Causal Inference: Causality and Potential Outcomes Framework

Dr. Ria Ivandić

17th January 2022

Causal Inference

"Felix, qui potuit rerum cognoscere causas"

- Virgil, 29BC.

Why Causal Inference?

- Our world is full of causal claims!
- This is inevitable!
- Can you think of anything that relates to knowledge that does not rely on a chain of cause and effect
- Can we answer questions about the (social) world, without making assumptions about whether X and Y are causally related?
- It will make you a better voter, informed citizen, allow you to make better decisions, and differentiate fake news.

Why Causal Inference

- The focus of Causal Inference is applied econometrics, developing tools to answer questions of the form “what is the effect of X on Y ”
  - What is the effect of reading the Daily Mail on voting intentions in the UK?
  - What is the effect of parents’ education on children’s income?
  - What is the effect of the minimum wage on unemployment?
  - What happens to a country if it withdraws from a trade agreement?
  - What is the effect of getting a high grade in Causal Inference on your likelihood of getting into a PhD?
- Answering such questions is difficult. Our world is full of data and observations, but sometimes careless interpretations.

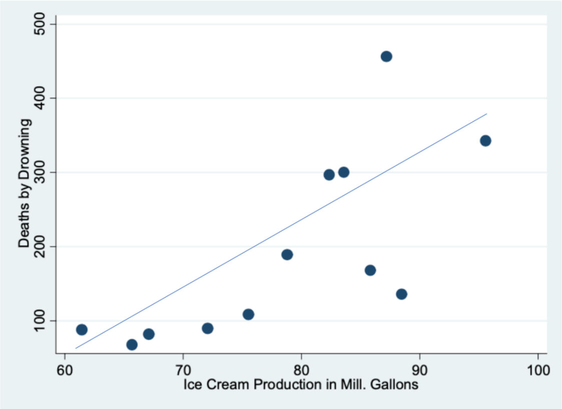
Why Causal Inference

- The focus of Causal Inference is applied econometrics, developing tools to answer questions of the form "what is the effect of X on Y "
  - What is the effect of reading the Daily Mail on voting intentions in the UK?
  - What is the effect of parents' education on children's income?
  - What is the effect of the minimum wage on unemployment?
  - What happens to a country if it withdraws from a trade agreement?
  - What is the effect of getting a high grade in Causal Inference on your likelihood of getting into a PhD?
- Answering such questions is difficult. Our world is full of data and observations, but sometimes careless interpretations.

Why Causal Inference

- The focus of Causal Inference is applied econometrics, developing tools to answer questions of the form "what is the effect of X on Y "
  - What is the effect of reading the Daily Mail on voting intentions in the UK?
  - What is the effect of parents' education on children's income?
  - What is the effect of the minimum wage on unemployment?
  - What happens to a country if it withdraws from a trade agreement?
  - What is the effect of getting a high grade in Causal Inference on your likelihood of getting into a PhD?
- Answering such questions is difficult. Our world is full of data and observations, but sometimes careless interpretations.

What is the Effect of Ice Cream Production on Drowning?

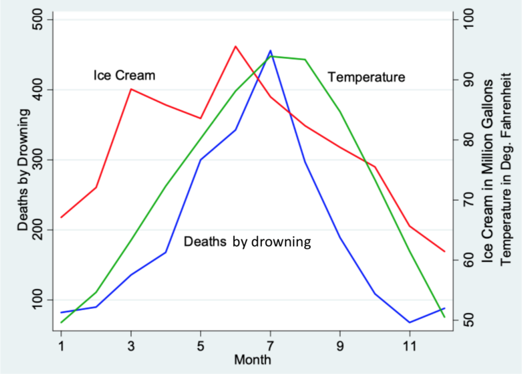


Monthly data for the USA in 2004 – a strong relationship between ice cream sales and drownings

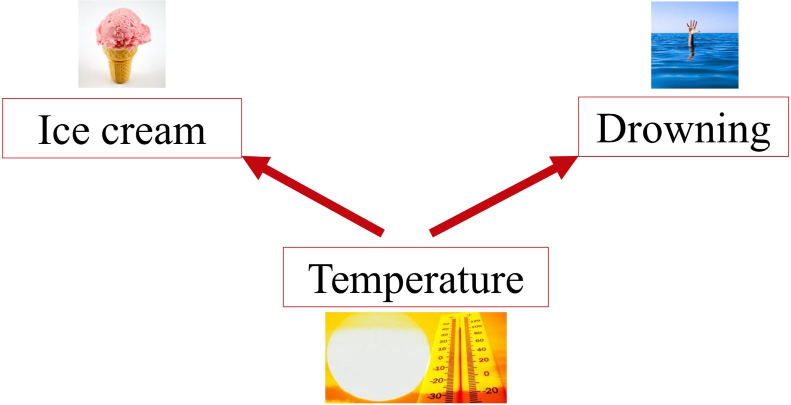
What Causes Drownings?

1. Ice cream
2. Lots of swimming in High temperatures
3. Irresponsible parents
4. Something else
5. I don't know

Drowning, Ice Cream, and the Heat



Correlation is Not Causation: A Third Variable can Cause Both

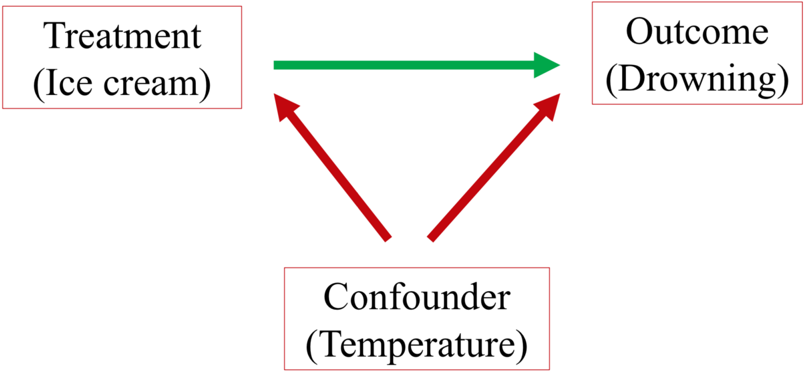


Ice cream

Drowning

Temperature

Correlation is Not Causation: A Third Variable can Cause Both



Treatment (Ice cream)

Outcome (Drowning)

Confounder (Temperature)

Syllabus: Causal Inference

1. Introduction, Causality, Potential Outcomes Framework
2. Randomised Control Experiments
3. Selection on Observables, Multiple Regression and Matching
4. Panel Data and Fixed Effects Models
5. Differences-In-Differences
6. Instrumental Variables and 2SLS Estimation
7. Regression Discontinuity Designs
8. Revisiting Instrumental Variables and Fuzzy Regression Discontinuity Design

Diff-in-diff
$y_{it} = \beta_0 + \beta_1 \cdot treated_i + \beta_2 \cdot post_t + \beta_3 \cdot treated_i \times post_t + u_{i,t}$
treated
non-treated
post
treat
reform
β_3 =
Ref
mptio
t-2
t-1

Dr. Ria Ivandić

I'm a Lecturer in Quantitative Methods at the Department of Politics and International Relations. I'm also affiliated with the London School of Economics and Political Science (Centre for Economic Performance, CEP). I have a PhD in Economics from the Department of Political Economy at King's College London, and a MSc in Economics from Barcelona Graduate School of Economics (Universitat Pompeu Fabra).

Summary of Current Research Agenda

- **Political behaviour: determinants of voter turnout**
  - Does the Number of Candidates Increase Turnout? Causal Evidence From Two-Round Elections (w/ Damien Bol)
  - From Taxes to Polling Booths: The Effect of Economic Policy on Turnout
- **Political economy of crime**
  - Wait and see? Public opinion dynamics after terrorist attacks (w/ Marialiesa Epifanio & Marco Giani)
  - Jihadi Attacks, Media and Local Hate Crime (w/ Steve Machin & Tom Kirchmaier)
  - Football, alcohol and domestic abuse (w/ Tom Kirchmaier and Neus Torres)

Welcome to Causal Inference

Setup of the course:

- 2 hours Lecture
- 1.5 hours Lab
- Assigned readings for both labs and lectures from next week

The lectures:

- Motivation
- Theory
- Examples

The lab sessions:

- Learning by Doing
- Using R, R Studio Cloud

Lab Sessions

- We strongly encourage you to become familiar with the lab assignment before coming to class.
- Download R, R Studio and try to familiarize yourself with the interface.

Two groups:

1. Fridays, 12:15-1:45pm in the Common Room Instructor: Kenneth Stiller (kenneth.stiller@nuffield.ox.ac.uk)
2. Thursdays at 12:00-1:30pm in the Skills Lab - Instructor: Felipe Torres Raposo (felipe.torres@politics.ox.ac.uk)

Auditors and attendance: Auditing students are only allowed to attend lectures, and **not labs**. Under exceptional circumstances, you may request to attend labs by emailing both me and your prospective lab instructor. You have to commit to attending all the eight labs in this case.

Assessments

- **Assignment 1: Distributed Week 1, Deadline Friday of Week 3 (20%)**
- Assignment 2: Distributed Week 3, Deadline Friday of Week 6 (20%)
- Assignment 3: Distributed Week 6, Deadline Friday of Week 9 (20%)
- Take Home Exam: Distributed Week 9, Deadline Friday of Week 0 (TT) (40%)

Assessments, II

- Assessments are intended to motivate you to keep up with the course material, and require work throughout but will ease your final exam preparation.
- Collaboration on problem sets is not allowed. Each student is required to produce his or her own final code and write-up, and to indicate on the write-up which classmates he or she collaborated with. Students will be penalized for violating this rule.
- All assignments should be submitted to the course canvas website in PDF format. In labs we will teach you to produce PDF reports using R markdown, which allows you to produce clearly formatted documents with math notations (using Latex), R code, and R output. Please make sure to keep within the word limit of each assignment.

Office Hours

Office Hours:

- Ria: Monday afternoons (4-5pm), by appointment on ria.ivandic@conted.ox.ac.uk
- Felipe: Thursday afternoons, by appointment on felipe.torres@politics.ox.ac.uk
- Ken: Tuesday afternoon, by appointment on kenneth.stiller@nuffield.ox.ac.uk

Questions:

- Slides and course materials can be found on canvas.ox.ac.uk. If you do not have the course registered in your account please contact DPIR postgraduate team (pg.studies@politics.ox.ac.uk).
- We will also enable an online message board for discussion in canvas. As a first point of contact, please do not send a general question to the instructors, but post directly to the message board so others can benefit.

Covid Regulations

- Lectures and labs remain **in person**. There will be no optional Zoom attendance.
- Please wear a mask in class. If you have an exemption, please let me know by email.
- Staff and students coming onsite are strongly encouraged to take an LFD test twice a week, every week. **I would appreciate if you took a test every Monday morning before coming to class.**
- If you are feeling unwell, even as a precaution, please do not come to class.
- For students who are self-isolating, we will provide hybrid teaching over Zoom. When you find out, please contact the admin team as soon as possible CC-ing me.

Assignment

- The world is full of incorrect causal claims.
- Look out for them!
- If you find a really bad one, send them to me by email.
- Best one (i.e. most ridiculous incorrect claim) gets an honorary mention at the end of lecture 8!

Table of Contents

Introduction

Course Specifics

Causality and Potential Outcomes

Selection bias

Definitions

Causality

Refers to the relationship between events where one set of events (the effects) is a direct consequence of another set of events (the causes). (Hidalgo & Sekhon 2012)

Causal Inference

The process by which one can use data to make claims about causal relationships. (Hidalgo & Sekhon 2012)

Inferring causal relationships is a central task of science.

Examples

- What is the effect of peace-keeping missions on peace?
- What is the effect of church attendance on social capital?
- What is the effect of minimum wage on employment?

A Counterfactual Logic

Counterfactual Logic

If X had/had not been the case, Y would/would not have happened

Example:*Does college education increase earnings?*

- If high school grads had instead obtained a college degree, how much would their income change?
- If college grads had only obtained a high school diploma, how much would their income change?

Mill & the Counterfactual Logic of Causality

Causal Inference as a counterfactual Problem

Rather than defining causality purely in reference to observable events, counterfactual models define causation in terms of a comparison of observable and unobservable events.

- We need to construct counterfactuals of the observed world as comparison units (Method of Difference).

Revisiting Linear Regression Models

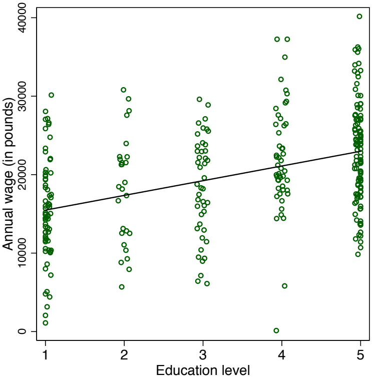
Question: What is the effect of schooling on earnings?

- We obtained some data
- Level of complete education: X
- Annual Income: Y

Linear Equation (OLS Review): The easiest, most parsimonious although not always most adequate, way to summarize the conditional expectation of Y, given X, is to specify a linear model between X and y:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

The Regression Model



- Think of Y as *Salary* and X as *years or schooling*. What we are interested in is β_1. We would like to interpret this coefficient as the *average change produced in individuals' wages by one more year of schooling.*

Linear Regression Model Assumptions (Gauss-Markov Theorem)

$E(Y|X) = \beta_0 + \beta X$

- What does β_0 stand for?
  - $E(Y|X = 0)$
- What does β_1 stand for?
  - The amount of change in Y in response to a unit change in X
  - Does it matter what the value of X is? No, because there is only one slope.

An Important OLS Assumption

OLS estimation and inference requires various assumptions but we are interested in one of them here. The one that would allow us to interpret β_1 as the average causal effect of education on economic well-being.

The zero-mean assumption

The error term, u_i, is centered around 0 across all values of X: $E[u|X] = 0$. Our X(s) and the error term are *independent*.

Would this assumption hold in our example?

An Important OLS Assumption

Omitted variable bias: This assumption means that all unexplained factors contributing in one's salary have the same value (trivially recoded to zero with the inclusion of a constant) no matter whether one holds a university degree or a high-school degree.

- To think whether this assumption is violated we need to think of additional determinants of income. How about parental socioeconomic status (SES)?
- Imagine, then, the true model generating income is the following:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + u_i$$

where $Y = \textit{Salary}$, $X = \textit{Years in Education}$, $Z = ParentalSES$.

The Problem

Since we now know the correct model, we can see what would have happened if we had estimated β_1 from the bivariate regression.

The True Model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + u_i$$

Model without including Parental SES

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

Putting the zero-mean assumption into context

The assumption would assume that the expected parental SES is the same among all people of different educational levels, for those with primary education and those with university degree: $E[u|X] = E[u] = 0$, equivalent to: $E[Z|X] = E[Z]$. This is unlikely to hold, so a model without parental SES (Z) would not estimate the causal effect.

The Selection-on-Observables Assumption

Including Covariates

— Do we really believe that controlling for parental SES is sufficient, i.e. that we know the true model of economic well-being?
— No, this is why we do not only include Parental SES, but other possible predictors of Y.
The reason for doing this is that we want to make the zero-mean assumption more plausible.

The Conditioning-on-Observables Assumptions

Conditioning on a vector of covariates, Z, we believe that this equality holds: $E[u|X] = E[u] = 0$.

Why do we need causal inference then? The pitfalls of observational research

- Clearly, we don't really know what Z includes. Many plausible candidates might not be measured and might even be measurable. How would one measure ability or aspiration?
- Estimating relationships with controls is certainly better than without them
- But, we never account for everything. We just can't!
- **If we are interested in causal links, we need a better, clearer, and stronger framework to understand the social world**

A hypothetical example

Imagine two students who are interested in getting a very high score on their thesis. They are considering the courses they should take and they are undecided between *Causal Inference* or sticking with MT's *Intro to Statistics*.

Y_i : Thesis score is the outcome variable of interest for unit i.

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment (taking Causal inference)} \\ 0 & \text{otherwise.} \end{cases}$$

$$Y_{di} = \begin{cases} Y_{1i} & \text{Potential thesis score for student } i \text{ with Causal Inference} \\ Y_{0i} & \text{Potential thesis score for student } i \text{ without Causal Inference} \end{cases}$$

Q: What is the effect of taking Causal Inference on your thesis score?

Defining the Potential Outcomes

Definition: Treatment

D_i : Indicator of treatment status for unit i

$$D_i = \begin{cases} 1 & \text{if unit } i \text{ received the treatment} \\ 0 & \text{otherwise.} \end{cases}$$

Definition: Observed Outcome

Y_i : Observed outcome variable of interest for unit i. (Realized after the treatment has been assigned)

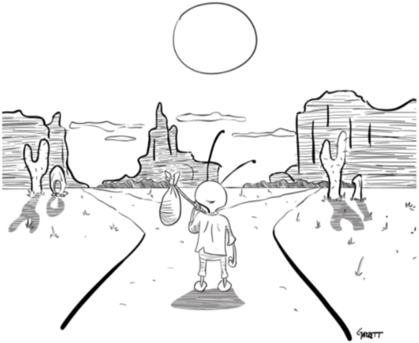
Defining the Potential Outcomes

Definition: Potential Outcomes

Y_{0i} and Y_{1i}: Potential Outcomes for unit i

$$Y_{di} = \begin{cases} Y_{1i} & \text{Potential outcome for unit } i \text{ with treatment} \\ Y_{0i} & \text{Potential outcome for unit } i \text{ without treatment} \end{cases}$$

The Road Not Taken (by Robert Frost)



It shows (1) the actual road that you chose, and (2) the counterfactual road that you could have chosen but did not. **How can we know it made the difference? We don't know what would have happened on the other path.**

Causality with Potential Outcomes

Let D_i denote a binary treatment for unit i, where $D_i \in 0, 1$. Let Y_i represent the observed outcome for unit i. The potential outcomes are thus, Y_{1i}, Y_{0i}

The causal effect of D on Y for i is $\tau_i = Y_{1i} - Y_{0i}$

Definition: Causal Effect

Causal Effect of the treatment on the outcome for unit i is the difference between its two potential outcomes:

$$\tau_i = Y_{1i} - Y_{0i}$$

The Fundamental Problem of Causal Inference

It is impossible to observe for the same unit i the values $D_i = 1$ and $D_i = 0$ as well as the values Y_{1i} and Y_{0i} and, therefore, it is impossible to observe the effect of D on Y for unit i.

This is why we call this a missing data problem. We cannot observe both potential outcomes, hence we cannot estimate:

$$\tau_i = Y_{1i} - Y_{0i}$$

		Y_{i1}	Y_{i0}
Person 1	Treatment Group ($D = 1$)	Observable as Y	Counterfactual
Person 2	Control Group ($D = 0$)	Counterfactual	Observable as Y

Quantities of Interest

Definition ATE

Average Treatment Effect:

$$\tau_{ATE} = E[Y_1 - Y_0]$$

Definition ATT

Average Treatment Effect of the Treated:

$$\tau_{ATT} = E[Y_1 - Y_0|D = 1]$$

Definition ATC

Average Treatment Effect of the Controls:

$$\tau_{ATC} = E[Y_1 - Y_0|D = 0]$$

An Example: ATE

Imagine a population of 4 units:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	?	?	?
2	1	1	?	?	?
3	0	0	?	?	?
4	1	0	?	?	?

What is the ATE?

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}]$$

An Example: ATE

Imagine a population of 4 units:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	3	?	?
2	1	1	1	?	?
3	0	0	?	0	?
4	1	0	?	1	?

What is the ATE?

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}]$$

Because we cannot observe two worlds at the same time, we cannot calculate the ATE.

An Example: ATE

Imagine a population of 4 units with the counterfactual values are made up! (As all other values in this example!)

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	1	0	1	1	0

What is the ATE?

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}] = 4/4 = 1$$

An Example: ATE (Continued)

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	3	0	3
2	1	1	1	1	0
3	0	0	1	0	1
4	1	0	1	1	0
$E[Y_1]$			1.5		
$E[Y_0]$				0.5	
$E[Y_1 - Y_0]$					1

$$\tau_{ATE} = E[Y_{1i} - Y_{0i}] = 1/4 \cdot (3 + 0 + 1 + 0) = 1$$

An Example: Incorrect ATE

In reality you only get the following:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	3	1	3	?	?
2	1	1	1	?	?
3	0	0	?	0	?
4	1	0	?	1	?

$$\textbf{Wrong } \tau_{ATE} = E[Y_{1i} - Y_{0i}] = 2 - 0.5 = 1.5$$

What is the identification problem?

$$\tau_{ATE} = E[Y_1 - Y_0]$$

$$= \pi(E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1])$$
$$+ \text{ (1-}\pi)(E[Y_{1i}|D_i = 0] - E[Y_{0i}|D_i = 0])$$

where π is the share of the treated units in our sample.

What can we observe from the above equation?

1. π
2. $E[Y_{1i}|D_i = 1]$
3. $E[Y_{0i}|D_i = 0]$

What can't we observe from the above equation?

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

Counterfactual outcomes!

What is the identification problem?

The observed difference in the outcome for the treatment and control group are:

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

$$\mathrm{E}[\mathrm{Y}_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] + E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0] =$$

$$\underbrace{E[Y_{i1} - Y_{i0}|D_i = 1]}_{ATT} + \underbrace{E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]}_{SelectionBias}$$

- ATT: Average treatment effect on the treated
- Selection Bias: Differences in the treated and control groups when assigned to the control group.

Both are unobserved and we need to make assumptions!

Real world selection bias: health insurance

- Health insurance in the US used to be voluntary. Obama's "Affordable Care Act" required citizens to buy insurance. Trump rescinded the mandate in 2019. Is mandatory health insurance a good policy? Similarly, is the NHS in the UK a good policy?
- **Question: What is the effect of health insurance on expenditures and health outcomes?**
- The US spends a larger fraction of GDP on health care than most other developed countries, but Americans are not necessarily healthier. Health insurance supposedly has health benefits. Yet, heath care is expensive and requiring it may be considered to violate freedom.

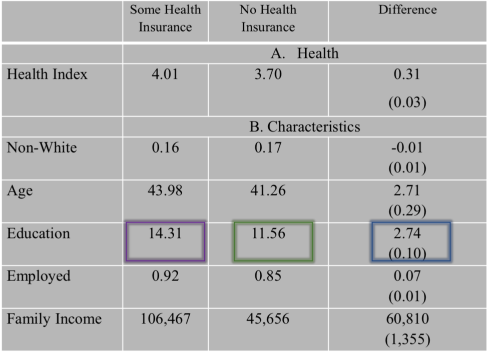
What is the Effect of Health Insurance on Health?

We want to compare:

- The health of someone with insurance.
- The health **the same person** without insurance.

We begin by comparing health outcomes between those with and without insurance.
Should we do this? Are there any potential issues?

What is the Effect of Health Insurance on Health?



	Some Health Insurance	No Health Insurance	Difference
	A. Health		
Health Index	4.01	3.70	0.31 (0.03)
	B. Characteristics		
Non-White	0.16	0.17	-0.01 (0.01)
Age	43.98	41.26	2.71 (0.29)
Education	14.31	11.56	2.74 (0.10)
Employed	0.92	0.85	0.07 (0.01)
Family Income	106,467	45,656	60,810 (1,355)

The observed difference in the outcome for the treatment and control group are:

$$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] = \underbrace{E[Y_{i1} - Y_{i0}|D_i = 1]}_{ATT} + \underbrace{E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]}_{SelectionBias}$$

What is the Effect of Health Insurance on Health?

What can cause a difference in health outcomes for individuals with and without health insurance?

1. Causation: having health insurance directly leads to better health.
2. Reverse causality: the less (or more) healthy are more likely to buy insurance.
3. Confounders: e.g., the more educated tend to buy insurance more often and they know how to live healthier.
4. Any other ideas?

What is the Effect of Health Insurance on Health? Selection bias

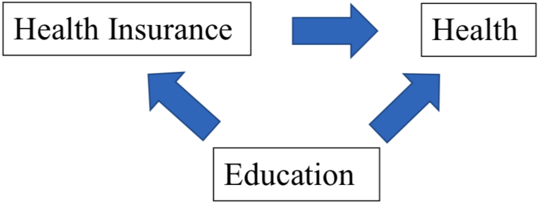
The observed difference in the outcome for the treatment and control group are:

$$E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] = \underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{ATT} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{Selection Bias}$$

Selection bias is the difference in the potential untreated outcomes between those who were treated and who were not treated. Can we approximate the selection bias here? Is it positive or negative?

$$\underbrace{E[Y_{i0}|D_i = 1] - E[Y_{i0}|D_i = 0]}_{Selection Bias}$$

What is the Effect of Health Insurance on Health? Selection bias



Selection bias is when treatment is assigned in a manner that also affects the outcome. In our example confounders, e.g., education levels, may affect health. Education may also affect the choice to attain insurance. Thus, potential outcomes differ for individuals with and without insurance.

From observed mean differences to ATT

We would like to turn this:

$$\underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{ATT} + \underbrace{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]}_{SelectionBias}$$

Into this:

$$\underbrace{E[Y_{1i} - Y_{0i}|D_i = 1]}_{ATT} + \underbrace{0}_{SelectionBias}$$

This means that:

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

This equality would make $ATT = ATE$. What assumptions do we need to claim this equality and how do we get there?

Randomization

Randomization and Selection Bias

With randomly assigned D_i there is No Selection Bias! Units i are similar on all (un)observed traits and only differ in terms of D_i

Because both conditional expectations $E[Y_{1i}|D=1]$ and $E[Y_{0i}|D=0]$ come from the same underlying population, we can claim that when D_i is randomly assigned, the units are interchangeable.

Is randomization always possible? Motivation for causal inference.

Key Assumptions for Identification

1. **Conditional Independence Assumption**

$$Y_{1i}, Y_{0i} \perp\!\!\!\perp D_i$$

This suggests that:

$$E[Y_{1i}|D=1] = E[Y_{0i} + D_i(Y_{1i} - Y_{0i})|D_i = 1] = E[Y_{1i}|D_i = 1] = E[Y_{1i}]$$

This assumptions allows the Expectations of the unobservables equal the conditional expectations of the observables for control and treatment. Hence, $E[Y_{i1}|D_i = 1] = E[Y_{i1}]$

Key Assumptions for Identification

2. **Stable unit treatment value assumption (SUTVA)**
    1. Consistency in the treatment group
    2. No Spillover across treatment groups. No interference!

3. **Unconfoundedness**

The causal effect only runs from D_i to Y_i

1. Common Cause ($D \perp\!\!\!\perp Y$, D causes Z, Z causes Y)
2. Common Effect (selection on the DV, post treatment bias)

Experimental vs Observational Studies

Definition: Observational Study

An observational study is an empiric investigation of the effects of exposure to different treatment regimes, in which the investigator cannot control the assignment of treatment.

- This means that control and treatment units are not automatically exchangeable.
- Does this mean we can only work with experimental data?
- Of course not. This is why we add controls to our regression models. To adjust for the observed covariates.
- What about the unobservables?
- Well, we hope they are also balanced.
- Is balance -or exchangeability- testable? NO!