Lecture 9: Regression and Causality

D. Alex Hughes

December 2, 2014

Questions on Project

1. How are they going?
2. Have you found data?
3. What are the implications of today's lecture for your project?

Potential Outcomes and Groundhog Day

Groundhog Day

- We see Ned come through several times.
- This gives Bill Murray an opportunity to potentially do different things
- He can do what comes naturally; ignore him; push him; punch him; hug him
- Each of these let him observe differences in reactions in Ned that are otherwise at the SAME point in Ned's life

Statement of Angriest and Pischeke's Goals

1. What is the causal relationship of interest?
2. What is the experiment that could potentially capture this causal relationship?
3. Given limitations of observational data - if you should choose to use it - what strategy will you use to *identify* the causal structure that you're interested in?

What regression gives us

- Covariance relationships between stacks of variables;
- We designate what the RHS variable and what is the LHS variable
  - This is designation is definitionally arbitrary; something we have run up against time and again with education income and schooling;
  - comes up with even greater regularity in the designation of political outcomes
  - Polarization in the American Congress did polarization happen as a result of (a) the southern realignment; (b) movement on the part of elites? The two are moving at the same time; and under a covariance examination there is no way that we can pull one apart
  - Does economic liberalization lead to economic and political development?
  - What is the lingering effect, if any, of the colonial-world?

What regression doesn't give us

- A statement about what is "causing" and what is "effecting"
- Or, even more broadly, a statement about the "causal structure" that exists in the data-generating-process (see e.g. Judea Pearl, *Causality*)

We will spend today carefully defining notions of causal structure and how we can then use regression as one potential mechanism to measure this causal-effect

What regression doesn't give us

- A statement about what is "causing" and what is "effecting"
- Or, even more broadly, a statement about the "causal structure" that exists in the data-generating-process (see e.g. Judea Pearl, *Causality*)

We will spend today carefully defining notions of causal structure and how we can then use regression as one potential mechanism to measure this causal-effect

Why we care about causal quantities

- Maybe this is in transition
- Possible that you fall into the purely descriptive camp (you want just to describe the world and the relationships that exist among its many actors)
- Don't care about testing what is causing what

However

- In these (possibly, aytheoretic) model fittings, the predictive researcher is implicitly assuming that whatever the mechanism that is "causing" the outcome, it will be constant over time
- See e.g. email Seth sent out to methods list with the chapter by Thad Dunning)

These contrasts are at the forefront of an *active* debate in the discipline.

The Ideal Experiment?

AP example: Early-entry to schooling

- What is the effect of delayed entry into school for young children?
- DK exists; so too does the option to hold your child out for an extra year.

Empirical Strategy

- Might compare: children who start K at age 6 and other who start at age 7
- Make it concrete: Measure performance in standardized tests

The Ideal Experiment?

AP example: Early-entry to schooling

- What is the effect of delayed entry into school for young children?
- DK exists; so too does the option to hold your child out for an extra year.

Empirical Strategy

- Might compare: children who start K at age 6 and other who start at age 7
- Make it concrete: Measure performance in standardized tests

Immediate Concerns?

- There are lots of ways that children who start at 6 might be different than 7
- Children's parents who start young might be working;
- Less time reading at home
- School districts that let delay are more progressive and have larger budgets... etc

Sure... But, suppose we can get around that and "randomly assign" students to start at 6 or 7

- Won't get around kids being different.
- One set is older than the other when they are taking the tests
- Older kids test better leads them to test better
- Clear, classic *maturation* effect

Immediate Concerns?

- There are lots of ways that children who start at 6 might be different than 7
- Children's parents who start young might be working;
- Less time reading at home
- School districts that let delay are more progressive and have larger budgets... etc

Sure... But, suppose we can get around that and "randomly assign" students to start at 6 or 7

- Won't get around kids being different.
- One set is older than the other when they are taking the tests
- Older kids test better leads them to test better
- Clear, classic *maturation* effect

Immediate Concerns?

- There are lots of ways that children who start at 6 might be different than 7
- Children's parents who start young might be working;
- Less time reading at home
- School districts that let delay are more progressive and have larger budgets... etc

Sure... But, suppose we can get around that and "randomly assign" students to start at 6 or 7

- Won't get around kids being different.
- One set is older than the other when they are taking the tests
- Older kids test better leads them to test better
- Clear, classic *maturation* effect

Immediate Concerns?

- There are lots of ways that children who start at 6 might be different than 7
- Children's parents who start young might be working;
- Less time reading at home
- School districts that let delay are more progressive and have larger budgets... etc

Sure... But, suppose we can get around that and "randomly assign" students to start at 6 or 7

- Won't get around kids being different.
- One set is older than the other when they are taking the tests
- Older kids test better leads them to test better
- Clear, classic *maturation* effect

Immediate Concerns?

- There are lots of ways that children who start at 6 might be different than 7
- Children's parents who start young might be working;
- Less time reading at home
- School districts that let delay are more progressive and have larger budgets... etc

Sure... But, suppose we can get around that and "randomly assign" students to start at 6 or 7

- Won't get around kids being different.
- One set is older than the other when they are taking the tests
- Older kids test better leads them to test better
- Clear, classic *maturation* effect

Ideal Experiment?

How, then, could we get around this? Could we?

Another question:

- How would you set up an ideal experiment to learn about the effect of race on perception of a police-homicide?

Setup from Holland (1986)

A note: this little bit of notation isn't consistent with *MHE*, but it's pretty close.

A Model

- Start with a *population* U of "units." Then a particular unit $u \in U$ makes sense.
- Suppose for each u there is some Y value, $Y(u)$. We'll call this the *response variable*
- Let A be some other variable defined on U, call A an *attribute* of the units $u \in U$.
- Then A and Y are just things that are $\in U$.

Then, the most information someone can have in the model is the joint values of $Y(u)$ and $A(u)$, and their joint distribution:

$$Pr(Y = y, A = a)$$

which is the proportion of $u \in U$ where $Y(u) = y$ and $A(u) = a$.

Conditional Probability

- Then, if A and Y are not independent, when we know something about A we also know something more about Y than before.
- We know $P(Y = y|A = a)$, and can calculate the $E[Y|A = a]$.

Causes and Treatments

- Throughout, and in everyday life *cause* $\equiv$ *treatment*
- When we say A causes B, what we really mean is that A causes B, relative to something else, "*not* A"
- The key notion in an experiment then is the *potential* of exposing each $u \in U$ to a cause
- Let S be a variable (schooling?) that indicates a treatment which can take two values, c and t.

Causes and Treatment

"In a controlled study, S is constructed by the experimenter, and in an uncontrolled study, S is determined by some factors beyond the experimenter's control."

- Critically, $S(u)$ *could have been different*
- $S(u)$ and $A(u)$ are on similar footing

The response variable:

- Causes must exist post-exposure (time...)
- Values of these post-exposure variables can change as result of exposure: *causes have effects*

Causes and Treatment

- So, what we really need is not one Y, but rather two?
- Y_t: The value of Y if exposed to treatment; and,
- Y_c: The value of Y if exposed to control

And with this laid out, we can say:

- $Y_t(u)$ is the value of Y person u would have if exposed to t; and,
- $Y_c(u)$ is the value of Y person u would have if exposed to c
- **This is the same unit** u.

Causes and Treatment

- So, what we really need is not one Y, but rather two?
- Y_t: The value of Y if exposed to treatment; and,
- Y_c: The value of Y if exposed to control

And with this laid out, we can say:

- $Y_t(u)$ is the value of Y person u would have if exposed to t; and,
- $Y_c(u)$ is the value of Y person u would have if exposed to c
- This is the same unit u.

Causes and Treatment

- So, what we really need is not one Y, but rather two?
- Y_t: The value of Y if exposed to treatment; and,
- Y_c: The value of Y if exposed to control

And with this laid out, we can say:

- $Y_t(u)$ is the value of Y person u would have if exposed to t; and,
- $Y_c(u)$ is the value of Y person u would have if exposed to c
- **This is the same unit** u.

Causes and Treatment

Then, the effect of t on u as measured on Y and relative to cause c is the difference between $Y_t(u)$ and $Y_c(u)$:

$$Y_t(u) - Y_c(u)$$

But,

Theorem

It is impossible to observe the value of $Y_1(u)$ and $Y_0(u)$ on the same unit, and therefore, it is impossible to observe the effect of t on u.

Examples of Fundamental Problem

Example

Image the case of a 4^{th} grader: she might get the new math class; or she might get the old matrix-tables class. We'll never see both for the same 4^{th} grader.

Example

Imagine election monitoring at a polling station in electoral Kenya (Clark). Either a polling station gets a monitor, or it doesn't; we can observe fraud at a polling station that has monitors on one that doesn't have monitors.

In both of these, we will not be able to *observe* the effect of giving a treatment to the polling station.

Examples of Fundamental Problem

Example

Image the case of a 4^{th} grader: she might get the new math class; or she might get the old matrix-tables class. We'll never see both for the same 4^{th} grader.

Example

Imagine election monitoring at a polling station in electoral Kenya (Clark). Either a polling station gets a monitor, or it doesn't; we can observe fraud at a polling station that has monitors on one that doesn't have monitors.

In both of these, we will not be able to *observe* the effect of giving a treatment to the polling station.

Scientific Solutions to Fundamental Problem

Example

- Imagine that we give some patient with hypertension some sildenafil.
- Then we measure his blood-pressure.
- Then we wait for his body to process the sildenafil and measure his blood-pressure again.
- Then we re-apply the sildenafil treatment...

Here we can re-create situations that are similar - think Ned - but can measure treatment-response and abatement.

Scientific Solutions to Fundamental Problem

The following defines the scientific solution:

$$Y_{c,t0}(u) = Y_{c,t1}(u)$$

But, there is an implicit *homogeneity* assumption. Maybe true. Maybe not.

- The *statistical* solution is to rely on population averages
- The average causal effect T of t (relative to c) over U is the expected value of of the difference $Y_t(u) - Y_c(u)$ over the u's in U.

$$E[Y_t - Y_c] = T$$

- Overcome the fundamental problem of causal inference
- We can learn something by observing different units that can, indeed, be observed.

Scientific Solutions to Fundamental Problem

The following defines the scientific solution:

$$Y_{c,t0}(u) = Y_{c,t1}(u)$$

But, there is an implicit*homogeneity* assumption. Maybe true. Maybe not.

- The *statistical* solution is to rely on population averages
- The average causal effect T of t (relative to c) over U is the expected value of of the difference $Y_t(u) - Y_c(u)$ over the u's in U.

$$E[Y_t - Y_c] = T$$

- Overcome the fundamental problem of causal inference
- We can learn something by observing different units that can, indeed, be observed.

Potential outcomes

Here, we fix ourselves to the notatitation in *MHE*

Definition

Y_{1i} and Y_{0i} will refer to *potential outcomes* and Y_i will refer to *observed outcomes*

Let's work with the hospital example that AP set up for us.

- We're interested in the effect of hospitalization on someone's health status;
- Does going to the hospital cause someone to become healthier (she get's care) or to become less healthy (she is around people who are sick)

Potential Outcomes

Here are the greatest hits:

Group	Health	SE
Hospital	3.21	0.014
$\neg$ Hospital	3.93	0.003

- Are people who do and don't go to the hospital the *same amount* of {sick,healthy} before they go to the hospital? Probably not.
- Even after getting care, someone who has had a (successful) quad-bypass may not be as healthy as the other person who, um, hasn't.

Potential Outcomes

- Imagine (*rainbow sounds...*) that a single person could either go to the hospital, or not go to the hospital.
- Think of going to the hospital as a binary RV $D_i = \{0, 1\}$.

Then for any person, there are two *potential outcomes*

$$\text{Potential outcome} = \begin{cases} Y_{1i} & \text{if } D_i = 1 \\ Y_{0i} & \text{if } D_i = 0 \end{cases}$$

So, the *potential outcome* is the *potential* health status if he had gone (or not gone) to the hospital, regardless of whether he goes to the hospital.

Potential Outcomes

Then, as long as we're still dreaming (or doctors applying sildenafil),

- Can easily imagine the causal effect as being the difference between the potential outcome where you don't go to the hospital (Y_{0i})
- And the potential outcome where you do go to the hospital (Y_{1i}):

$$Y_{1i} - Y_{0i}$$

.

Of course, we don't get to see this, but we do, in fact - we get to see some of these.

- In a binary case, we get to see as many as half,
- In more complex cases, we see fewer of them.

Group		$\mathbf{Y_{1i}}$		$\mathbf{Y_{0i}}$
T = 1	Observable:	$Y_{1i} \vert T = 1$		
T = 0			Observable:	$Y_{0i} \vert T = 0$

Potential Outcomes

Then, as long as we're still dreaming (or doctors applying sildenafil),

- Can easily imagine the causal effect as being the difference between the potential outcome where you don't go to the hospital (Y_{0i})
- And the potential outcome where you do go to the hospital (Y_{1i}):

$$Y_{1i} - Y_{0i}$$

.

Of course, we don't get to see this, but we do, in fact - we get to see some of these.

- In a binary case, we get to see as many as half,
- In more complex cases, we see fewer of them.

Group		$\mathbf{Y_{1i}}$		$\mathbf{Y_{0i}}$
T = 1	Observable:	$Y_{1i} \vert T = 1$	Counterfactual:	$Y_{0i} \vert T = 1$
T = 0	Counterfactual:	$Y_{1i} \vert T = 0$	Observable:	$Y_{0i} \vert T = 0$

Potential Outcomes

Then, we might think of the observed outcome (Y_i) as a step-wise function of the potential outcomes.

$$
\begin{aligned}
Y_i &= \begin{cases} Y_{1i} & if D_i = 1 \\ Y_{0i} & if D_i = 0 \end{cases} \\
&= Y_{0i} + (Y_{1i} - Y_{0i})D_i \\
&= Y_{0i}(1 - D_i) + (Y_{1i} - Y_{0i})D_i
\end{aligned}
$$

Because we're in the *real world* we can only learn about the effects of going to the hospital by comparing averages across populations of people who *went* to the hospital and those who didn't go to the hospital.

Potential Outcomes

Then, we might think of the observed outcome (Y_i) as a step-wise function of the potential outcomes.

$$
\begin{aligned}
Y_i &= \begin{cases} Y_{1i} & if D_i = 1 \\ Y_{0i} & if D_i = 0 \end{cases} \\
&= Y_{0i} + (Y_{1i} - Y_{0i})D_i \\
&= Y_{0i}(1 - D_i) + (Y_{1i} - Y_{0i})D_i
\end{aligned}
$$

Because we're in the *real world* we can only learn about the effects of going to the hospital by comparing averages across populations of people who *went* to the hospital and those who didn't go to the hospital.

Naive Comparison

- What if we just compare the health statuses of people who ever went to the hospital against those who *n*ever went to the hospital?
- Then we're, in the *observed* world, and...
- Missing an important part to draw a causal inference

We observe the average observed difference in health, conditional on the treatment status:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

Naive Comparison

- What if we just compare the health statuses of people who ever went to the hospital against those who *n*ever went to the hospital?
- Then we're, in the *observed* world, and...
- Missing an important part to draw a causal inference

We observe the average observed difference in health, conditional on the treatment status:

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$$

Naive Comparison

- Say we are interested in the **ATT** - the effect of treatment on those who received treatment.
- This is the difference in potential outcomes for people who received treatment: $E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$.
- Read aloud this might be: "*Of people who received treatment, the ATT is the difference in potential outcomes of potentially receiving and potentially not receiving the treatment.*"
- Statement about effects within a specific group of people: people who received the treatment.

Naive Comparison

If we just have observed outcomes, what we see is

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] \\ &= \{\mathbf{E[Y_{1i}|D_i = 1]} - E[Y_{0i}|D_i = 1]\} \\ &\quad - \{E[Y_{0i}|D_i = 1] - \mathbf{E[Y_{0i}|D_i = 0]}\} \end{aligned}$$

And so,

- We're not seeing the internal $E[Y_{0i}|D_i = 1]$ because this is a counterfactual case.
- Including it makes it clear that we have two components:
  1. ATT; and,
  2. Selection into treatment

Random Assigment

- Random assignment solves this because it makes D_i independent of the potential outcomes.
  - People who get $D_i = 1$ have potential outcomes Y_{1i} & Y_{0i}
  - People who get $D_i = 0$ have potential outcomes Y_{1i} & Y_{0i}
- They're the same!
- This makes two critical statements true:

$$E[Y_{1i}|D_i = 1] = E[Y_{1i}|D_i = 0]$$
$$E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$$

- And, because of this, we can make the following substitution:

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= \mathbf{E[Y_{1i}|D_i = 1]} - \mathbf{E[Y_{0i}|D_i = 0]} \\ &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \end{aligned}$$

1. Questions on Project
2. Causality Basics
   - Motivation
   - The Ideal Experiment
   - Holland (1986)
3. Potential Outcomes
   - Introduction
   - Missing Data Problem
   - Now What?
4. Regression and Experiments
   - Regression and Experiemnts
5. Observational Data
   - CEF
   - Conditional Independence Assumption
   - Broader Treatments

currentsection]

Missing Data Problem

- Another way that we could think of this issue is as a missing data problem:
- For example, take the following example from Patrick Lam's slides.

i	D_i	Y_{1i}	Y_{0i}	$Y_{1i} - Y_{0i}$
1	0	3	5	2
2	1	2	5	3
3	1	5	4	-1
4	0	2	7	5
5	1	1	2	1

- We observe some of these, but we don't observe others of them.
- The concern, when we think about this observation as a missing data problem, is that the values that are missing are not just missing completely at random (MCAR)
- Rather that there might be some systematic difference.

Example

Given the data from the last slide: Let's be concrete with the quantities we want:

1. What is the ATT?
2. What is the ATE?

Ok, so you randomized treatment

Now what?

1. Are the subjects' characteristics "balanced" across the treatments?
2. Did the randomization *work*?

This isn't actually comparing potential outcomes

- Comparing that that PO could have been the same
- Sometimes shown with a balance table and F-tests for indicators: AP 2.2.1
- Other times show with MNP predicting treatment status on RHS (a la Green & Gerber)

Ok, so you're balanced

Regression and Experiments Suppose that the treatment effect is the same for everyone: $Y_{1i} - Y_{0i} = \tau$ Then, we can rewrite the causal statement from earlier in terms of this τ.

$$\begin{aligned} Y_i &= \alpha + \tau D_i + \epsilon_i \\ &= Y_{0i} + (Y_{1i} - Y_{0i})D_i + (Y_{0i} - E(Y_{0i})) \end{aligned}$$

And,

$$\begin{aligned} E[Y_i|D_i = 1] &= \alpha + \tau + E[\epsilon_i|D_i = 1] \\ E[Y_i|D_i = 0] &= \alpha + E[\epsilon_i|D_i = 0] \end{aligned}$$

Then, the treatment effect $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$*is*

$$(\alpha - \alpha) + (\tau(1) - \tau(0)) + (E[\epsilon_i|D_i = 1] - E[\epsilon_i|D_i = 0])$$
$$\tau + \textit{SelectionBias}$$

This selection bias is defined solely in terms of potential outcomes for Y_{0i} (recall, the relationship between the intercept and mean of ϵ_i and can be written as)

$$E[\epsilon_i|D_i = 1] - E[\epsilon_i|D_i = 0], \text{ or,}$$
$$(E[Y_{0i}|D_i = 1] - E[Y_{0i}]) - (E[Y_{0i}|D_i = 0] - E[Y_{0i}])$$
$$E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

Regression and Experiments

- This is the difference between (no-treatment) potential outcomes when one is (vs. isn't) treated
- If this is different, then there is selection into treatment,
- Which we will see as correlation between regression residual and treatment indicator
- People who went to hospital (likely) had potential outcomes that would have been worse in the potential they did not go to the hospital

What about including other regressors?

- If we have randomized treatment, solving the selection problem, why would we include any other regressors?
- Tennessee case include regressors about student income, teachers' race, teachers' experience
- Why?
- What should be the relationship between D_i and these variables?
- If D_i is assigned at random, then $Cov(D_i, Z_i) \equiv 0$
- By OVB formula, $\Delta\tau = 0$
- But, including these other RHS variables will decrease the MSE of the regression,
- Reducing σ^2_ϵ, and therefore, increasing precision in our estimate (lower SEs) of τ.
- Acknowledgement that τ is a distribution of effects, potentially conditional on other observable (or unobservable) characteristics

Why isn't everything a RCT?

1. Expensive
2. Time consuming
3. Logistically...challenging

But, if we're interested in causal statements, want to use regression, but are a graduate student, what do we do?

- We search for "as-if" randomization

Conditional Expectation Function

We started this quarter suggesting that regression was a useful tool to make statements about conditional expectations of variables. How far we've come!

Definition

The **conditional expectation function** (CEF) for one realization Y_i of Y is the expectation of Y (population average) at some X_i.

- Given some X's, what would we expect Y to be?
- What if the X_i's are *good* at helping us make a conditional statement about Y_i?
- What if they're... bad?

CEF

Classic statements from Math Camp:

- $Y = f(X)$
- $E[Y] = \int X * f(X)$
- $E[Y|X \in \{x_0, x_1\}] = \int_{x_0}^{x_1} x * f(X)$

Partitioning the CEF

Theorem

We can partition the CEF into parts explained by X and parts not explained by X (residuals).

$$Y_i = E[Y_i|X_i] + \epsilon_i$$

Benefits of Regression

For a number of reasons, regression does a good job estimating the CEF

1. If the CEF is linear, then regression is that CEF: $E[\hat{\beta}] = \beta$
2. Even if the CEF isn't linear, regression gives the best *linear* predictor of the CEF - $(y_i = \hat{y}_i)^2)$ is a small under regression as any possible linear form
3. Even if the CEF isn't linear, regression gives the best *linear* approximation of the CEF - very similar demonstration

Regression and Causality

- So, regression does well at approximating the CEF
- When can we think of regression as possessing a *causal* interpretation?
- Is it ever possible that we can use observational data and still get around the selection effect to make a causal claim?

Regression and Causality

When can we say a regression has a causal interpretation? Regression is just a mechanical process.

Theorem

A regression can be said to have a causal interpretation when the CEF it is estimating, itself, has a causal interpretation.

Theorem

Ok, so... a CEF has a causal interpretation when it describes differences in potential outcomes.

Regression and Causality

When can we say a regression has a causal interpretation? Regression is just a mechanical process.

Theorem

A regression can be said to have a causal interpretation when the CEF it is estimating, itself, has a causal interpretation.

Theorem

Ok, so... a CEF has a causal interpretation when it describes differences in potential outcomes.

Regression and Causality

When can we say a regression has a causal interpretation? Regression is just a mechanical process.

Theorem

A regression can be said to have a causal interpretation when the CEF it is estimating, itself, has a causal interpretation.

Theorem

Ok, so... a CEF has a causal interpretation when it describes differences in potential outcomes.

Regression and Causality

When can we say a regression has a causal interpretation? Regression is just a mechanical process.

Theorem

A regression can be said to have a causal interpretation when the CEF it is estimating, itself, has a causal interpretation.

Theorem

Ok, so... a CEF has a causal interpretation when it describes differences in potential outcomes.

Examples

Example

- Bill Murray & Ned in Groundhog day
- What potential friendship outcomes might we see in Ned conditional on Murray's behavior?
- Would we conditionally expect friendship after a hug? After a punch?

Examples

Example

- Take the classic example of income and schooling
- What potential income would a given (fixed) person earn under possibly different amounts of schooling

- The causal effect of schooling tells us what a person (or people) would earn on average if we could change their schooling in a perfectly controlled setting (scientifically, with assumption of homogeneity); or,
- Compare different groups of people with schooling amounts assigned randomly so that schooling is independent potential outcomes

Examples

Example

- Take the classic example of income and schooling
- What potential income would a given (fixed) person earn under possibly different amounts of schooling

- The causal effect of schooling tells us what a person (or people) would earn on average if we could change their schooling in a perfectly controlled setting (scientifically, with assumption of homogeneity); or,
- Compare different groups of people with schooling amounts assigned randomly so that schooling is independent potential outcomes

Conditional Independence Assumption

Definition

The **Conditional Independence Assumption** asserts that, conditional on observable characteristics (X_i), there is no selection into treatment. Formally,

$$\{Y_{0i}, Y_{1i}\} \perp D_I | X_i$$

This is also sometimes called "*selection on observables*"

Examples

CIA and the AP' Schooling Example **Recall**: Income, Schooling (college)

- We anticipate that potential outcomes are not independent of schooling decision
- Are there some X_i, that if included, would estimate the selection into college?
- Imagine the minimal case of (a) ability; and (b) family background.
- We can get a long ways down the road using potential outcomes framework.

Examples

CIA and the AP' Schooling Example **Recall**: Income, Schooling (college)

- We anticipate that potential outcomes are not independent of schooling decision
- Are there some X_i, that if included, would estimate the selection into college?
- Imagine the minimal case of (a) ability; and (b) family background.
- We can get a long ways down the road using potential outcomes framework.

Potential Outcomes and College

- Let c_i signal treatment "college". Then,

$$\text{Potential outcome } = \begin{cases} Y_{1i} & \text{if } c_i = 1 \\ Y_{0i} & \text{if } c_i = 0 \end{cases}$$

- Y_{0i} are potential earnings for person i without college
- Y_{1i} are potential earnings for same i with college.

We can state the observed causal effect as:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})c_i$$

Potential Outcomes and College

- We never get to see both Y_{1i} and Y_{0i}, and on casual observation we would just observe:

$$E[Y_i|c_i = 1] - E[Y_i|c_i = 0 = E[Y_{1i} - Y_{0i}|c_i = 1] + E[Y_{0i}|c_i = 1] + E[Y_{0i}|c_i = 0]$$

- Potential earnings (to not going to college) of people who went to college are greater than the potential earnings (to not going to college) of people who didn't go to college
- Positive selection bias overstates ATT

Potential Outcomes and College

CIA

- Conditional on observed characteristics (ability and family), selection bias disappears.
- $E[Y_i|X_i, c_i = 1] - E[Y_i|X_i, c_i = 0] = E[Y_{1i} - Y_{0i}|X_i]$
- $E[Y_{0i}|X_i, c_i = 1] = E[Y_{0i}|X_i, c_i = 0]$

This is great!

- If we can claim that the potential outcomes to not receiving the treatment were the same (there was no selection into treatment)
- We can interpret regression as causal

Broadening the Example

- What if schooling has a different effect for everyone?
- People can take on different amounts of the treatment?

$$Y_{si} \equiv f_i(s)$$

Where

- f_i is the individuals' return to schooling
- (s) is the amount of schooling received

- This is a considerably more general statement than the equivalence across all and only one level of treatment

The **CIA** now becomes:

$$Y_{si} \perp s_i | X_i, \forall s$$

This will help us assess situations where s is assigned conditional on X

Broadening the Example

- In this setup, the *average causal effect* of one more year of schooling is
$$E[f_i(s) - f_i(s-1)|X_i]$$
- And four years is
$$E[f_i(s) - f_i(s-4)|X_i]$$
- We will only ever observe $Y_i = f_i(s)$, but if the **CIA** holds, then average earnings across schooling levels have a causal interpretation.
$$E[Y_i|X_i, s_i = s] - E[Y_i|X_i, s_i = (s-1)$$
$$= E[f_i(s) - f_i(s-1)|X_i]$$
- **Stop.** Look how powerful that is!

Broadening the Example

If the variable you are assessing is independent of potential earnings conditional on X_i, then selection bias vanishes, and and that variable has a causal interpretation.

- Notice that we have the ability - with this set up - to estimate causal effects at *all* values of education.
- This is a lot of potential effects.

This dynamic system leads us back to regression because some τ can summarize the effect of education across all the values.

- Suppose we have a linear, constant effects model of the form:

$$f_i(s) = \alpha + \tau s + \epsilon_i$$

- This says that $f_i(s)$ is linear, and the effect (τ) is the same for everyone
- Then, the only individual specific term in $f_i(s)$ is ϵ_i the unobserved things that determine earnings

Broadening the Example

If we make the substitution of an individual's education into the equation above we have:

$$Y_i = \alpha + \tau s_i + \epsilon_i$$

Which looks like a bi-variate regression, but it is *explicitly* linked to the causal model above!

Causal Models

- Of course, there is likely to be selection into s_i
- So there is some $cov(s_i, epsilon_i)$ (call ability and family).

What if we break ϵ_i into parts?

- Some observable parts X_i and some unobservable υ_i
- Then: $\epsilon_i = X_i'\gamma + \upsilon_i$

Here, γ are regression coefficients of ϵ_i, on X_i, and so $cov(\upsilon_i, X_i) = 0$

Causal Models

Finally, we can write:

$$\begin{aligned} E[f_i(s)|X_i, s_i] = E[f_i(s)|X_i] &= \alpha + \tau s + E[\epsilon_i|X] \\ &= \alpha + \tau s + X_i'\gamma \end{aligned}$$

Then, through the regression setup, $cov(X_i, s_i) = cov(X_i, \epsilon_i) = 0$.