Week 1: Potential Outcomes and Causal Inference

PUBL0050 Causal Inference

Julia de Romémont

Term 2 2023-24

UCL Departement of Political Science

This Lecture

Causal Questions

Course Outline and Logistics

Counterfactuals and Causality

The "Potential Outcomes" Framework

Selection Bias

Causal Questions

Social science questions

Quantitative social science addresses many different types of question:

1. **Descriptive questions**
   - E.g. What is democracy?
2. **Measurement questions**
   - E.g. Which countries are democratic? How democratic are they?
3. **Prediction questions**
   - E.g. If we know other things about a country, can we predict how democratic it is?
4. **Causation questions**
   - E.g. What causes a country to become democratic?

Knowledge and Causality

> *We do not have knowledge of a thing until we have grasped its why, that is to say, its cause.*
>
> *–Aristotle, Physics*

In this course, we will be learning to make **inferences** about **causal relationships**:

Causality

The relationship between events where one set of events (the effects) is a direct consequence of another set of events (the causes).

Causal Inference

The process by which one can use data to make claims about causal relationships.

Causal Inference is **important**

1. **Social science theories** are inherently causal.
2. **Improving public policy** requires knowing "what works".
3. **Being good citizens** requires understanding when there is sufficient causal evidence to endorse a given political argument.

Causal Questions

We will focus on measuring the effect of some treatment on some outcome:

- What is the effect of canvassing on vote intention?
- What is the effect of class size on test scores?
- What is the effect of social distancing on COVID infection rates?
- What is the effect of peace-keeping missions on peace?
- What is the effect of austerity on support for Brexit?
- What is the effect of institutions on growth?
- What is the effect of smoking bans on public health?
- What is the effect of job training on employment?
- What is the effect of education on anti-immigrant views?

Causal Questions

Note the difference between these two forms of causal questions:

- Does phenomenon X have a causal effect on phenomenon Y
  - 'Effects of causes' questions
- What are the phenomena that cause phenomenon Y?
  - 'Causes of effects' questions

Our focus will be exclusively on questions of the first type.

Course Outline and Logistics

Teaching staff

Course Convenor

Dr. Julia de Romémont

- E-mail: j.romemont@ucl.ac.uk
- Student support and feedback hours: Wednesday 11.45-12.45; Thursdays 11-12

Seminar Leader

Tom Barton

- Email: t.barton@ucl.ac.uk
- Student Support and feedback hours: Tuesdays 11-12; Fridays 12.30-1.30 (Drop in)

Prerequisites

This is a course for students with **at least one** prior module in quantitative methods. The course builds directly from the material that covered PUBL0055.

In particular, it assumes you have a good working knowledge of:

- Linear regression
- T-tests
- Statistical inference
- `R` and `Rstudio`

Learning Objectives

The goal of quantitative empirical research is to **learn from data about how the world works**.

To this end, we aim to **uncover patterns and regularities** within our data and

- Generalise them (**Population Inference**)
- Characterise them (**Measurement Inference**)
- Explain them (**Causal Inference**)

Learning objectives

This course will provide you with:

- An understanding of the assumptions required to support causal claims made with quantitative data
- An overview of the most commonly used statistical methods which aim to estimate causal effects
- The practical skills required to implement these methods using `R`

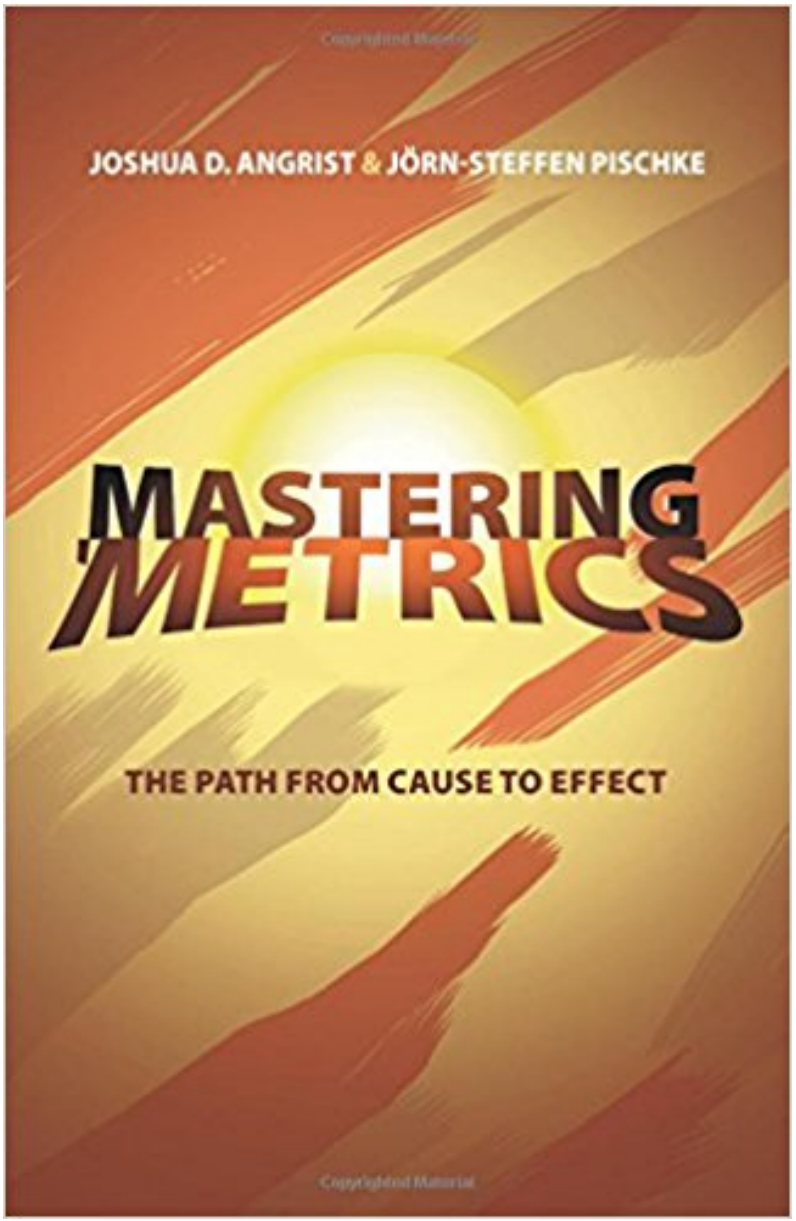
Schedule

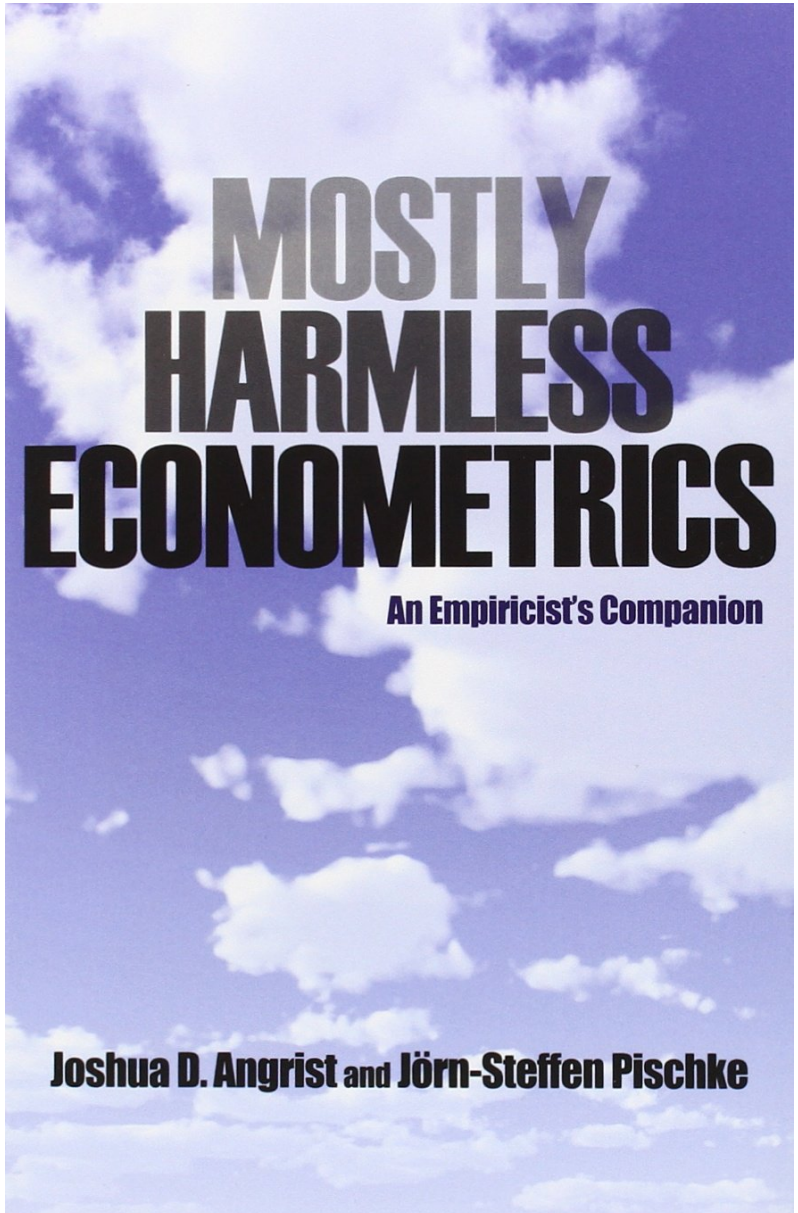
1. Causal Inference and Potential Outcomes
2. Randomised Experiments
3. Selection on Observables I: Subclassification & Matching
4. Selection on Observables II: Regression
5. Panel Data and Difference-in-Differences

Reading Week

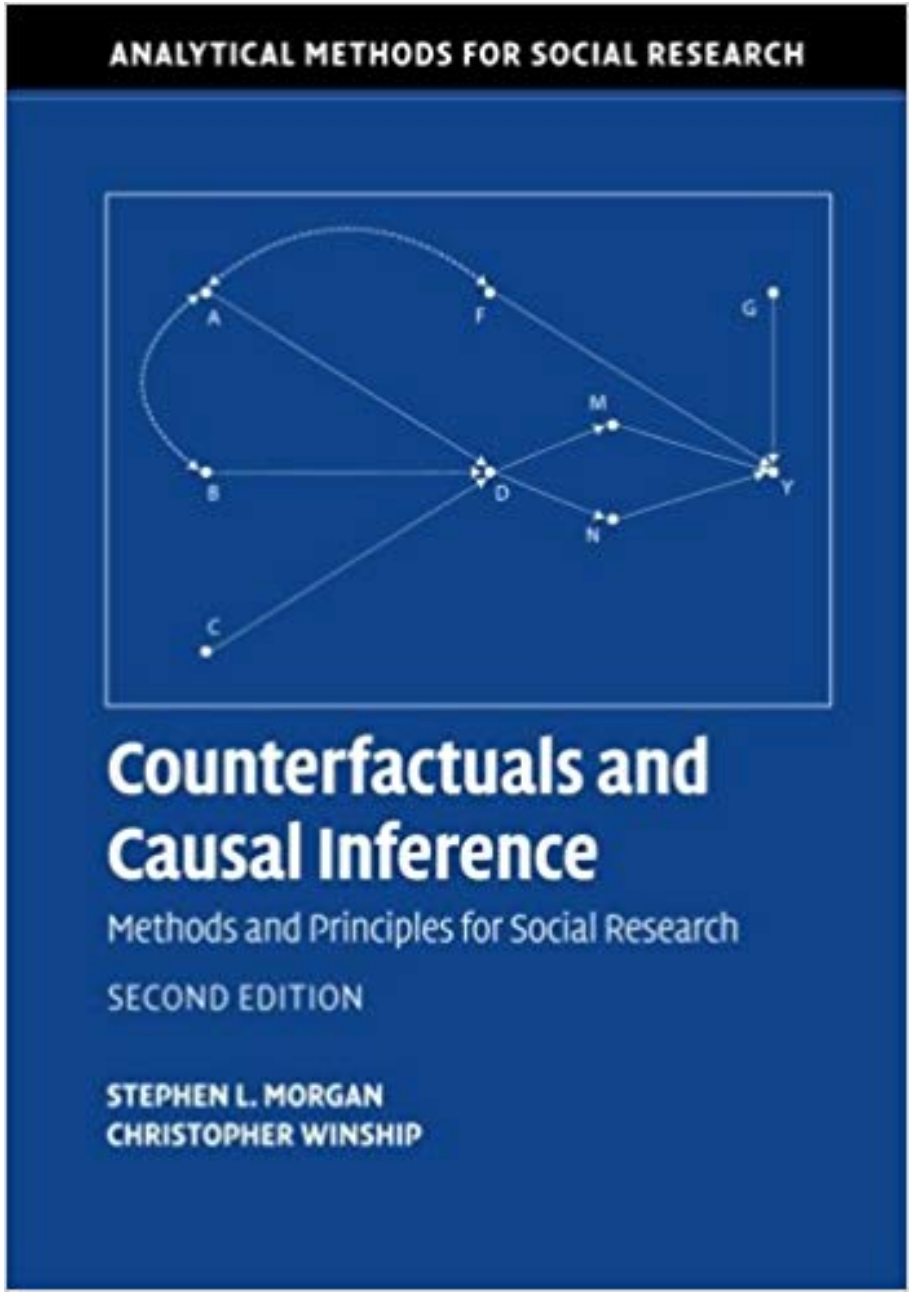
6. Synthetic Control
7. Instrumental Variables I
8. Instrumental Variables II
9. Regression Discontinuity Designs
10. Overview and Review

Course Texts I





Course Texts II



Assessment

100% of the assessment will be via a 3000 word research paper due on **April 22th, 2024 at 2pm**.

- Design an original research study to answer a causal question
- Any substantive area in the social sciences is fine
- Must use (at least) one technique from the course

Guidance on the structure of these papers can be found on the course website.

Course Website

Course Logistics

Lectures

- Lectures will be held on Wednesday 9-11am. Lectures will be recorded and uploaded to Moodle

Seminars

- Weekly one-hour practical session on lecture topic
- Fridays 10am, 11am, 2pm, 3pm
- Please attempt the seminar questions *before* attending the seminar!
- Solutions will be made visible on the Monday after

Recommended Workflow

1. Complete required reading
2. Attend the lecture
3. Attempt the seminar assignments[1]
4. Attend the seminar
5. Go back through the seminar assignment and complete it by using the provided solutions
6. Make note of any unanswered questions, ask them in the seminars, lectures or SSF
7. Regularly go back through slides, assignments and readings of earlier weeks, and, *help each other*!

[1] At the *very* least, read the prompt and load the data into R

How to participate in this course

Academic freedom

- Everyone must respect freedom of thought and freedom of expression
- You are explicitly prohibited from recording, publishing, distributing or transferring any class material/content

Expectations

- We are happy to answer any questions and explain things (again)
  - But we expect you to be up to date with the material, and read instructions and informational material diligently
- You can ask us questions in person (lectures, seminars, SSF) and via the Moodle forum
  - Only contact us via email if
    - We asked you to
    - Your question is urgent
    - *Your question can be answered with two sentences or less*

On the use of AI

The Department has developed guidelines specific to quantitative methods courses which you should read and can find here.

The bottom line:

- ChatGPT & other tools *can* be useful to understand, write or correct ('debug') or explain some of the concepts discussed in the course
- **However**, you will not be able to know if what a given AI tool tells you is correct if you don't engage with the course material yourself!
- You can use AI tools in the assignment but
  - Only for certain tasks; and
  - With appropriate acknowledgment and referencing

Counterfactuals and Causality

The Meaning of Causation

> *We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would have happened without it.*
>
> – *David Lewis, Causation, 1973*

This is a **counterfactual** view of causality:

- One variable, X, is understood to cause another variable, Y, if the value for Y *would have been different* for a different value of X

The Road Not Taken

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I -
I took the one less traveled by,
And that has made all the difference.

–Robert Frost (1874 - 1963)

- Potential outcomes
- Causal effects
- Fundamental problem of causal inference
- Inference

What If..?

There are **a lot** of movies, series, books etc, devoted to exploring the counterfactual! E.g.

- Everything Everywhere all at Once
- The Man in the High Castle
- A slightly older example: Sliding Doors

> *Time. Space. Reality. It's more than a linear path. It's a prism of endless possibility, where a single choice can branch out into infinite realities, creating alternate worlds from the ones you know. I am the Watcher. I am your guide through these vast new realities. Follow me, and ponder the question... What if?*
>
> *– The Watcher in Marvel's What If...?*

Potential Outcomes

What does it mean for a **causal factor** to affect an **outcome** for an **individual**?

- e.g. time spent studying
- e.g. final grade on this course
- e.g. you

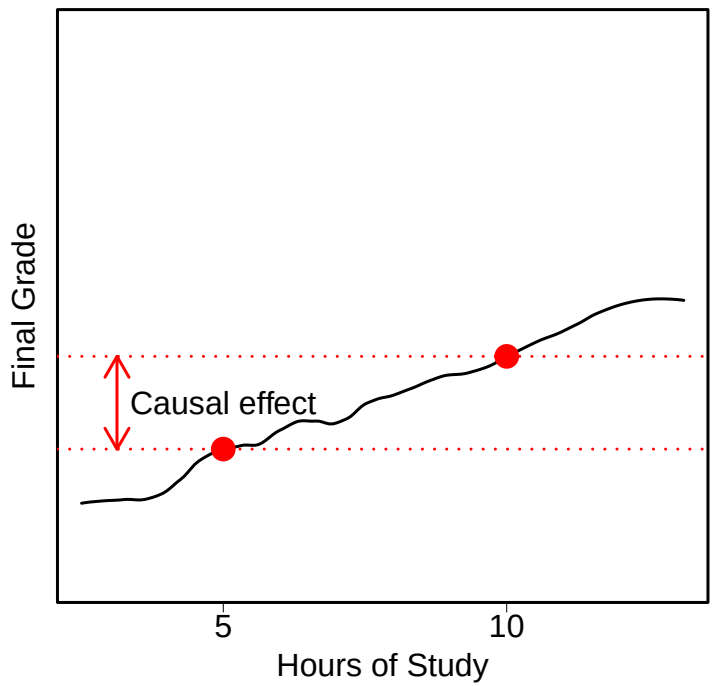
In the **counterfactual** approach to causation: if the outcome for the individual would be different for different hypothetical values of the causal factor.

These hypothetical outcomes (associated with hypothetical values of the causal factor) are known as **potential outcomes.**

Potential Outcomes Illustration

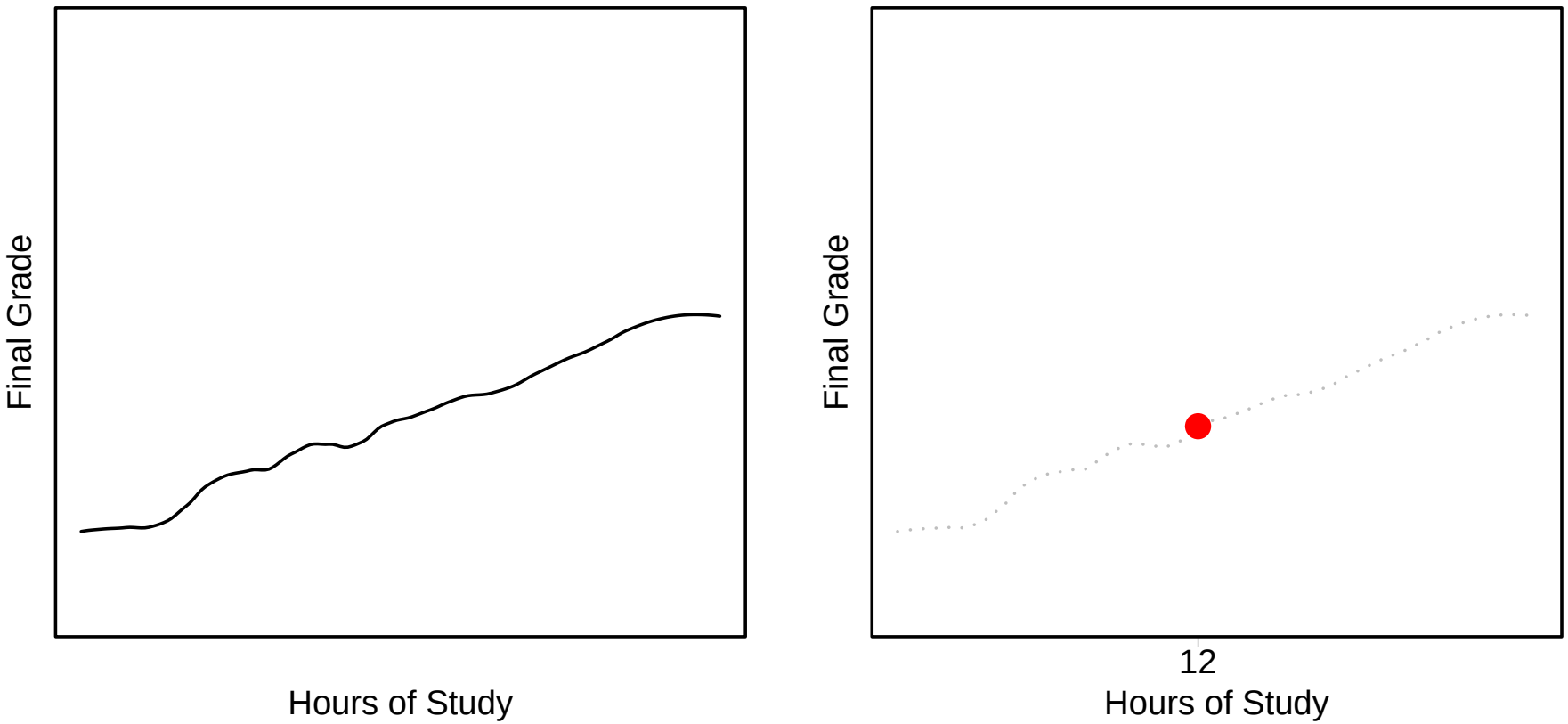
Imagine we knew the grade a particular individual **would** receive for different amounts of study time:

- Each point on the line represents a potential outcome (the hypothetical outcome associated with a each value of our causal factor)
- **Causal effects** are defined in terms of potential outcomes
- If we could observe all potential outcomes for an individual, we could quantify...
  - ...the effect of spending 10 hours per week studying rather than 5
  - ...the average effect of spending an additional hour studying
  - ...and so on...



Fundamental Problem of Causal Inference

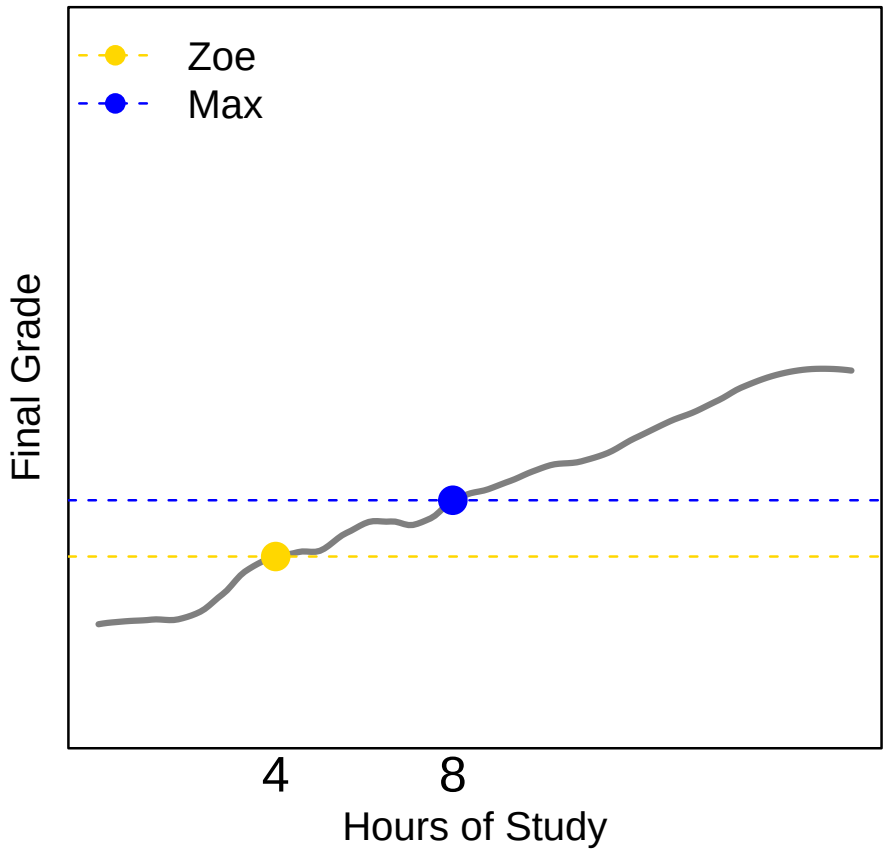
For any given unit/individual we **only observe one** potential outcome:



→ To make causal inferences we have to impute (make educated guesses about) the other unrealised potential outcomes

Making Comparisons Across Units

- Tempting: make comparisons of **realised outcomes** across units with different values of the **causal factor**
- This can lead to erroneous inferences (conclusions)



What's the Problem?

1. **Heterogeneity**: Different units have different potential outcomes
2. **Confounding**: when the values of the causal factor are systematically related to the potential outcomes
    - **(e.g. smarter students study more/less, students who benefit more from studying study more/less)**

These are the two central problems that the tools we study on the course aim to overcome.

Learning about things we cannot see

A central goal in statistics is **inference**: to draw conclusions about things we cannot see from things that we can see.

Statistical inference

The process of drawing conclusions about features/properties of the **population** on the basis of a **sample**.

Causal inference

The process of drawing conclusions about features/properties of the **full set of potential outcomes** on the basis of some **observed outcomes**.

Generally we will be attempting to do both: using a sample to draw conclusions about the full set of potential outcomes in the population.

The "Potential Outcomes" Framework

Potential Outcomes History

> *If a person eats of a particular dish, and dies in consequence, that is,* ***would not have died if he had not*** *eaten of it, people would be apt to say that eating of that dish was the cause of his death*
>
> *–John Stuart Mill, "The Law of Causation" in Logic, 1843*

Counter-factual definitions of causality have a long intellectual history:

- Neyman (1923) introduced the potential outcomes notation for experiments.
- Fisher (1925) proposed randomizing treatments to units.
- Rubin (1974) then extended the potential outcomes framework ("Rubin Causal Model", Holland, 1986) to observational studies

Potential Outcomes Notation I

Definition: Treatment

D_i: Indicator of treatment intake for unit i

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

- D_i denotes the **treatment** (**causal variable**) for unit i e.g.
  - $D_i = 1$: asprin; $D_i = 0$: no asprin
  - $D_i = 1$: encouragement to vote; $D_i = 0$: no encouragement to vote
- Defined for binary case, but we can (and will) generalise to continuous treatments

Potential Outcomes Notation II

Definition: Outcome

Y_i: Observed outcome variable of interest for unit i

Definition: Potential Outcome

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

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

- If $D_i = 1$, Y_{0i} is what the outcome **would** have been if D_i had been 0
- If $D_i = 0$, Y_{1i} is what the outcome **would** have been if D_i had been 1

$\rightarrow$ potential outcomes are fixed attributes for each i and represent the outcome that would be observed hypothetically if i were treated/untreated

Potential Outcomes Notation III

Given Y_{0i} and Y_{1i}, it is straightforward to define a causal effect.

Definition: Causal Effect

For each unit i, the causal effect of the treatment on the outcome is defined as the difference between its two potential outcomes:

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

- τ_i is the difference between two hypothetical states of the world
  - One where i receives the treatment
  - One where i does not receive the treatment

Potential Outcomes Notation IV

Assumption

The outcome (Y_i) is connected to the potential outcomes (Y_{0i},Y_{1i}) via:

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}$$

so

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

- *A priori* each potential outcome **could be** observed
- After treatment, one outcome **is** observed, the other is *counterfactual*

Fundamental Problem of Causal Inference

Definition: Fundamental Problem of Causal Inference

We cannot observe both potential outcomes (Y_{1i}, Y_{0i}) for the same unit i

Causal inference is difficult because it is about something we can never see.

–Paul Rosenbaum, Observation and Experiment

'No Interference' Assumption

Recall that observed outcomes are realized as

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}$$

- The 'non-interference' assumption, implies that potential outcomes for unit i are unaffected by treatment assignment for unit j
- Also known as the Stable Unit Treatment Value Assumption (SUTVA)
- Rules out interference among units
- Examples:
  - Effect of GOTV on spouse's turnout
  - Effect of medication when patients share drugs

Illustration: Individual Treatment Effects τ_i

Imagine a population with 4 units, where we observe both potential outcomes for each unit:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	5	1	5	2	3
2	2	1	2	1	1
3	0	0	1	0	1
4	1	0	1	1	0

If we know both potential outcomes for each unit, we can easily estimate the **Average Treatment Effect**:

$$\begin{aligned} \tau_{\text{ATE}} &\equiv E[Y_{1i} - Y_{0i}] \\ &= \frac{1}{N}\sum_{i=1}^{N}(Y_{1i} - Y_{0i}) \\ &= E[Y_{1i}] - E[Y_{0i}] \end{aligned}$$

$$\textbf{ATE} = \frac{3+1+1+0}{4} = \frac{5+2+1+1}{4} - \frac{2+1+0+1}{4} = 1.25$$

Illustration: FPOCI

We never actually observe **both** potential outcomes.

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	5	1	5	?	?
2	2	1	2	?	?
3	0	0	?	0	?
4	1	0	?	1	?

$$\textbf{ATE} = \frac{? + ? + ? + ?}{4} = ?$$

Estimating individual effects or average effects requires observing **unobservable** potential outcomes and therefore cannot be accomplished without additional assumptions.

The Value of Potential Outcomes

The strength of [the potential outcomes framework] is that it allows us to make these assumptions more explicit than they usually are. When they are explicitly stated, the analyst can then begin to look for ways to evaluate or partially test them.

–Holland, 1986

Selection Bias

Illustration: Selection Bias

One intuitive approach is to **make comparisons across units** using Y_i.

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	5	1	5	?	?
2	2	1	2	?	?
3	0	0	?	0	?
4	1	0	?	1	?

i.e. Compare the average observed outcome under treatment to the average observed outcome under control.

- However, recalling that the 'true' **ATE** $= 1.25$
- But the **difference in means** $= \frac{5+2}{2} - \frac{0+1}{2} = 3$
- so, in this example at least, **difference in means** $\neq$ **ATE**

Statistics Concepts and Notation

To proceed to a more general statement of selection bias, we will need some key building blocks of statistics

- **Population** – all the units of interest
- **Sample** – a subset of the units in the population
- **Variable (e.g. Y_{1i}, X_i)** – a numerical measure
- **Random Variable** – a variable whose outcome is the result of a random process
  - number on a six-sided dice
  - treatment status of unit i in a randomized experiment
  - Y_{1i} for a unit randomly selected from the population

Expectations

- The **expectation** of a random variable X is denoted $E[X]$, where $E[X] \equiv \sum xPr[X = x]$
- This is the average outcome of the random variable
  - e.g. for a six-sided dice: $E[X] = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = 3.5$

Conditional Expectations

- The **conditional expectation** of a variable refers to the expectation of that variable amongst some subgroup
- e.g. $E[Y_{1i}|D_i = 1]$ gives the conditional expectation of the random variable Y_{1i} when $D_i = 1$

Statistical Terms

- **Parameter/Estimand**

  $\rightarrow$ a fixed feature of the population (e.g. $E[Y_{1i}]$)

- **Sample statistic**

  $\rightarrow$ a random variable whose value depends on the sample (e.g. the sample mean, $\bar{Y}$)

- **Estimator**

  $\rightarrow$ any function of sample data used to estimate a parameter (e.g. $\frac{1}{n}\sum_{i=1}^{n} Y_i$)

- **Unbiased estimator**

  $\rightarrow$ A statistic is an unbiased estimator of a parameter if the **estimate** it produces is on average **(i.e. across an infinite number of samples)** equal to the parameter (e.g. $E[\frac{1}{n}\sum_{i=1}^{n} Y_i] = E[Y_i]$)

Definitions of Common Quantities of Interest

- An intuitive quantity of interest **(parameter)** is the **Average Treatment Effect**
  - i.e. how outcomes would change, on average, if every unit were to go from untreated to treated (defined by potential outcomes).

Average Treatment Effect (ATE)

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

Definitions of Common Quantities of Interest

Average Treatment Effect (ATE)

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

Average Treatment Effect on the Treated (ATT)

$$\tau_{\text{ATT}} \equiv E[Y_{1i} - Y_{0i} | D_i = 1]$$

Average Treatment Effect on the Controls (ATC)

$$\tau_{\text{ATC}} \equiv E[Y_{1i} - Y_{0i} | D_i = 0]$$

Average Treatment Effects for Subgroups (CATE)

$$\tau_{\text{ATE}_X} \equiv E[Y_{1i} - Y_{0i} | X_i = x]$$

Difference in Group Means and the ATE

- For a given sample, one obvious **estimator** of the ATE is the **difference in group means (DIGM)**.
- Label units such that $D_i = 1$ for $i \in \{1, 2, ..., m\}$ and $D_i = 0$ for $i \in \{m+1, m+2, ..., n\}$

$$\text{DIGM} \equiv \frac{1}{m} \sum_{i=1}^{m} Y_i - \frac{1}{n-m} \sum_{i=m+1}^{n} Y_i$$

- i.e. the difference in observed outcomes between treatment and control

Key question $\rightarrow$ Is the DIGM an **unbiased estimator** for the ATE?

DIGM - an Unbiased Estimator for the ATE?

Remember our toy example:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	5	1	5	2	3
2	2	1	2	1	1
3	0	0	1	0	1
4	1	0	1	1	0

$$\tau_{\textbf{ATE}} = \frac{3+1+1+0}{4} = 1.25$$

$$\textbf{DIGM} = \frac{5+2}{2} - \frac{0+1}{2} = 3$$

Is this true generally?

DIGM - an unbiased estimator for the ATE?

Let's redefine the DIGM using $\tau_i \equiv Y_{1i} - Y_{0i}$:

$$\begin{aligned} DIGM &= \frac{1}{m}\sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\ &= \frac{1}{m}\sum_{i=1}^{m}(Y_{0i}+\tau_i) - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\ &= \frac{1}{m}\sum_{i=1}^{m}\tau_i + \frac{1}{m}\sum_{i=1}^{m} Y_{0i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \end{aligned}$$

$$\begin{aligned} E[DIGM] &= E[\tau_i|D_i=1] + \{E[Y_{0i}|D_i=1] - E[Y_{0i}|D_i=0]\} \\ &= \tau_{\text{ATT}} + \text{selection bias} \end{aligned}$$

where τ_{ATT} is the average treatment effect for the treated group

DIGM - an unbiased estimator for the ATE?

Remember our example:

i	Y_i	D_i	Y_{1i}	Y_{0i}	τ_i
1	5	1	5	2	3
2	2	1	2	1	1
3	0	0	1	0	1
4	1	0	1	1	0

$$\tau_{\textbf{ATE}} = \frac{3+1+1+0}{4} = 1.25$$

$$\tau_{\textbf{ATT}} = \frac{3+1}{2} = 2$$

$$\textbf{Selection bias} = \frac{2+1}{2} - \frac{0+1}{2} = 1$$

$$\tau_{\textbf{ATT}} + \textbf{Bias} = 2+1 = 3 = \textbf{DIGM}$$

DIGM - an unbiased estimator for the ATE?

Implication

The DIGM is only an unbiased estimator of τ_{ATE} when

1. $\tau_{\text{ATT}} = \tau_{\text{ATE}}$ *and*
2. $E[Y_{0i}|D = 1] = E[Y_{0i}|D = 0]$ i.e. there is no selection bias

Does $E[Y_{0i}|D_i = 1]$ normally equal $E[Y_{0i}|D_i = 0]$?

- Does job training improve employment outcomes?
  - Participants are self-selected from a population of individuals in difficult labor situations
  - Post-training period earnings for participants would be higher than those for nonparticipants *even in the absence of the program*
  - **i.e.** $E[Y_0|D = 1] - E[Y_0|D = 0] > 0$
- Does asprin releive headache symptoms?
  - Those who take asprin self-select from the population of individuals who are suffering from headaches
  - Drug-takers would have worse headaches than non-drug-takers *in the absence of pain-relief*
  - **i.e.,** $E[Y_0|D = 1] - E[Y_0|D = 0] < 0$

No! $E[Y_{0i}|D_i = 1]$ is not normally equal to $E[Y_{0i}|D_i = 0]$

$E[Y_{0i}|D_i = 1]$ doesn't normally equal $E[Y_{0i}|D_i = 0]$

Implications

1. Selection into treatment is often associated with potential outcomes.
2. Selection bias can be positive or negative.
3. (In general) Do not believe causal arguments based on simple differences between groups!

"Solving" the Problem of Selection

- To assess (and correct for) selection bias, we need to know something about the potential outcomes that we do not observe.
- In order to infer these unobservable outcomes, we make assumptions about how certain units come to be "selected" for treatment.

Definition: Assignment Mechanism

The assignment mechanism is the procedure that determines which units are selected for treatment. Examples include:

- Random assignment
- Selection on observable characteristics
- Selection on unobservable characteristics

Assignment Mechanisms

We will link these assignment mechanisms to specific empirical approaches, discussing what they tell us about causal relationships.

- Random assignment
  - Randomized experiments (week 2)
- Selection on observables
  - Subclassification, matching (week 3)
  - Regression (week 4)
- Selection on unobservables
  - Difference-in-differences; synthetic control (weeks 5 and 6)
  - Instrumental variables (weeks 7 and 8)
  - Regression discontinuity designs (week 9)

Conclusion

- **Potential outcomes**: Causality is defined by potential outcomes, not by observed outcomes
- **FPOCI**: We only ever observe one potential outcome per unit. Then how can we find $\tau_i = Y_{1i} - Y_{0i}$?
- **Inference**:
  - Statistical inference $\rightarrow$ learning about the population from a sample
  - Causal inference $\rightarrow$ learning about the population of potential outcomes from the observed outcomes
- **Selection bias**: The difference in means is only an unbiased estimator for the ATE when there is no selection bias