, $Y^a \perp\!\!\!\perp A$. Often, we have conditional exchangeability, which means that groups are only exchangeable within certain subsets of the population, $Y^a \perp\!\!\!\perp A|C$. We can think about this using our example. In Figure 2, we see that we have conditional exchangeability such that within levels of C, the observed value of treatment is independent of the counterfactual outcome. If we don't condition on C, then we would not have exchangeability, since there would be a backdoor path from Y through C to A.

2.1.3 Positivity

Lastly, **positivity** is needed in order to quantify the causal effect using observational data since we need data in which some individuals received treatment and some who did not receive treatment for every level of of our confounder. Another way to think about this is that we cannot have any strata defined by the exposure and confounder in which there are no individuals.

Now that we have reviewed our causal assumptions, we can identify our causal effect. The goal of this process is to re-write our counterfactual statement in terms of observed data. We will not only be using the above principles but also properties of expectations and the law of total probability to identify the causal effect:

Average Causal Effect: $E[Y^{a=1} - Y^{a=0}]$

2.2 Identification Steps

2.2.1 Using properties of expectations

We can re-write our average causal effect as the difference between two expectations.

$$\begin{aligned} &E[Y^{a=1} - Y^{a=0}] \\ =&E[Y^{a=1}] - E[Y^{a=0}] \end{aligned}$$

2.2.2 Using law of total probability

Since we only have conditional exchangeability, $Y^a \perp\!\!\!\perp A|C$, we need to bring C into the conditioning statement. C is also not independent of Y^a so we cannot just add C. We must marginalize over C, meaning that we will be summing the $E[Y^a]$ over all values of C.

$$\begin{aligned} &E[Y^{a=1}] - E[Y^{a=0}] \\ =&\sum_c E[Y^{a=1}|C]Pr(C=c) - \sum_c E[Y^{a=0}|C]Pr(C=c) \end{aligned}$$

Law of Total Probability
The marginal probability of a variable, X, is equal to the joint probability of that variable, X, and a second, Z, summed over the second variable Z.

$$Pr(X) = \sum_z Pr(X,Z) = \sum_z Pr(X|Z)Pr(Z)$$

2.2.3 Using exchangeability

Now that we have C in the conditioning statement, we can bring in A, since A is independent of our counterfactual outcome Y^a within levels of C. This is because of *conditional exchangeability*.

$$\begin{aligned} &\sum_c E[Y^{a=1}|C]Pr(C=c) - \sum_c E[Y^{a=0}|C]Pr(C=c) \\ =&\sum_c E[Y^{a=1}|A,C]Pr(C=c) - \sum_c E[Y^{a=0}|A,C]Pr(C=c) \end{aligned}$$

Exchangeability and Independence
Independence and exchangeability are used interchangeably to describe the situations in which the flow of probability is blocked.

Two variables are independent when the marginal probability is equal to the conditional probability, i.e., if $X \perp\!\!\!\perp W$, then $Pr(X) = Pr(X|W)$.

If $Y^a \perp\!\!\!\perp A|C$, then we have $Pr[Y^a|C] = Pr[Y^a|A,C]$.

2.2.4 Using consistency

On the left, when we are identifying the outcome under treatment, we can set $A=1$, and on the right, when we are identifying the effect no treatment, we are identifying the outcome under no treatment, we can set $A=0$, then use consistency to remove the counterfactuals from Y.

$$\begin{aligned} &\sum_c E[Y^{a=1}|A=1,C=c] - \sum_c E[Y^{a=0}|A=0,C=c] \\ =&\sum_c E[Y|A=1,C=c]Pr(C=c) - \sum_c E[Y|A=0,C=c]Pr(C=c) \end{aligned}$$

2.3 Counterfactuals as observed data

We have successfully identified our causal effect of interest expressed using observed data! Remember that this is the **marginal average treatment effect (ATE)**.

$E[Y^{a=1} - Y^{a=0}] = \sum_c E[Y|A=1,C]Pr(C=c) - \sum_c E[Y|A=0,C]Pr(C=c)$

In the next section, we explore how to use regression to estimate components of this equation.

As an aside, consider if our causal effect of interest was the conditional average treatment effect (CATE), conditional on C. In counterfactual notation this would be $E[Y^{a=1} - Y^{a=0}|C=c]$. Since C is already in the statement, we would not to use the law of total probability from 2.2.2, and thus would not have to marginalize over C. Our causal effect of interest could be expressed in terms of observed data as: $E[Y^{a=1} - Y^{a=0}|C=c] = E[Y|A=1,C=c] - E[Y|A=0,C=c]$.

3 Connection to Regression

We now have our causal effect identified! So how can we use our observed data to estimate it?

We can apply properties of summation to re-write our causal effect in a way that may look a little more familiar.

$$E[Y^{a=1} - Y^{a=0}] = \sum_c E[Y|A=1,C]Pr(C=c) - \sum_c E[Y|A=0,C]Pr(C=c)$$

$$= \sum_c (E[Y|A=1,C] - E[Y|A=0,C])Pr(C=c)$$

$$= [(E[Y|A=1,C=0] - E[Y|A=0,C=0]) * Pr(C=0)] + [(E[Y|A=1,C=1] - E[Y|A=0,C=1]) * Pr(C=1)]$$

$(E[Y|A=1,C=0] - E[Y|A=0,C=0])$ is the conditional average treatment effect for males (i.e., $C=0$)

$(E[Y|A=1,C=1] - E[Y|A=0,C=1])$ is the conditional average treatment effect for females (i.e., $C=1$)

We can estimate these values using a linear regression model!

3.1 Linear regression assuming no effect modification

Consider a linear regression model to assess the effect of statin treatment on blood cholesterol level, where we account for the confounder, biological sex. This model makes the assumption of no effect modification, since we have **not** included an interaction term between statin treatment and biological sex, $A * C$.

$$E[Y|A, C] = \beta_0 + \beta_1 A + \beta_2 C$$

Conditioning on C in the model is necessary given the steps above that allowed us to use the exchangeability $Y^a \perp\!\!\!\perp A|C$ to identify the average causal effect. Here, β_1 is the average causal effect, i.e., had everyone been treated with statins, the average difference in blood cholesterol is β_1 compared to had no one been treated with statins, holding sex constant. This is a **conditional causal effect** since we are holding C, biological sex, constant. As previously noted, this model makes the assumption that there is no effect modification; the effect of statins on blood cholesterol level does not differ across males and females. We can see this by identifying the conditional average treatment effect among males (CATE(0)) and among females (CATE(1)) separately:

CATE(0) Average Treatment Effect among Males: $E[Y|A = 1, C = 0] - E[Y|A = 0, C = 0]$

$E[Y|A = 1, C = 0] = \beta_0 + \beta_1(A = 1) + \beta_2(C = 0)$

$E[Y|A = 0, C = 0] = \beta_0 + \beta_1(A = 0) + \beta_2(C = 0)$

$E[Y|A = 1, C = 0] - E[Y|A = 0, C = 0] = [\cancel{\beta_0} + \beta_1(1) + \cancel{\beta_2(0)}] - [\cancel{\beta_0 + \beta_1(0) + \beta_2(0)}]$

$= \beta_1(1)$

CATE(1) Average Treatment Effect among Females: $E[Y|A = 1, C = 1] - E[Y|A = 0, C = 1]$

$E[Y|A = 1, C = 1] = \beta_0 + \beta_1(A = 1) + \beta_2(C = 1)$

$E[Y|A = 0, C = 1] = \beta_0 + \beta_1(A = 0) + \beta_2(C = 1)$

$E[Y|A = 1, C = 1] - E[Y|A = 0, C = 1] = [\cancel{\beta_0} + \beta_1(1) + \cancel{\beta_2(1)}] - [\cancel{\beta_0 + \beta_1(0) + \beta_2(1)}]$

$= \beta_1(1)$

3.1.1 Putting it all together

We can set hypothetical values for each quantity and see how the CATE relates to the marginal ATE.

Let $\beta_1 = 2$, $Pr(C = 0) = 0.35$, and $Pr(C = 1) = 0.65$

$E[Y^{a=1} - Y^{a=0}] = [(E[Y|A = 1, C = 0] - E[Y|A = 0, C = 0]) * Pr(C = 0)]$

$$+[(E[Y|A = 1, C = 1] - E[Y|A = 0, C = 1]) * Pr(C = 1)]$$

$$= [\beta_1 * Pr(C = 0)] + [\beta_1 * Pr(C = 1)]$$

$$= [2 * 0.35] + [2 * 0.65]$$

$$= 2$$

When there is no effect modification, the CATE = ATE!

3.2 Linear regression with effect modification

Consider a linear regression model to assess the effect of statin treatment on blood cholesterol level, where we account for the confounder, biological sex, and effect modification. The β_3 term allows for the effect of statins on blood cholesterol level to differ among males and females. If there were truly no effect modification, then $\beta_3 = 0$ and 'drops out' making our equation look like the equation in Section 3.1.

$$E[Y|A, C] = \beta_0 + \beta_1 A + \beta_2 C + \beta_3 (A * C)$$

In this section, for the purpose of this exercise, we will assume that there is in fact effect modification and then $\beta_3 \neq 0$. We will see that β_3 will contribute to the causal effect by identifying the conditional average treatment effect among males and females separately.

CATE(0) Average Treatment Effect among Males: $E[Y|A = 1, C = 0] - E[Y|A = 0, C = 0]$

$E[Y|A = 1, C = 0] = \beta_0 + \beta_1(A = 1) + \beta_2(C = 0) + \beta_3(A = 1)(C = 0)$

$E[Y|A = 0, C = 0] = \beta_0 + \beta_1(A = 0) + \beta_2(C = 0) + \beta_3(A = 0)(C = 0)$

$E[Y|A = 1, C = 0] - E[Y|A = 0, C = 0] = [\cancel{\beta_0} + \beta_1(1) + \cancel{\beta_2(0) + \beta_3(1)(0)}] - [\cancel{\beta_0 + \beta_1(0) + \beta_2(0) + \beta_3(0)(0)}]$

$= \beta_1(1)$

CATE(1) Average Treatment Effect among Females: $E[Y|A = 1, C = 1] - E[Y|A = 0, C = 1]$

$E[Y|A = 1, C = 1] = \beta_0 + \beta_1(A = 1) + \beta_2(C = 1) + \beta_3(A = 1)(C = 1)$

$E[Y|A = 0, C = 1] = \beta_0 + \beta_1(A = 0) + \beta_2(C = 1) + \beta_3(A = 1)(C = 0)$

$E[Y|A = 1, C = 1] - E[Y|A = 0, C = 1] = [\cancel{\beta_0} + \beta_1(1) + \cancel{\beta_2(1)} + \beta_3(1)(1)] - [\cancel{\beta_0 + \beta_1(0) + \beta_2(1) + \beta_3(0)(1)}]$

$= \beta_1(1) + \beta_3(1)(1)$

Note the difference from above - previously, the average causal effect for males and females were both identified by β_1. Now, only the average causal effect among males ($C = 0$) is identified by β_1. Among females, the effect is identified by $\beta_1 + \beta_3$. Again, since there is effect modification, ($\beta_3 \neq 0$), we find that the average causal effects among males and females differ, i..e., the effect of treatment on cholesterol differs for males and for females.

3.2.1 Putting it all together

We can set hypothetical values for each quantity and see how the marginal ATE differs from the CATE in the presence of effect modification.

Let $\beta_1 = 2$, $\beta_3 = 4$, $Pr(C = 0) = 0.35$, and $Pr(C = 1) = 0.65$

$E[Y^{a=1} - Y^{a=0}] = [(E[Y|A = 1, C = 0] - E[Y|A = 0, C = 0]) * Pr(C = 0)]$

$$+[(E[Y|A = 1, C = 1] - E[Y|A = 0, C = 1]) * Pr(C = 1)]$$

$$= [\beta_1 * Pr(C = 0)] + [(\beta_1 + \beta_3) * Pr(C = 1)]$$

$$= [2 * 0.35] + [(2 + 4) * 0.65]$$

$$= 4.6$$

When there is effect modification, the CATE $\neq$ ATE!