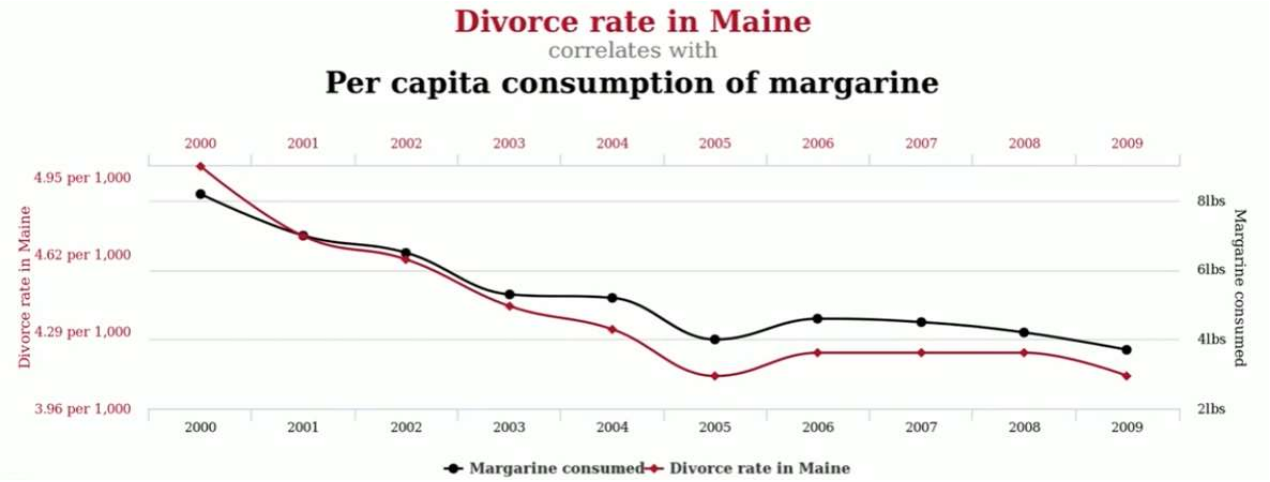
Module 1

Wednesday, October 2, 2024 2:03 PM

Confusion over Causality

Spurious Correlation: Causally unrelated variables might happen to be highly correlated with each other over some period of time.

- People may attribute causality to stuff that is actually just an anecdote and there isn't actually any proof that there is any causality.
- Something may be reported in such a way that they don't use any form of the word "cause", but the information is interpreted as causal.



Reverse Causality: Even if there is a causal relationship, sometimes the direction is unclear.

- Are active people more likely to prioritize near green space or does green space cause people to exercise more?

The field of causal inference or causal modeling attempts to clear up this causality confusion by proposing:

- Formal definitions of causal effects
- Assumptions necessary to identify causal effects from data
- Rules about what variables need to be controlled for
- Sensitivity analyses to determine the impact of violations of assumptions on conclusions

Potential Outcomes and Counterfactuals

Treatments and Outcomes: Suppose we are interested in the causal effect of some exposure/treatment A on some observed outcome Y, then an example of treatments and outcomes could be:

- $A = \begin{cases} 1 = \ received\ active\ drug \\ 0 = \ received\ placebo \end{cases}, Y = \begin{cases} 1 = developed\ disease\ within\ 2\ years \\ 0 = otherwise \end{cases}$

Potential Outcomes: The outcomes we would see under each possible treatment option.

- Y^a is the observed outcome if the treatment was set to $A = a$
- The potential outcomes are Y^1, Y^0

Counterfactuals: The outcomes that would have been observed had the treatment been different.

- If treatment was $A = 1$, then the counterfactual outcome is Y^0

Note:

- Before the treatment decision is made, any outcome is a potential outcome.
- After the study is done, there is an observed outcome ($Y = Y^A$) and a counterfactual outcome ($Y = Y^{1-A}$) .
- Counterfactual outcomes are typically assumed to be the same as potential outcomes, and so the terms are often used interchangeably.

Suppose that we are observing the effects of the flu vaccine. We want to know if receiving the vaccine prevented a patient from getting the flu.

What actually happened:

- The patient got the vaccine and did not get sick
  - Actual Treatment: $A = 1$
  - Observed Outcome: $Y = Y^1$

What would have happened (contrary to fact):

- If the patient didn't get the vaccine, would they have gotten sick?
  - Counterfactual Treatment: $A = 0$
  - Counterfactual Outcome: $Y = Y^0$

Hypothetical Interventions

It is cleanest to think of causal effects of interventions/actions, which are causal effects of variables that can be manipulated. Causal effects of hypothetical interventions are generally well defined, such as the outcome if prescribed a drug versus the outcome if prescribed a placebo.

One Version of Treatment: it is common to assume that there are no hidden versions of treatment. Say we are interested in the causal effect of BMI on health outcomes; then there is a problem because there are many potential ways that one could achieve a certain BMI and these different ways might be associated with different outcomes. Thus there isn't one version of treatment. Weight is not directly manipulable, so it's better to think of causal effects of interventions that aim at manipulating weight instead.

Immutable Variables: It is less clear what a causal effect of an immutable variable would mean, things that cannot be hypothetically changed like race, gender, or age. When we consider a potential outcome, we imagine that we could hypothetically set the treatment and then see the outcome. Instead we consider things that are manipulable:

No direct intervention	Manipulable (intervention)
Race	Name on resume
Obesity	Bariatric surgery
Socioeconomic status	Gift of money

Causal Effects: treatments that we can imagine being randomized (manipulated) in a hypothetical trial. We focus on causal effects of hypothetical interventions because their meaning is well defined and potentially actionable. In general, we say that *A* had a causal effect on *Y* if Y^1 is different from Y^0.

- There is only a causal effect if $Y^1 \neq Y^0$

The Fundamental Problem of Causal Inference is that we can only observed one potential outcome for each person. However, with certain assumptions, we can estimate the population level (average) causal effect.

Suppose we want to test the causal effect of medicine on clearing a headache.

- $Y = \begin{cases} 1, & \text{headache is gone} \\ 0, & \text{headache is not gone} \end{cases}$
- $A = \begin{cases} 1, & \text{took medicine} \\ 0, & \text{did not take medicine} \end{cases}$

"I took medicine and my headache is gone, therefore the medicine worked" → this is not proper causal reasoning since it's just an anecdotal statement. The statement is equivalent to $Y^1 = 1$, but we don't know what the counter factual is: $Y^0 = ?$ Perhaps the headache would have gone away without the medicine, too.

We can't know what would have happened if I hadn't taken the medication (unit level causal inference), but we can calculate the rate of headache remission if everyone took ibuprofen when they had a headache versus if no one did.

Causal Effects

Consider some hypothetical Population of Interest, which is the set of all people who are the entire population that you're interested. For example, we are researching diabetes treatments, then this population would be the set of all diabetics.

Now we have two hypothetical worlds, where World 1 is everyone in the Population of Interest that gets Treatment $A = 0$ and World 2 is everyone in the Population of Interest that gets Treatment $A = 1$. Both World 1 and World 2 contain 100% of the Population of Interest.

We can calculate the average of each world and take the difference to get the Average Causal Effect: $E(Y^1 - Y^0)$

- This is the average value of the Outcome *Y* if everyone was treated with $A = 1$ minus the average value of Y if everyone was treated with $A = 0$.
- If the Outcome is binary, this is a Risk Difference.

Say we want to calculate the Average Causal Effect of regional ($A = 1$) versus general ($A = 0$) anesthesia for hip fracture surgery on risk of major pulmonary complications.

Suppose that $E(Y^1 - Y^0) = -0.1$.

This means that the probability of major pulmonary complications is lower by 1% if given regional anesthesia compared with general anesthesia. Thus if 1000 people had this surgery, we would expect about 100 fewer people to have pulmonary complications with regional anesthesia compared to general anesthesia.

Say we want to calculate the Average Causal Effect of thiazide diuretic ($A = 1$) versus no treatment ($A = 0$) among hypertensive patient, where the outcome Y is systolic blood pressure.

Suppose that $E(Y^1 - Y^0) = -20\ mm\ Hg$

This means that if the population of hypertensive patients took thiazide diuretics, their average systolic blood pressure would be 20 mm Hg lower than if they didn't take the medication.

Conditioning on Treatment

In general: $E(Y^1 - Y^0) \neq E(Y|A = 1) - E(Y|A = 0)$

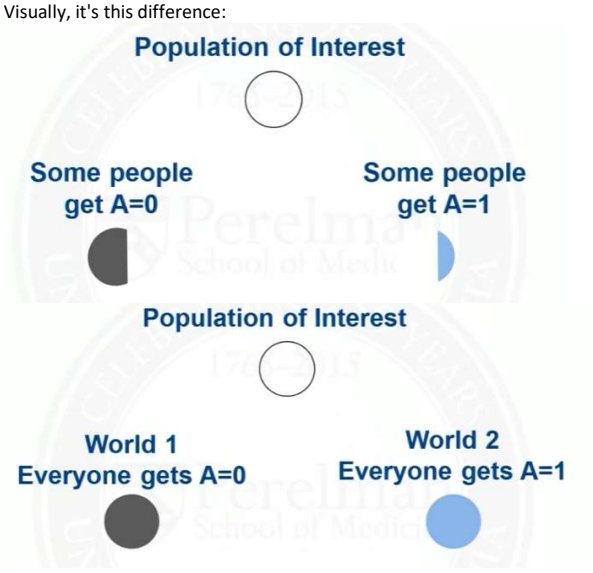
The conditional expectation restricts to the subpopulation of people who actually had a specific treatment, so the average isn't taken from the same population as the non-conditional expectation. Thus it isn't a causal effect as it is comparing two separate populations of people.

In layman's terms:

- $E(Y^1 - Y^0)$: the average value of the outcome if the entire population of interest has Treatment 1 minus the average value of the outcome if the entire population has Treatment 0.
- $E(Y|A = 1)$: of those who had Treatment 1, what is the average outcome?
- $E(Y|A = 0)$: of those who had Treatment 0, what is the average outcome?

Taking the average across the entire population of interest versus just a subpopulation is the key difference here. The people in the subpopulation may differ from the entire population in key ways. For example, people who have a higher risk for flu may be more likely to get the flu vaccine.

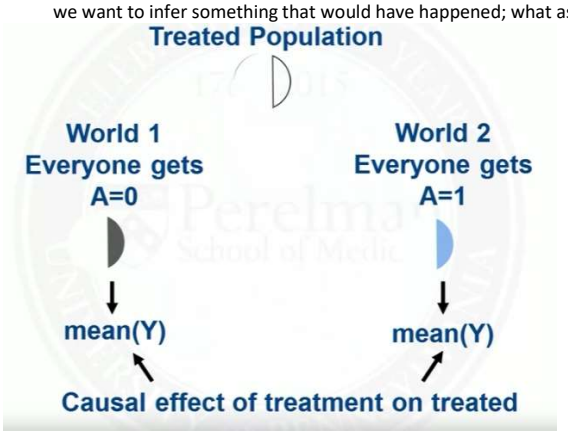
Visually, it's this difference:



Causal Relative Risk: $E\left(\frac{Y^1}{Y^0}\right)$

Causal effect of treatment on the treated: $E(Y^1 - Y^0|A = 1)$

- How well does the treatment work among those who are treated?
- In both worlds, we restrict to the same group of people. In real life, we have the challenge that we can only observe one treatment and one outcome for each person. Instead, we want to infer something that would have happened; what assumptions are necessary to estimate causal effects from observed data?



Average causal effect in the subpopulation with Covariate V=v: $E(Y^1 - Y^0|V = v)$

Practice Quiz

1. What is the average outcome if no one in the population had been treated?
   a. $E(Y^0)$
2. We can only observe one potential outcome for each person is
   a. The Fundamental Problem of Causal Inference

Causal Assumptions

Identifiability of causal effects requires making some untestable assumptions, generally called Causal Assumptions:

1. Stable Unit Treatment Value Assumption (SUTVA)

2. Consistency
3. Ignorability
4. Positivity

These assumptions are about the observed data (Y, A) and a set of pretreatment covariates (X).

SUTVA: a statistical assumption that ensures a unit's response is only dependent on the treatment it was assigned and not the treatment of other units.

1. No interference: units (members of the population) do not interfere with each other; treatment assignment of one unit does not affect the outcome of another unit. Interference is also known as spillover or contagion.
2. One version of treatment: there are no different forms or versions of each treatment level that lead to different potential outcomes

SUTVA allows us to write the potential outcome for the *ith* person in terms of only that person's treatment.

Consistency: the potential outcome under treatment $A = a$, Y^a, is equal to the observed outcome if the actual treatment received is $A = a$.

Ignorability: given pre-treatment covariates X, treatment assignment is independent from potential outcomes. Among people with the same values of X, we can think of treatment A as being randomly assigned. This is sometimes referred to as the No Unmeasured Confounders assumption.

$Y^0, Y^1 \perp A|X$

As an example:

- X is a single variable (age) that can take values "older" or "younger".
- Older people are more likely to get treatment A=1.
- Older people are more likely to have outcome (hip fracture), regardless of treatment.

Age is therefore related to the treatment and the risk of the outcome, so treatment isn't randomly assigned since older people are more likely to get treatment A=1. Thus outcomes Y^0, Y^1 are not independent from treatment A. However, within levels of X, treatment might be randomly assigned; among people who are younger, treatment is random; among people who are older, treatment is random. Thus treatment assignment is ignorable given age.

Positivity: for every set of values for covariates X, treatment assignment was not deterministic; at every level of X and for every treatment, people had a nonzero probability of getting treatment.

$P(A = a|X = x) > 0$ for all a and x

These assumptions can be put together to identify causal effects. Starting with only observed data:

$$E(Y|A = a, X = x)$$
$$= E(Y^a|A = a, X = x) \leftarrow \text{by consistency}$$
$$= E(Y^a|X = x) \leftarrow \text{by ignorability}$$

Stratification

Previously, we needed to condition on X to be able to link the observed outcome to the potential outcome; recall that under certain causal assumptions: $E(Y|A = a, X = x) = E(Y^a|X = x)$. But if we want to get rid of the X and get the marginal causal effect, we can average over the distribution of X.

Let's say that we have a single categorical X variable, then we can calculate the expected value of the observed outcome if the treatment A=a as the summation of the Expected value of the outcome given the treatment A=a and covariate X=x multiplied by the probability that the covariate equals x.

First recall that the Expected Value is calculated as the sum of all possible values multiplied by the probability of its occurrence: $E(Z) = \sum_i z_i P(Z = z)$

Since we have that $E(Y|A = a, X = x) = E(Y^a|X = x)$, then we can say that the expected value of the observed outcome Y^a is the sum of all possible outcomes multiplied by their probability of occurrence:

$$E(Y^a) = \sum_x E(Y|A = a, X = x)P(X = x)$$

This is a standardized mean, which happens to be the same as the average potential outcome.

This is a process known as Standardization, which is just conditioning (stratifying) and marginalizing (averaging) over the distribution of X. In other words, it is the process of obtaining a treatment effect within each stratum and then pooling across the stratum, weighting by the probability (size) of each stratum. From the data, you could estimate a treatment effect by computing means under each treatment within each stratum, and then pooling across the stratum.

Also recall the Law of Total Probability, which states that for disjoint events W such that their probabilities sum up to 1, then the probability of event Z is: $P(Z) = \sum_{i=1}^{n} P(Z \cap W_i)$

Say we are doing a study comparing two diabetes treatments:

- Treatment A: new initiators of saxagliptin versus sitagliptin
- Outcome Y: Major Adverse Cardiac Event (MACE).
- Challenges:
  - saxagliptin users were more likely to have had past use of some other oral antidiabetic (OAD) drug
  - patients with past use of OAD drugs are at higher risk for MACE

The steps for standardizing this problem are as follows:

1. compute the rate of MACE for saxagliptin and sitagliptin initiators in two subpopulations:
   a. Patients who have had no prior OAD use
   b. Patients who have had prior OAD use
2. Take the weighted average, where weights are based on the proportion of people in each population

This is the causal effect if, within levels of prior OAD use, treatment can be thought of as randomized (by ignorability given prior OAD use).

	MACE = yes	MACE = no	Total
Saxa=yes	350	3650	4000
Saxa=no	500	6500	7000
Total	750	10250	11000

$$P(MACE|saxagliptin = yes) = \frac{350}{4000} = .0875$$
$$P(MACE|saxagliptin = no) = \frac{500}{7000} = .0714$$

This is the raw data of 8.75% of saxagliptin users having MACE and 7.14% of sitagliptin users having MACE, it appears that saxagliptin users suffer more from MACE. But we don't know if that difference is caused by previous OAD use / being assigned to "worse off" people. So now we stratify the data by prior OAD use:

OAD USE = no	MACE = yes	MACE = no	Total
Saxa=yes	50	950	1000
Saxa=no	200	3800	4000
Total	250	4750	5000

OAD USE = yes	MACE = yes	MACE = no	Total
Saxa=yes	300	2700	3000
Saxa=no	300	2700	3000
Total	600	5400	6000

We can see that the majority of saxagliptin users had prior OAD use and those people are at higher risk for MACE.

$$P(MACE|saxagliptin = yes \cap OAD\ use = no) = \frac{50}{1000} = .05$$
$$P(MACE|saxagliptin = no \cap OAD\ use = no) = \frac{200}{4000} = .05$$

$$P(MACE|saxagliptin = yes \cap OAD\ use = yes) = \frac{300}{3000} = .10$$
$$P(MACE|saxagliptin = no \cap OAD\ use = yes) = \frac{300}{3000} = .10$$

Thus among users with prior OAD use, the risk of MACE is 10% regardless of treatment. There appears to be no difference in treatment effectiveness between saxagliptin and sitagliptin, whereas when we looked at the raw data before it appeared that saxagliptin was less effective.

Next we want to calculate the means of the potential outcomes:

Probability of MACE if everyone was given saxagliptin:

$$E(Y^{saxa}) = E(Y|A = saxa \cap OAD = yes)P(OAD = yes) + E(Y|A = saxa \cap OAD = no)P(OAD = no) = \frac{300}{3000} * \frac{6000}{11000} + \frac{50}{1000} * \frac{5000}{11000} = 0.077$$

Probability of MACE if everyone was given sitagliptin:

$$E(Y^{sita}) = E(Y|A = sita \cap OAD = yes)P(OAD = yes) + E(Y|A = sita \cap OAD = no)P(OAD = no) = \frac{300}{3000} * \frac{6000}{11000} + \frac{200}{4000} * \frac{5000}{11000} = 0.077$$

So once we marginalized, we find that the mean for both treatments is exactly the same, so the potential outcome is the same if you gave everyone saxagliptin versus giving everyone sitagliptin. That is an effective causal effect.

Problems with standardization:

- Typically, there will be many X variables needed to achieve ignorability and problems won't be simple.
- Stratification would lead to many empty cells (data that is missing) because there will be many combinations of factors for which we just don't have the data

Thus we need alternatives to standardization and ways to estimate causal effects:

- Matching
- Inverse probability of treatment weighting
- Propensity score methods

Practice Quiz

1. Which assumption is also referred to as the "no unmeasured confounders assumption"?
   a. Ignorability
2. Which causal assumption requires no interference between units?
   a. SUTVA

Incident User and Active Comparator Designs

Let's consider a cross-sectional look at treatments:

Suppose we were interested in whether yoga affects blood pressure. At any time, some people regularly practice yoga while others do not:

- Those who do not practice may have in the past
- Those who do might have been practicing for a long time or be beginners
- Why did some people stop while others continued? What if those that quit did so because it wasn't working for them?

This is a type of selection bias that is difficult to control for.

One way to handle this is to use Incident User Design, which restricts the treated population to those who are newly initiating treatment. This is also known as New User Design and we want to know "what is the causal effect of starting the treatment for the first time". One challenge with this approach is if the comparison group is "no treatment", it isn't obvious when follow-up should start for the "no treatment" group. Having an active comparator makes this easier; i.e., instead of comparing yoga benefits to "no treatment" benefits, compare yoga to some other kind of fitness activity like zumba. The downside of this approach is that it makes the causal question more narrow.

Other considerations:

- It isn't always possible to implement an incident user design, like the causal effect of air pollution
- Sometimes no treatment is the comparison group of interest
- Causal methods exist that can handle time-varying treatments; the causal effects of treatment regimens over time

Quiz

1. The Fundamental Problem of Causal Inference is that:
    a. We can only observed one potential outcome for each subject
2. What represents the causal effect of treatment on the treated?
    a. $E(Y^1 - Y^0|A = 1) = E(Y^1|A = 1) - E(Y^0|A = 1)$
3. What represents the average causal effect for the population?
    a. $E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$
4. Which causal assumption would be violated if the effectiveness of treatment on an individual depended on the treatment status of other individuals?
    a. SUTVA
5. Which causal assumption would be violated if we were interested in the causal effect of treatment for people age 40-80, but everyone over age 70 received the treatment?
    a. Positivity
6. If the consistency assumption holds, then the observed outcome for a treated subject is equal to the potential outcome under the treatment:
    a. True
7. Which of the following can most easily be thought of as an intervention?
    a. Changing medication
8. Treatment assignment being ignorable given confounders X means:
    a. Within levels of X, treatment assignment is independent from the potential outcomes
9. Computing means within levels of covariates and then combining these estimates is known as:
    a. Standardization