Causal inference

Andy Eggers

Oxford DPIR

HT 2018

Getting started

Causal inference is the process by which one can use data to make claims about causal relationships (Hidalgo and Sehkon 2012).

Getting started

Causal inference is the process by which one can use data to make claims about causal relationships (Hidalgo and Sehkon 2012).

Mostly this means measuring the effect of a **treatment** on some **outcome**:

- What is the effect of peace-keeping missions on peace?
- What is the effect of class size on test scores?
- What is the effect of canvassing on vote intention?

Getting started

Causal inference is the process by which one can use data to make claims about causal relationships (Hidalgo and Sehkon 2012).

Mostly this means measuring the effect of a **treatment** on some **outcome**:

- What is the effect of peace-keeping missions on peace?
- What is the effect of class size on test scores?
- What is the effect of canvassing on vote intention?

Measuring causal relationships also important for answering **why** questions:

- explaining variation in outcomes (e.g. why are some countries so poor?)
- explaining relationships between variables (e.g. why do more nationalistic countries redistribute less?)

What's new?

Didn't we already do this? Isn't "measuring the effect of D on Y" the purpose of regression?

What's new?

Didn't we already do this? Isn't "measuring the effect of D on Y" the purpose of regression?

Yes, but regression also good for description. Next week: regression from causal inference perspective.

What's new?

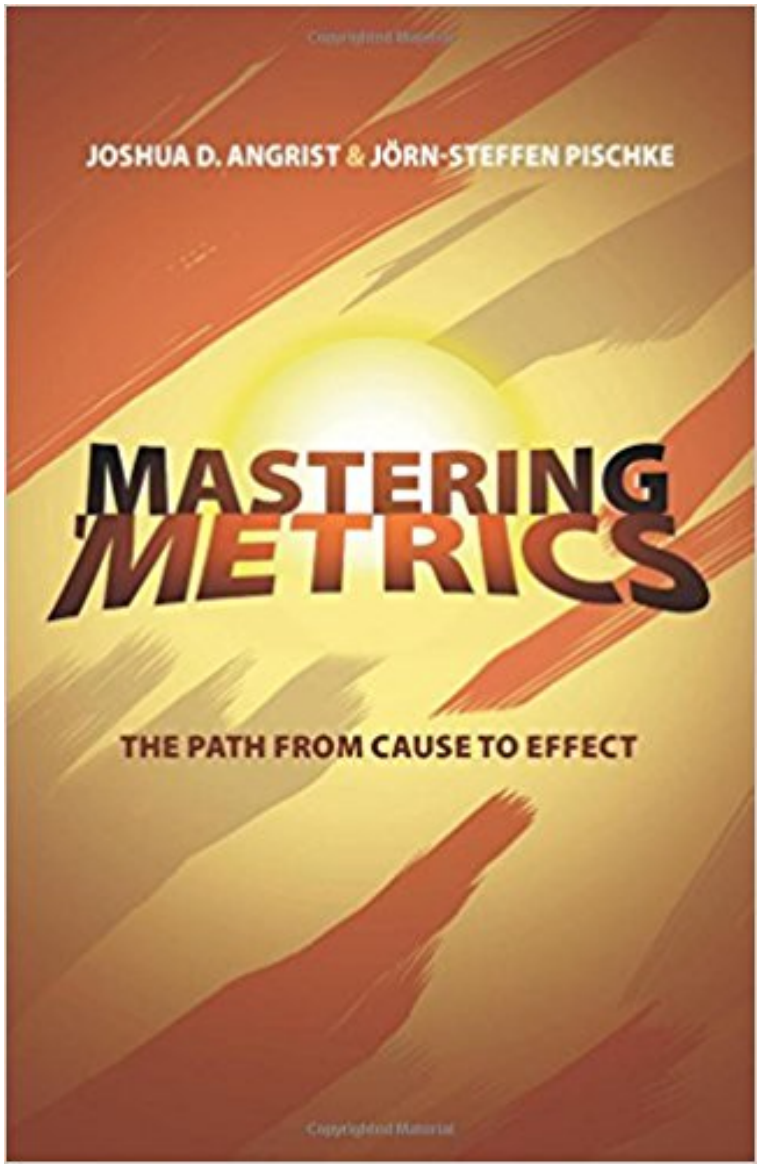
Didn't we already do this? Isn't "measuring the effect of D on Y" the purpose of regression?

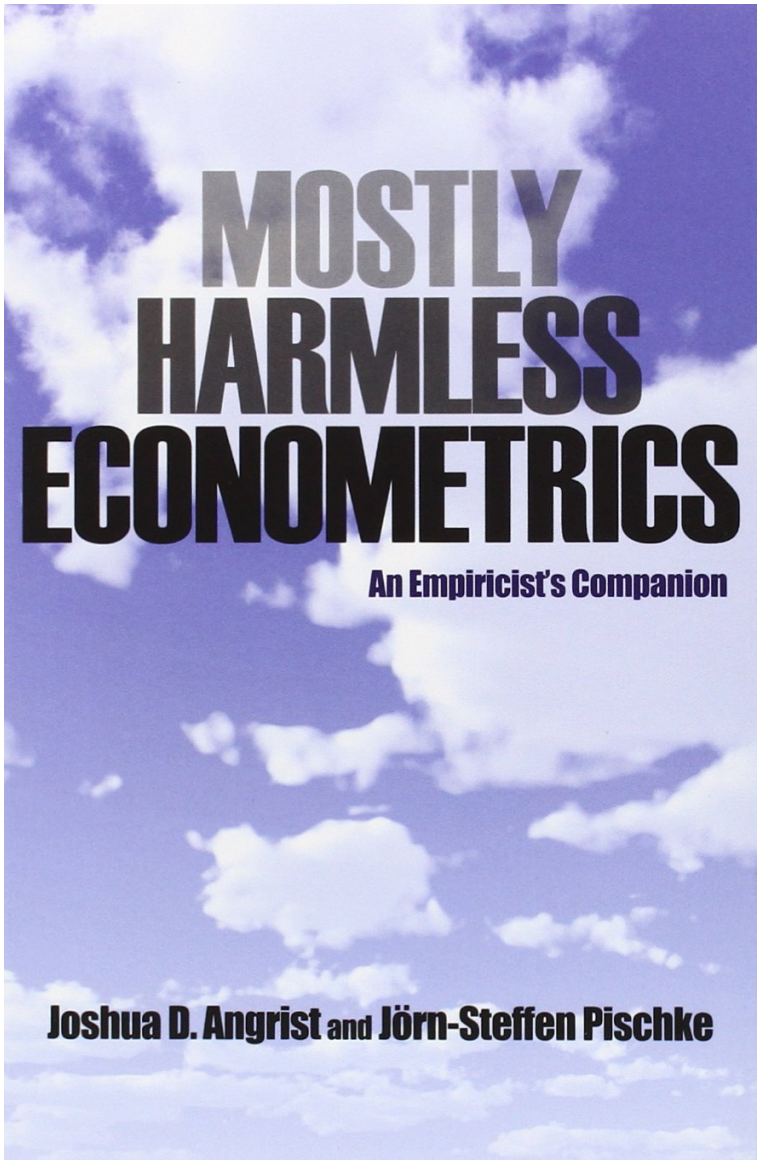
Yes, but regression also good for description. Next week: regression from causal inference perspective.

Plan by week:

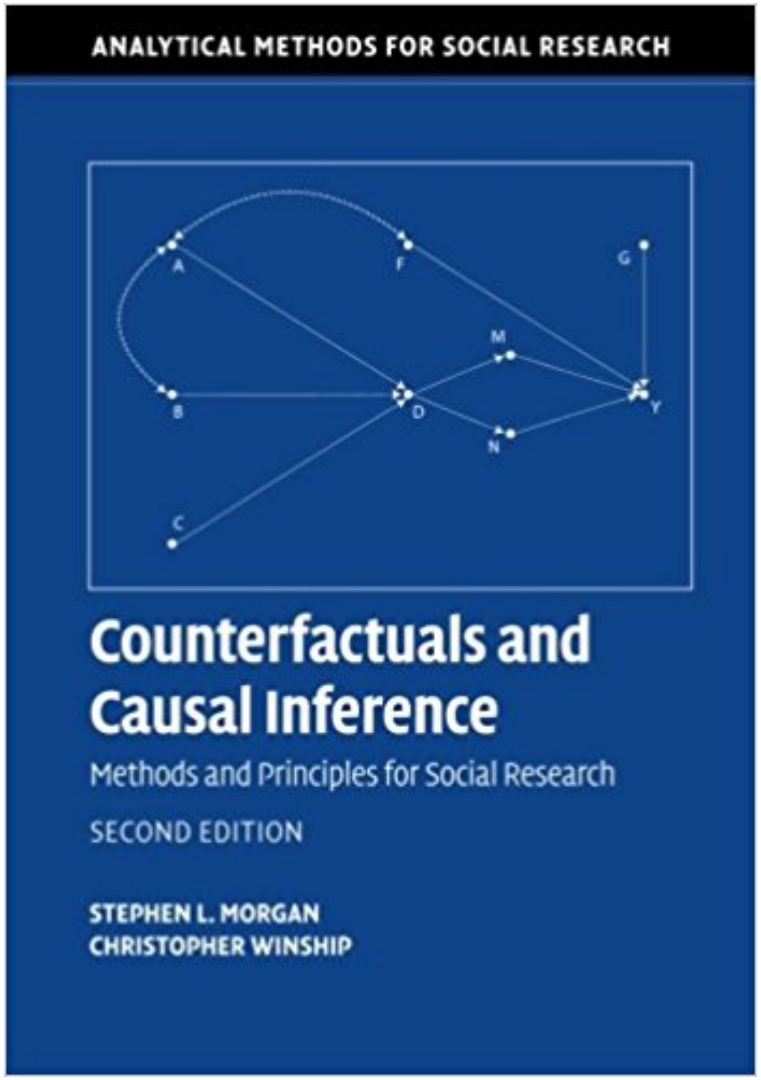
1. Potential outcomes framework and randomized experiments
2. Covariate adjustment 1 (Regression, matching, sub-classification)
3. Covariate adjustment 2 (Regression, matching, sub-classification)
4. Instrumental variables
5. Regression discontinuity design (RDD)
6. Differences-in-differences
7. Panel fixed-effects analysis
8. Treatment effect heterogeneity, mechanisms, and mediation

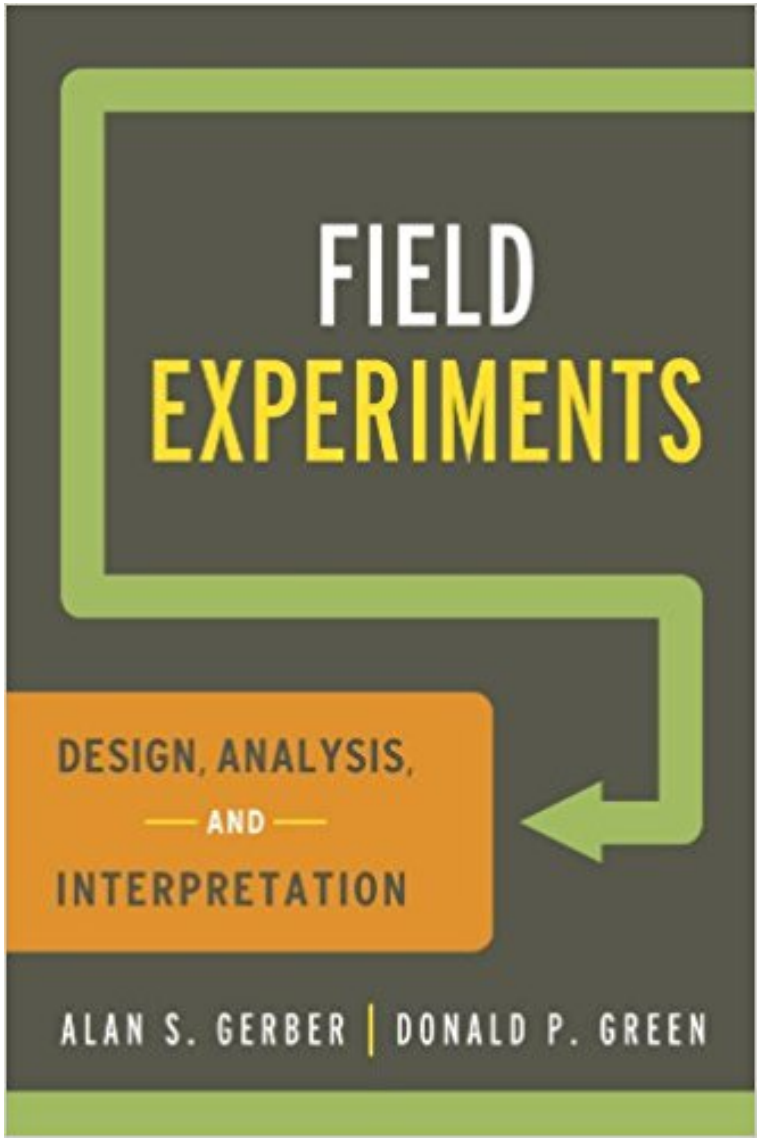
Textbooks


JOSHUA D. ANGRIST & JÖRN-STEFFEN PISCHKE
MASTERING 'METRICS
THE PATH FROM CAUSE TO EFFECT


MOSTLY HARMLESS ECONOMETRICS
An Empiricist's Companion
Joshua D. Angrist and Jörn-Steffen Pischke

Textbooks (2)


ANALYTICAL METHODS FOR SOCIAL RESEARCH
A
F
G
M
B
D
Y
N
C
Counterfactuals and Causal Inference
Methods and Principles for Social Research
SECOND EDITION
STEPHEN L. MORGAN
CHRISTOPHER WINSHIP


FIELD EXPERIMENTS
DESIGN, ANALYSIS, AND INTERPRETATION
ALAN S. GERBER | DONALD P. GREEN

Assessment

Problem sets (each 20% of final mark, total of 60%):

- PS 1: Distributed week 1, due noon on Thursday of Week 3.
- PS 2: Distributed week 3, due noon on Thursday of Week 6.
- PS 3: Distributed week 6, due noon on Thursday of Week 9.

Take-home exam (40% of final mark): Distributed by Friday of Week 8, due noon on Friday of week 0 of Trinity Term.

Weekly lab sessions

Spyros Kosmidis

Vuk Vukovic

Schedule:

- Group 1: Thursday 1500-1630
- Group 2: Thursday 1630-1800

Basics (in plain-ish language)

What does it mean for some **causal factor** (e.g. time studying) to **affect** some **outcome** (e.g. final mark) for an **individual** (e.g. you)?

Basics (in plain-ish language)

What does it mean for some **causal factor** (e.g. time studying) to **affect** some **outcome** (e.g. final mark) for an **individual** (e.g. you)?

In the counterfactual approach to causation, it means that the outcome for this individual would be different for different hypothetical values of the causal factor.

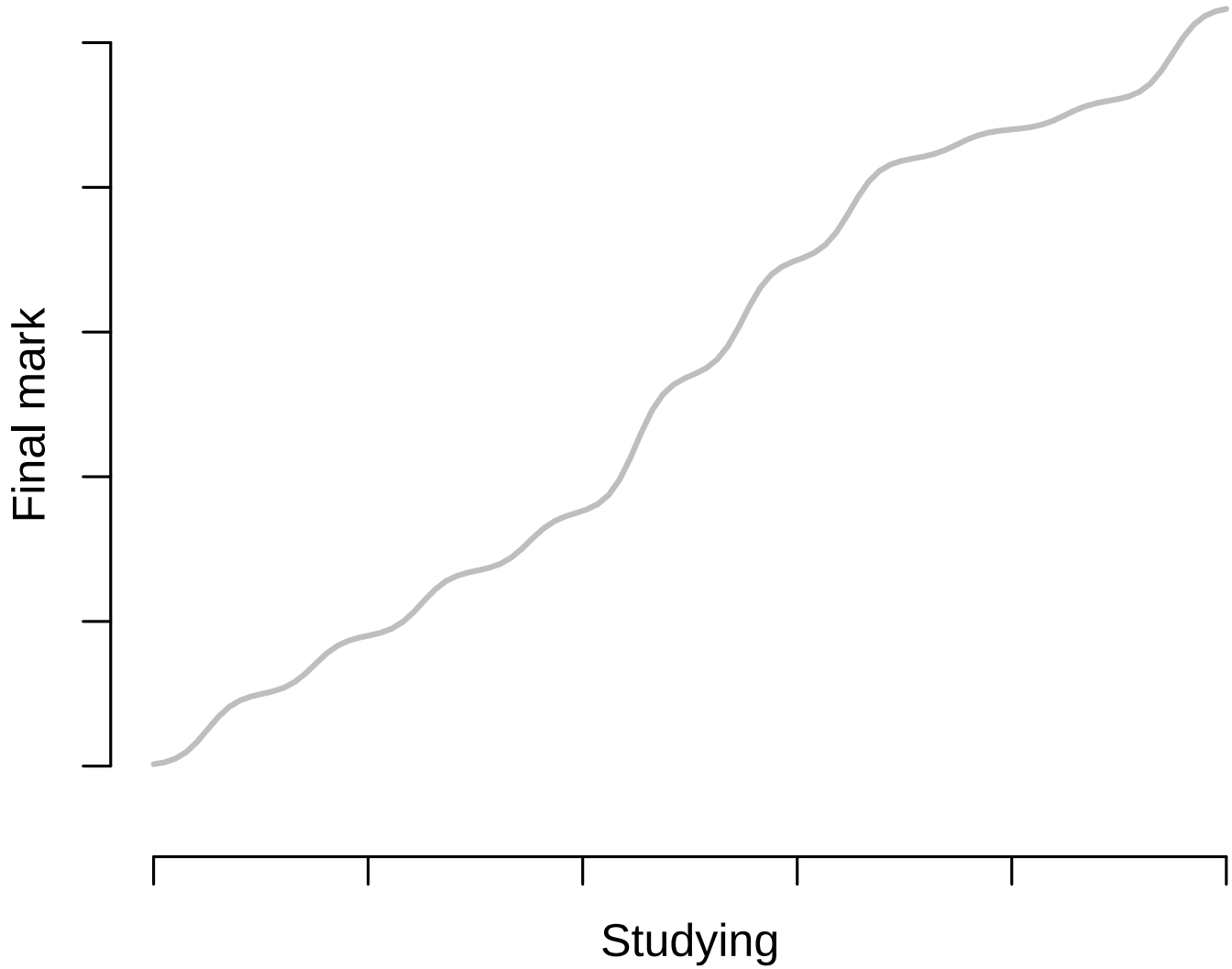
Basics (in plain-ish language)

What does it mean for some **causal factor** (e.g. time studying) to **affect** some **outcome** (e.g. final mark) for an **individual** (e.g. you)?

In the counterfactual approach to causation, it means that the outcome for this individual would be different for different hypothetical values of the causal factor.

We call these hypothetical outcomes (one for each hypothetical value of the causal factor) **potential outcomes**.

Potential outcomes (continuous case)



Effects from potential outcomes

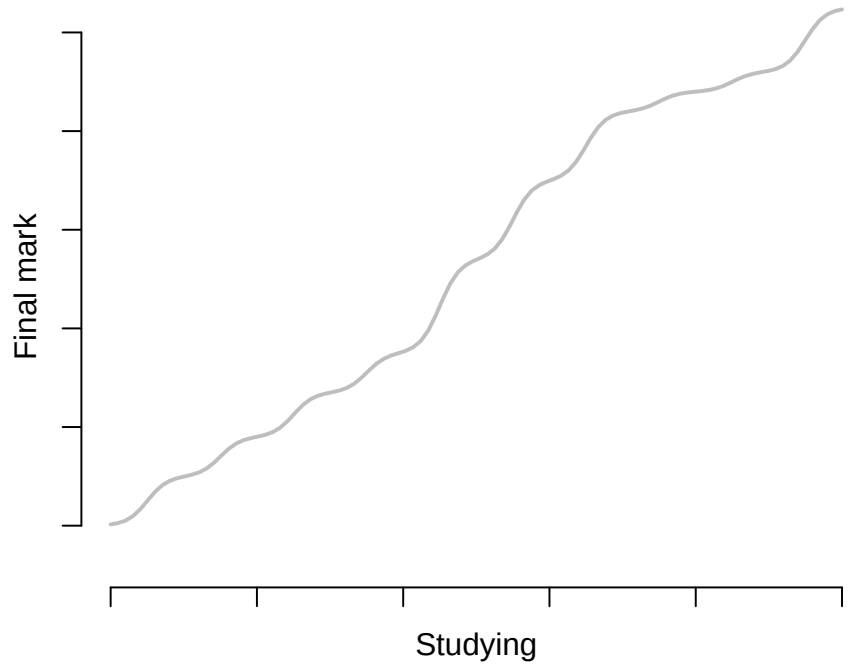
In the potential outcomes framework,
(causal) effects are defined in terms
of these potential outcomes.

Effects from potential outcomes

In the potential outcomes framework, (causal) effects are defined in terms of these potential outcomes.

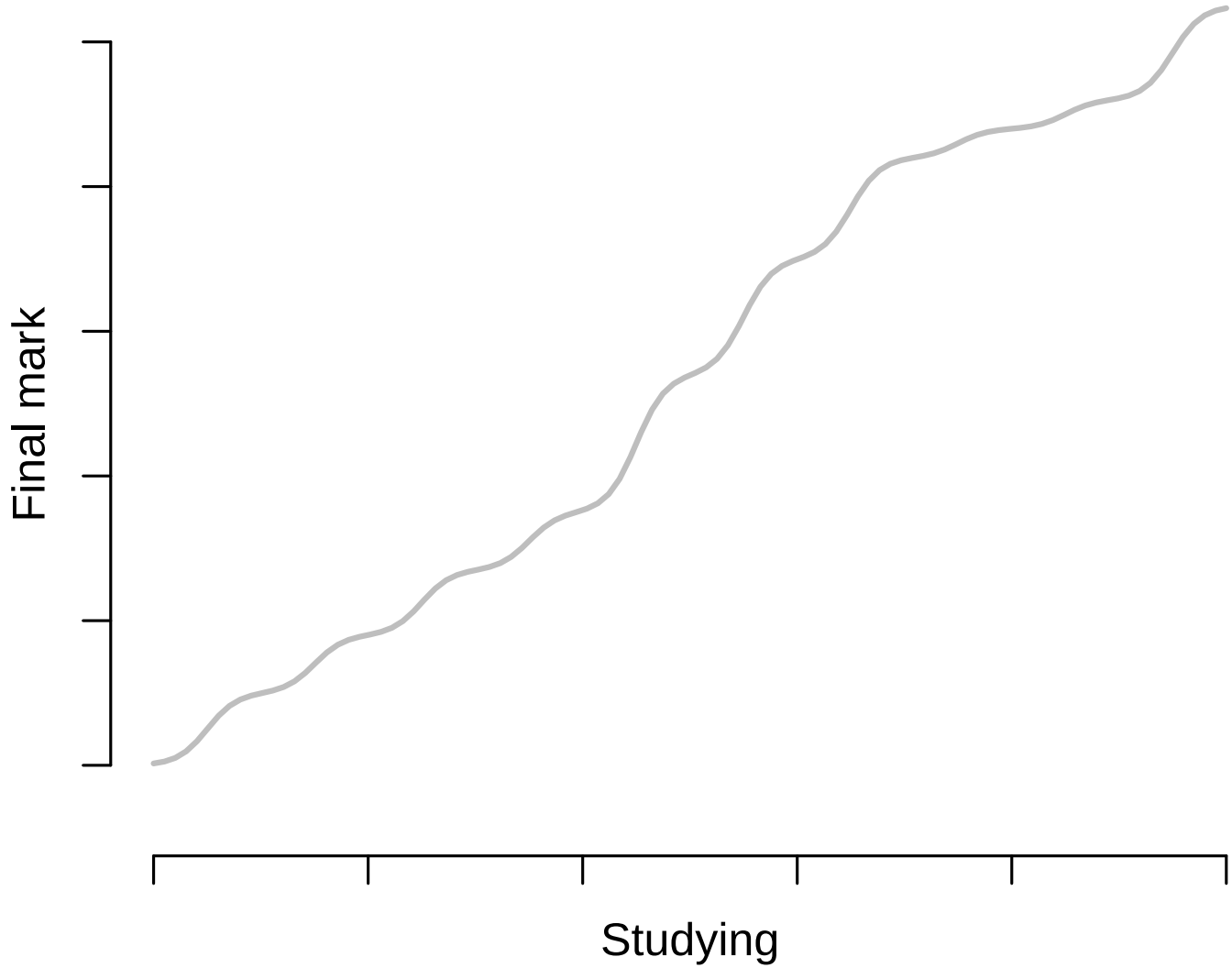
If we knew all potential outcomes for an individual, we could also know

- the effect of studying 100 hours vs. 1 hour on final mark
- the average effect of studying an additional hour on final mark
- etc.



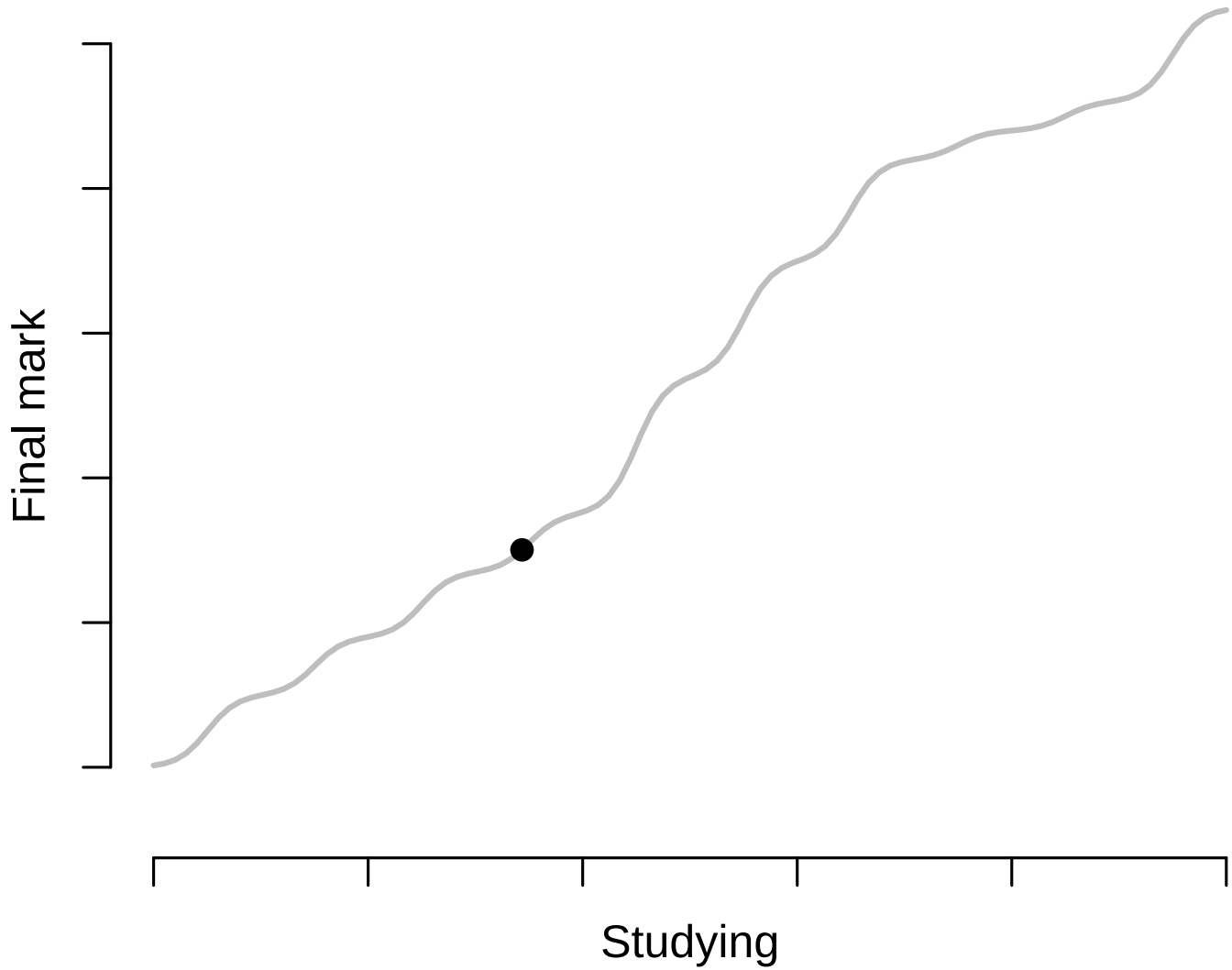
Fundamental problem of causal inference

We only observe **one** potential outcome per unit.



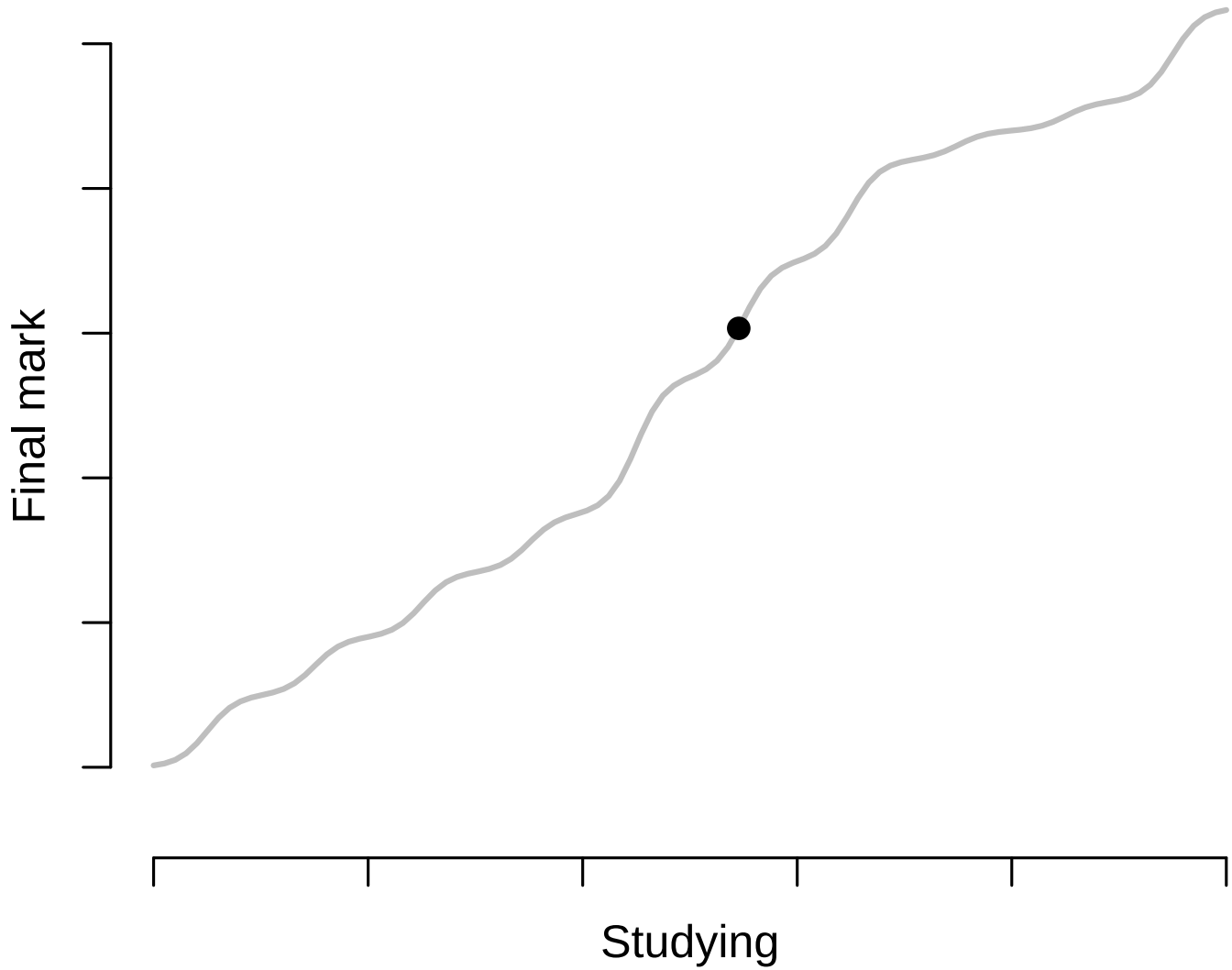
Fundamental problem of causal inference

We only observe **one** potential outcome per unit.



Fundamental problem of causal inference

We only observe **one** potential outcome per unit.



Comparisons over time?

What about observing the same individual at various points in time?

Comparisons over time?

What about observing the same individual at various points in time?

This can be very useful.

Comparisons over time?

What about observing the same individual at various points in time?

This can be very useful.

Does it work for measuring the effect of

- state of light switch on brightness of room?
- taking an aspirin on headache?
- studying on final mark?
- GDP on democracy?

Comparisons over time?

What about observing the same individual at various points in time?

This can be very useful.

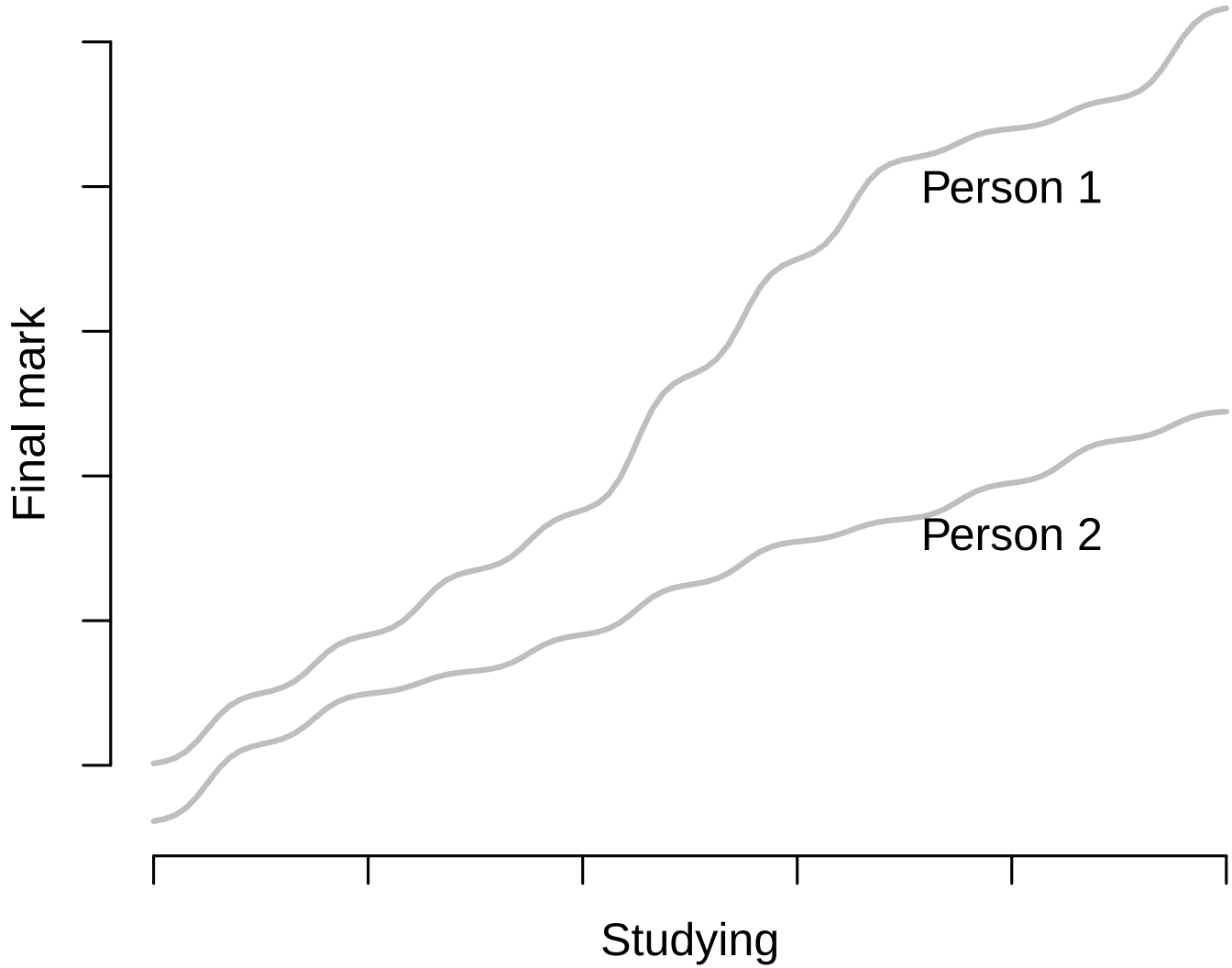
Does it work for measuring the effect of

- state of light switch on brightness of room?
- taking an aspirin on headache?
- studying on final mark?
- GDP on democracy?

Depends on whether (a) other things change over time and (b) outcomes depend on past treatments

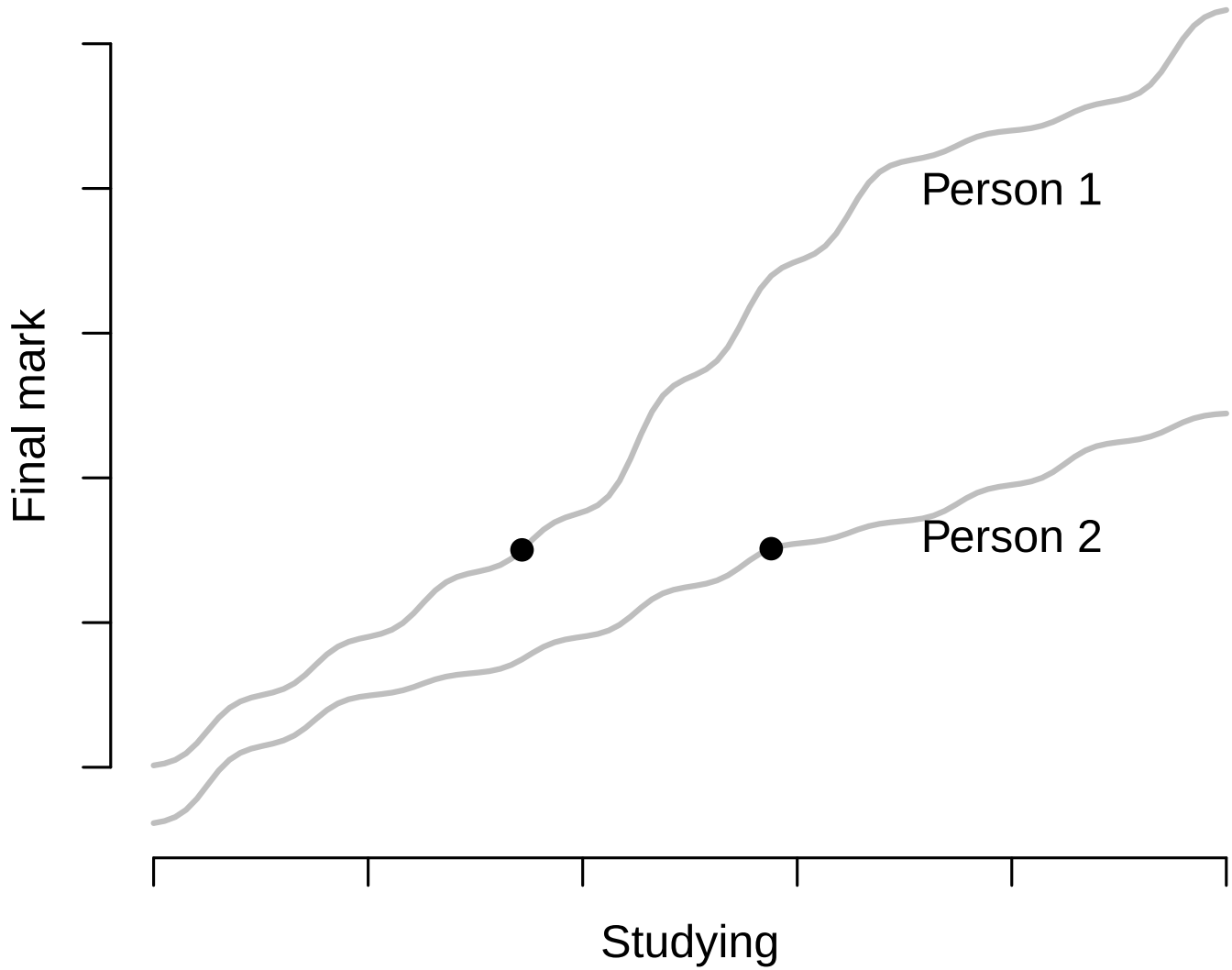
Comparisons across units

The only other option is to compare across units. But be careful!



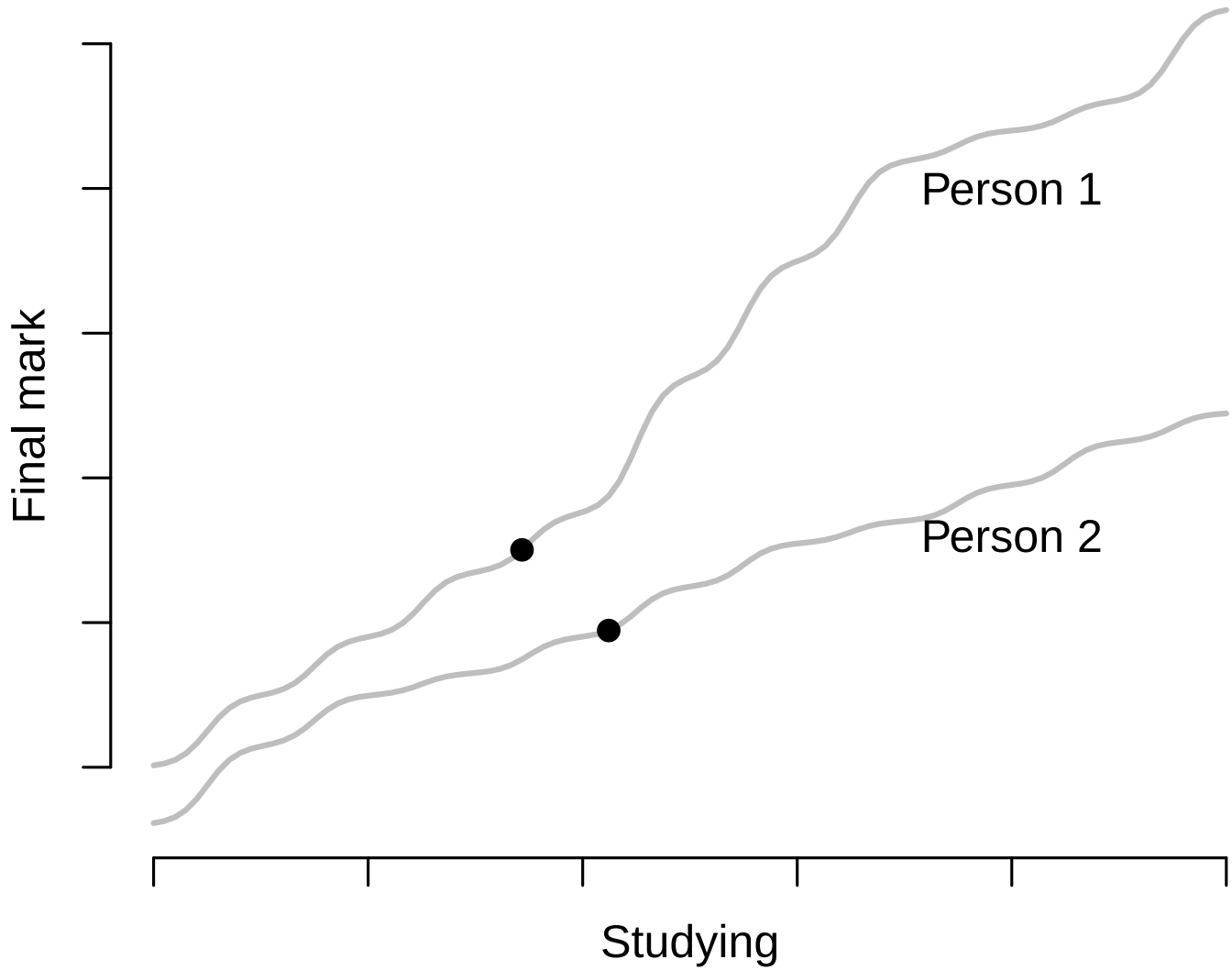
Comparisons across units

The only other option is to compare across units. But be careful!



Comparisons across units

The only other option is to compare across units. But be careful!



Comparisons across units: what's the problem?

1. **Heterogeneity**: Different units have different potential outcomes

Comparisons across units: what's the problem?

1. **Heterogeneity**: Different units have different potential outcomes

2. **Confounding/endogeneity**: Typically, observed treatment is related to the potential outcomes (e.g. smarter students study more/less, students who benefit more from studying study more/less)

Comparisons across units: what's the problem?

1. **Heterogeneity**: Different units have different potential outcomes

2. **Confounding/endogeneity**: Typically, observed treatment is related to the potential outcomes (e.g. smarter students study more/less, students who benefit more from studying study more/less)

Causal inference is partly about managing **heterogeneity**, mostly about addressing **confounding**.

Causal inference and statistical inference

To **infer** is to draw conclusions about things unseen from things seen.

Causal inference and statistical inference

To **infer** is to draw conclusions about things unseen from things seen.

In **statistical inference** we draw conclusions about the **population** based on the **sample**.

Causal inference and statistical inference

To **infer** is to draw conclusions about things unseen from things seen.

In **statistical inference** we draw conclusions about the **population** based on the **sample**.

e.g. how likely is it that two variables are uncorrelated in the population, given the correlation I observe in my sample?

Causal inference and statistical inference

To **infer** is to draw conclusions about things unseen from things seen.

In **statistical inference** we draw conclusions about the **population** based on the **sample**.

e.g. how likely is it that two variables are uncorrelated in the population, given the correlation I observe in my sample?

In **causal inference** we draw conclusions about the **full set of potential outcomes** based on the **observed potential outcomes**.

Causal inference and statistical inference

To **infer** is to draw conclusions about things unseen from things seen.

In **statistical inference** we draw conclusions about the **population** based on the **sample**.

e.g. how likely is it that two variables are uncorrelated in the population, given the correlation I observe in my sample?

In **causal inference** we draw conclusions about the **full set of potential outcomes** based on the **observed potential outcomes**.

In practice, we are usually doing both: using a sample to draw conclusions about the full set of potential outcomes in the population.

Potential outcomes framework: activity 1!

Think of a treatment and an outcome it might affect for some unit.
(Examples from political science, nutrition, fiction, etc. all welcome.)

1. Make a figure showing how the potential outcomes might depend on the treatment for one unit. Identify the unit.
2. Which potential outcome do we observe for this unit?
3. Repeat (1) and (2) for additional unit(s).

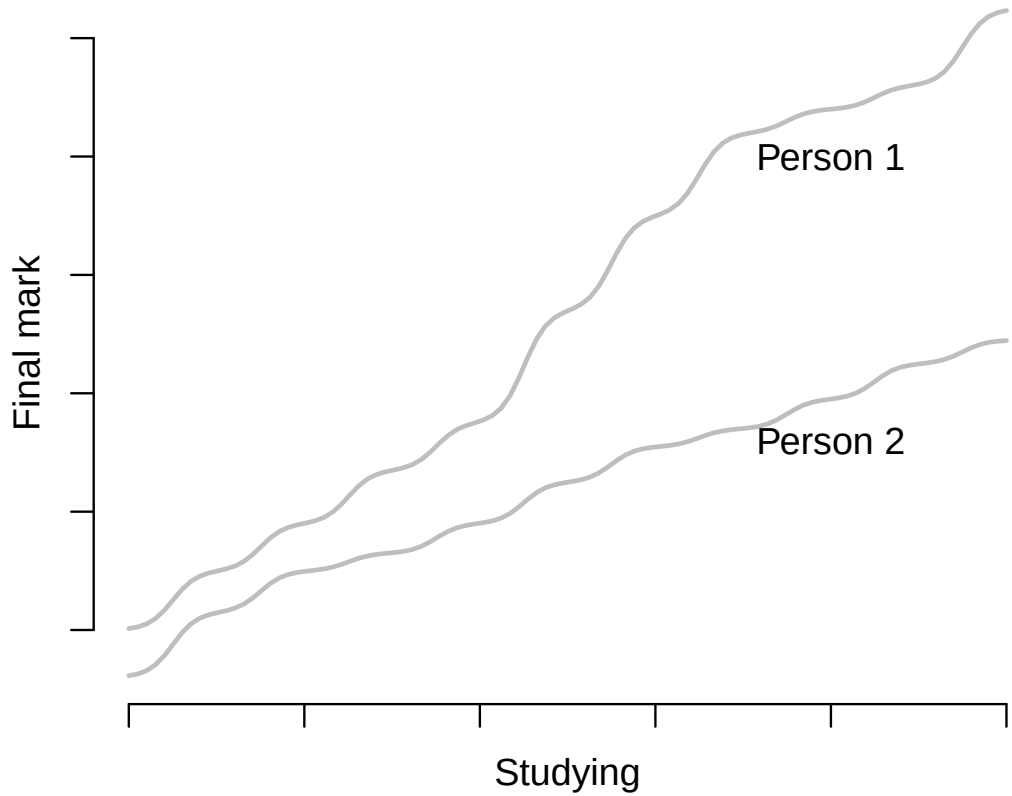
A note on the standard "no interference" assumption

We will typically assume "no interference" (SUTVA): my final mark does not depend on how much you study (Gerber & Green, 43-44).

A note on the standard "no interference" assumption

We will typically assume "no interference" (SUTVA): my final mark does not depend on how much you study (Gerber & Green, 43-44).

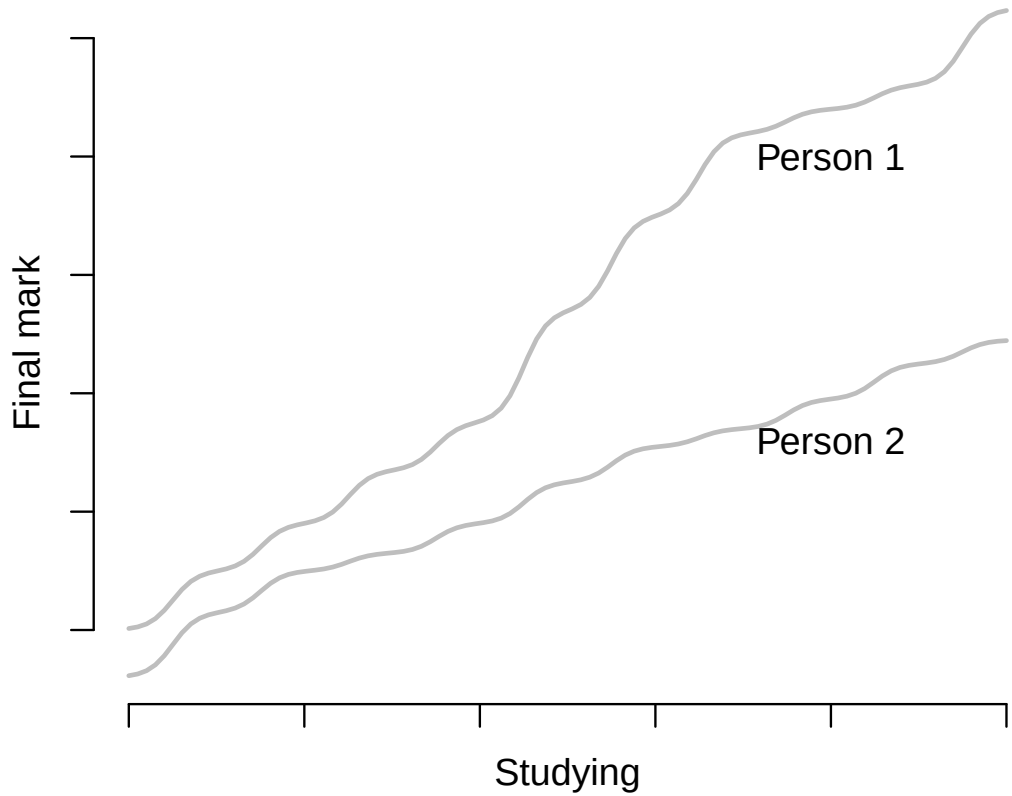
Does this figure make that assumption?



A note on the standard "no interference" assumption

We will typically assume "no interference" (SUTVA): my final mark does not depend on how much you study (Gerber & Green, 43-44).

Does this figure make that assumption?



When might this assumption be invalid?

The potential outcomes framework: informal introduction

The potential outcomes framework: more formally

Randomized experiments

Potential outcomes framework: notation

Warning

Every author uses different notation. (Angrist & Pischke, Gerber & Green, Morgan & Winship, Imbens & Rubin, . . .).

I will use Angrist & Pischke's notation from *Mastering 'Metrics* (**MM**).

Potential outcomes framework: notation

Warning

Every author uses different notation. (Angrist & Pischke, Gerber & Green, Morgan & Winship, Imbens & Rubin, . . .).
I will use Angrist & Pischke's notation from *Mastering 'Metrics* (**MM**).

Let D_i denote the **treatment** ("causal variable") for unit i.

Potential outcomes framework: notation

Warning

Every author uses different notation. (Angrist & Pischke, Gerber & Green, Morgan & Winship, Imbens & Rubin, . . .).
I will use Angrist & Pischke's notation from *Mastering 'Metrics* (**MM**).

Let D_i denote the **treatment** ("causal variable") for unit i.

In general D_i may take many values, but we focus on the binary case ("treatment" and "control"): $D_i \in \{0, 1\}$.

Potential outcomes framework: notation

Warning

Every author uses different notation. (Angrist & Pischke, Gerber & Green, Morgan & Winship, Imbens & Rubin, . . .).
I will use Angrist & Pischke's notation from *Mastering 'Metrics* (**MM**).

Let D_i denote the **treatment** ("causal variable") for unit i.

In general D_i may take many values, but we focus on the binary case ("treatment" and "control"): $D_i \in \{0, 1\}$.

Let Y_i denote the observed **outcome** for unit i.

Potential outcomes framework: notation

Warning

Every author uses different notation. (Angrist & Pischke, Gerber & Green, Morgan & Winship, Imbens & Rubin, . . .).
I will use Angrist & Pischke's notation from *Mastering 'Metrics* (**MM**).

Let D_i denote the **treatment** ("causal variable") for unit i.

In general D_i may take many values, but we focus on the binary case ("treatment" and "control"): $D_i \in \{0, 1\}$.

Let Y_i denote the observed **outcome** for unit i.

Potential outcomes: Y_{0i}, Y_{1i}.

Potential outcomes framework: notation

Warning

Every author uses different notation. (Angrist & Pischke, Gerber & Green, Morgan & Winship, Imbens & Rubin, . . .).
I will use Angrist & Pischke's notation from *Mastering 'Metrics* (**MM**).

Let D_i denote the **treatment** ("causal variable") for unit i.

In general D_i may take many values, but we focus on the binary case ("treatment" and "control"): $D_i \in \{0, 1\}$.

Let Y_i denote the observed **outcome** for unit i.

Potential outcomes: Y_{0i}, Y_{1i}.

Effect of D **on** Y for i: $\tau_i = Y_{1i} - Y_{0i}$ (treatment effect, causal effect).

Individual treatment effects: no FPOCI

Imagine a population of 4 units in which both potential outcomes are observed:

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

Individual treatment effects: no FPOCI

Imagine a population of 4 units in which both potential outcomes are observed:

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

Individual treatment effects: no FPOCI

Imagine a population of 4 units in which both potential outcomes are observed:

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	
4	1	0	1	1	

Individual treatment effects: no FPOCI

Imagine a population of 4 units in which both potential outcomes are observed:

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	

Individual treatment effects: no FPOCI

Imagine a population of 4 units in which both potential outcomes are observed:

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

Individual treatment effects: no FPOCI

Imagine a population of 4 units in which both potential outcomes are observed:

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

Could average these to get the **average treatment effect (ATE)**:

Individual treatment effects: no FPOCI

Imagine a population of 4 units in which both potential outcomes are observed:

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

Could average these to get the **average treatment effect (ATE)**: 1

Fundamental problem of causal inference

What you actually observe:

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

Fundamental problem of causal inference

What you actually observe:

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

Fundamental problem of causal inference

What you actually observe:

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

Fundamental problem of causal inference

What you actually observe:

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

Fundamental problem of causal inference

What you actually observe:

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

Fundamental problem of causal inference

What you actually observe:

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

Fundamental problem of causal inference

What you actually observe:

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

Fundamental problem of causal inference

What you actually observe:

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			

Fundamental problem of causal inference

What you actually observe:

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			

Fundamental problem of causal inference

What you actually observe:

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			

Fundamental problem of causal inference

What you actually observe:

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	

Fundamental problem of causal inference

What you actually observe:

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	

Fundamental problem of causal inference

What you actually observe:

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	?

Fundamental problem of causal inference

What you actually observe:

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	?

Average treatment effect (ATE) $=$?

Fundamental problem of causal inference

What you actually observe:

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	?

$$\text{Average treatment effect (ATE)} = ?$$

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

Fundamental problem of causal inference

What you actually observe:

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	?

$$\text{Average treatment effect (ATE)} = ?$$

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

Quick check: Name some unobservable potential outcomes.

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure
- **Random variable**, e.g. X: a variable whose outcomes are results of a random/stochastic phenomenon

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure
- **Random variable**, e.g. X: a variable whose outcomes are results of a random/stochastic phenomenon
  - number on a six-sided dice (equal prob. of 1, 2, . . . , 6)

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure
- **Random variable**, e.g. X: a variable whose outcomes are results of a random/stochastic phenomenon
  - number on a six-sided dice (equal prob. of 1, 2, ..., 6)
  - treatment status of unit i in randomized experiment (e.g. $\Pr(D_i = 1) = .4$)

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure
- **Random variable**, e.g. X: a variable whose outcomes are results of a random/stochastic phenomenon
  - number on a six-sided dice (equal prob. of 1, 2, ..., 6)
  - treatment status of unit i in randomized experiment (e.g. $\Pr(D_i = 1) = .4$)
  - Y_{1i} for a unit randomly selected from the population

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure
- **Random variable**, e.g. X: a variable whose outcomes are results of a random/stochastic phenomenon
  - number on a six-sided dice (equal prob. of 1, 2, ..., 6)
  - treatment status of unit i in randomized experiment (e.g. $\Pr(D_i = 1) = .4$)
  - Y_{1i} for a unit randomly selected from the population
- $E[X]$ denotes **expectation** (or **mean**) of a random variable X

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure
- **Random variable**, e.g. X: a variable whose outcomes are results of a random/stochastic phenomenon
  - number on a six-sided dice (equal prob. of 1, 2, ..., 6)
  - treatment status of unit i in randomized experiment (e.g. $\Pr(D_i = 1) = .4$)
  - Y_{1i} for a unit randomly selected from the population
- $E[X]$ denotes **expectation** (or **mean**) of a random variable X
  - for a six-sided dice, $E[X] = 3.5$

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure
- **Random variable**, e.g. X: a variable whose outcomes are results of a random/stochastic phenomenon
  - number on a six-sided dice (equal prob. of 1, 2, . . . , 6)
  - treatment status of unit i in randomized experiment (e.g. $\Pr(D_i = 1) = .4$)
  - Y_{1i} for a unit randomly selected from the population
- $E[X]$ denotes **expectation** (or **mean**) of a random variable X
  - for a six-sided dice, $E[X] = 3.5$
  - given $\Pr(D_i = 1) = .4$, $E[D_i] = .4$

Statistics concepts and notation

- A **population** contains all units of interest; a **sample** is a subset of these
- **Variable** (e.g. Y_{1i}, X): a numerical measure
- **Random variable**, e.g. X: a variable whose outcomes are results of a random/stochastic phenomenon
  - number on a six-sided dice (equal prob. of 1, 2, . . . , 6)
  - treatment status of unit i in randomized experiment (e.g. $\Pr(D_i = 1) = .4$)
  - Y_{1i} for a unit randomly selected from the population
- $E[X]$ denotes **expectation** (or **mean**) of a random variable X
  - for a six-sided dice, $E[X] = 3.5$
  - given $\Pr(D_i = 1) = .4$, $E[D_i] = .4$
  - $E[Y_{1i}]$ is the average of Y_{1i} in the population

Average treatment effect

Most important **causal estimand** is the **average treatment effect (ATE)**.

Average treatment effect

Most important **causal estimand** is the **average treatment effect (ATE)**.

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

Average treatment effect

Most important **causal estimand** is the **average treatment effect (ATE)**.

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

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

Average treatment effect

Most important **causal estimand** is the **average treatment effect (ATE)**.

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

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

Average treatment effect

Most important **causal estimand** is the **average treatment effect (ATE)**.

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

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

Average treatment effect

Most important **causal estimand** is the **average treatment effect (ATE)**.

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

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

(assuming population of N units).

A bit more statistics and notation

- $E[Y_{1i}|D_i = 1]$ denotes the **conditional expectation** of random variable Y_{1i} where $D_i = 1$, i.e. $E[Y_{1i}]$ once we restrict the population to units where $D_i = 1$

A bit more statistics and notation

- $E[Y_{1i}|D_i = 1]$ denotes the **conditional expectation** of random variable Y_{1i} where $D_i = 1$, i.e. $E[Y_{1i}]$ once we restrict the population to units where $D_i = 1$
- A **parameter** is a fixed feature of a population (e.g. $E[Y_{1i}]$)

A bit more statistics and notation

- $E[Y_{1i}|D_i = 1]$ denotes the **conditional expectation** of random variable Y_{1i} where $D_i = 1$, i.e. $E[Y_{1i}]$ once we restrict the population to units where $D_i = 1$
- A **parameter** is a fixed feature of a population (e.g. $E[Y_{1i}]$)
- A **sample statistic** is a random variable whose value depends on the sample, e.g. the **sample mean**

A bit more statistics and notation

- $E[Y_{1i}|D_i = 1]$ denotes the **conditional expectation** of random variable Y_{1i} where $D_i = 1$, i.e. $E[Y_{1i}]$ once we restrict the population to units where $D_i = 1$
- A **parameter** is a fixed feature of a population (e.g. $E[Y_{1i}]$)
- A **sample statistic** is a random variable whose value depends on the sample, e.g. the **sample mean**
- An **estimator** is any function of sample data used to estimate a parameter

A bit more statistics and notation

- $E[Y_{1i}|D_i = 1]$ denotes the **conditional expectation** of random variable Y_{1i} where $D_i = 1$, i.e. $E[Y_{1i}]$ once we restrict the population to units where $D_i = 1$
- A **parameter** is a fixed feature of a population (e.g. $E[Y_{1i}]$)
- A **sample statistic** is a random variable whose value depends on the sample, e.g. the **sample mean**
- An **estimator** is any function of sample data used to estimate a parameter
- A sample statistic is an **unbiased estimator** of a parameter if its expectation (across samples) equals the parameter, e.g.

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = E[X_i]$$

A bit more statistics and notation

- $E[Y_{1i}|D_i = 1]$ denotes the **conditional expectation** of random variable Y_{1i} where $D_i = 1$, i.e. $E[Y_{1i}]$ once we restrict the population to units where $D_i = 1$
- A **parameter** is a fixed feature of a population (e.g. $E[Y_{1i}]$)
- A **sample statistic** is a random variable whose value depends on the sample, e.g. the **sample mean**
- An **estimator** is any function of sample data used to estimate a parameter
- A sample statistic is an **unbiased estimator** of a parameter if its expectation (across samples) equals the parameter, e.g.

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = E[X_i]$$

i.e. sample mean is unbiased estimator of mean/expectation

Difference in group means (naive estimate)

Recall that average treatment effect is

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

Difference in group means (naive estimate)

Recall that average treatment effect is

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

Given a sample, one possible estimator of the ATE is **difference in group means (DIGM)**, i.e. the difference in observed outcomes between treatment group and control group.

Difference in group means (naive estimate)

Recall that average treatment effect is

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

Given a sample, one possible estimator of the ATE is **difference in group means (DIGM)**, i.e. the difference in observed outcomes between treatment group and control group.

More formally, label units such that $D_i = 1$ for $i \in \{1, 2, \ldots, m\}$ and $D_i = 0$ for $i \in \{m+1, m+2, \ldots, n\}$.

Difference in group means (naive estimate)

Recall that average treatment effect is

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

Given a sample, one possible estimator of the ATE is **difference in group means (DIGM)**, i.e. the difference in observed outcomes between treatment group and control group.

More formally, label units such that $D_i = 1$ for $i \in \{1, 2, \ldots, m\}$ and $D_i = 0$ for $i \in \{m+1, m+2, \ldots, n\}$.

Then **difference in group means** is

$$\text{DIGM} \equiv \frac{1}{m} \sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m} \sum_{i=m+1}^{n} Y_{0i}.$$

Difference in group means (naive estimate)

Recall that average treatment effect is

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

Given a sample, one possible estimator of the ATE is **difference in group means (DIGM)**, i.e. the difference in observed outcomes between treatment group and control group.

More formally, label units such that $D_i = 1$ for $i \in \{1, 2, \ldots, m\}$ and $D_i = 0$ for $i \in \{m+1, m+2, \ldots, n\}$.

Then **difference in group means** is

$$\text{DIGM} \equiv \frac{1}{m} \sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m} \sum_{i=m+1}^{n} Y_{0i}.$$

By the unbiasedness of the sample mean, we know that

$$E\left[\text{DIGM}\right] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

Difference in group means (naive estimate)

Recall that average treatment effect is

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

Given a sample, one possible estimator of the ATE is **difference in group means (DIGM)**, i.e. the difference in observed outcomes between treatment group and control group.

More formally, label units such that $D_i = 1$ for $i \in \{1, 2, \ldots, m\}$ and $D_i = 0$ for $i \in \{m+1, m+2, \ldots, n\}$.

Then **difference in group means** is

$$\text{DIGM} \equiv \frac{1}{m} \sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m} \sum_{i=m+1}^{n} Y_{0i}.$$

By the unbiasedness of the sample mean, we know that

$$E\big[\text{DIGM}\big] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0]$$

which can be written $E[Y_i|D_i = 1] - E[Y_i|D_i = 0]$, where Y_i is Y_{1i} if $D_i = 1$ and otherwise Y_{0i}.

Difference in group means (example)

Recall our toy example:

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

Difference in group means (example)

Recall our toy example:

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	?

Difference in group means (example)

Recall our toy example:

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	?

ATE was actually 1 (generally unknown), but DIGM is

Difference in group means (example)

Recall our toy example:

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	?

ATE was actually 1 (generally unknown), but DIGM is $\frac{3+1}{2} - \frac{0+1}{2} =$

Difference in group means (example)

Recall our toy example:

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	?

ATE was actually 1 (generally unknown), but DIGM is $\frac{3+1}{2} - \frac{0+1}{2} = 1.5$.

Difference in group means (example)

Recall our toy example:

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	?

ATE was actually 1 (generally unknown), but DIGM is $\frac{3+1}{2} - \frac{0+1}{2} = 1.5$.

But let's try to be more general.

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

We can rewrite DIGM (using $\tau_i \equiv Y_{1i} - Y_{0i}$) to get an answer.

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

We can rewrite DIGM (using $\tau_i \equiv Y_{1i} - Y_{0i}$) to get an answer.

$$\text{DIGM} \quad \equiv \quad \frac{1}{m}\sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i}$$

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

We can rewrite DIGM (using $\tau_i \equiv Y_{1i} - Y_{0i}$) to get an answer.

$$\begin{aligned} \text{DIGM} &\equiv \frac{1}{m}\sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\ &= \frac{1}{m}\sum_{i=1}^{m} \left\{ Y_{0i} + \tau_i \right\} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \end{aligned}$$

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

We can rewrite DIGM (using $\tau_i \equiv Y_{1i} - Y_{0i}$) to get an answer.

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

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

We can rewrite DIGM (using $\tau_i \equiv Y_{1i} - Y_{0i}$) to get an answer.

$$
\begin{aligned}
\text{DIGM} &\equiv \frac{1}{m}\sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\
&= \frac{1}{m}\sum_{i=1}^{m} \{Y_{0i} + \tau_i\} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\
&= \frac{1}{m}\sum_{i=1}^{m} \tau_i + \{\frac{1}{m}\sum_{i=1}^{m} Y_{0i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i}\} \\
E\left[\text{DIGM}\right] &= E[\tau_i | D_i = 1] + \{E[Y_{0i} | D_i = 1] - E[Y_{0i} | D_i = 0]\}
\end{aligned}
$$

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

We can rewrite DIGM (using $\tau_i \equiv Y_{1i} - Y_{0i}$) to get an answer.

$$
\begin{aligned}
\text{DIGM} &\equiv \frac{1}{m}\sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\
&= \frac{1}{m}\sum_{i=1}^{m} \{Y_{0i} + \tau_i\} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\
&= \frac{1}{m}\sum_{i=1}^{m} \tau_i + \{\frac{1}{m}\sum_{i=1}^{m} Y_{0i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i}\} \\
E\big[\text{DIGM}\big] &= E[\tau_i | D_i = 1] + \{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\} \\
&\equiv \quad \text{ATT} \quad + \quad \{\text{selection bias}\}.
\end{aligned}
$$

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

We can rewrite DIGM (using $\tau_i \equiv Y_{1i} - Y_{0i}$) to get an answer.

$$
\begin{aligned}
\text{DIGM} &\equiv \frac{1}{m}\sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\
&= \frac{1}{m}\sum_{i=1}^{m} \{Y_{0i} + \tau_i\} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\
&= \frac{1}{m}\sum_{i=1}^{m} \tau_i + \{\frac{1}{m}\sum_{i=1}^{m} Y_{0i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i}\} \\
E\big[\text{DIGM}\big] &= E[\tau_i | D_i = 1] + \{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\} \\
&\equiv \quad \text{ATT} \quad + \quad \{\text{selection bias}\}.
\end{aligned}
$$

where *ATT* is the *ATE* in the treated group (average treatment effect for the treated).

Selection bias

Is DIGM an **unbiased estimator** of the ATE?

We can rewrite DIGM (using $\tau_i \equiv Y_{1i} - Y_{0i}$) to get an answer.

$$
\begin{aligned}
\text{DIGM} &\equiv \frac{1}{m}\sum_{i=1}^{m} Y_{1i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\
&= \frac{1}{m}\sum_{i=1}^{m} \{Y_{0i} + \tau_i\} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i} \\
&= \frac{1}{m}\sum_{i=1}^{m} \tau_i + \{\frac{1}{m}\sum_{i=1}^{m} Y_{0i} - \frac{1}{n-m}\sum_{i=m+1}^{n} Y_{0i}\} \\
E\big[\text{DIGM}\big] &= E[\tau_i | D_i = 1] + \{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\} \\
&\equiv \quad \text{ATT} \quad + \quad \{\text{selection bias}\}.
\end{aligned}
$$

where ATT is the ATE in the treated group (average treatment effect for the treated).

DIGM is an unbiased estimator of ATE if ATT = ATE and if there is no selection bias, i.e. if $E[Y_{0i}]$ is the same in the treatment group and control group.

Potential outcomes framework: activity 2!

Selection bias: the difference in average Y_{0i} between the treatment group and the control group.

Potential outcomes framework: activity 2!

Selection bias: the difference in average Y_{0i} between the treatment group and the control group.

1. Think of a (new) binary treatment and an outcome it might affect.
2. What are the units in your example?
3. What are the potential outcomes?
4. What would selection bias mean in this case?
5. In an observational study seeking to measure the effect of your treatment on your outcome (i.e. not a randomized experiment), do you think selection bias would be positive or negative?

Randomization and selection bias

Recall that

$$E\left[\text{DIGM}\right] \quad = \quad E[\tau_i|D_i = 1] + \left\{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\right\}$$

Randomization and selection bias

Recall that

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

Randomization and selection bias

Recall that

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

Randomization and selection bias

Recall that

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

Suppose D_i is assigned **at random**: $\Pr(D_i = 1) = c \ \forall \ i$.

Randomization and selection bias

Recall that

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

Suppose D_i is assigned **at random**: $\Pr(D_i = 1) = c \ \forall \ i$.

Then across randomizations,

- $E[\text{ATT}] = \text{ATE}$

Randomization and selection bias

Recall that

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

Suppose D_i is assigned **at random**: $\Pr(D_i = 1) = c \ \forall \ i$.

Then across randomizations,

- $E[\mathrm{ATT}] = \mathrm{ATE}$
- $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$, i.e. no selection bias.

Randomization and selection bias

Recall that

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

Suppose D_i is assigned **at random**: $\Pr(D_i = 1) = c \ \forall \ i$.

Then across randomizations,

- $E[\text{ATT}] = \text{ATE}$
- $E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0]$, i.e. no selection bias.

This is true even for small samples (though the variance across randomizations will be smaller for large samples).

Another way to explain why randomization works

Our target:

$$\text{ATE} \equiv E[\tau_i]$$

Another way to explain why randomization works

Our target:

$$\begin{aligned} \text{ATE} &\equiv E[\tau_i] \\ &= E[Y_{1i} - Y_{0i}] \end{aligned}$$

Another way to explain why randomization works

Our target:

$$\begin{aligned} \text{ATE} &\equiv E[\tau_i] \\ &= E[Y_{1i} - Y_{0i}] \\ &= E[Y_{1i}] - E[Y_{0i}] \end{aligned}$$

Another way to explain why randomization works

Our target:

$$\begin{aligned} \text{ATE} &\equiv E[\tau_i] \\ &= E[Y_{1i} - Y_{0i}] \\ &= E[Y_{1i}] - E[Y_{0i}] \end{aligned}$$

If we randomize treatment and control status, we observe Y_{1i} for a **representative sample** of the population, which gives us an unbiased estimate of $E[Y_{1i}]$.

Another way to explain why randomization works

Our target:

$$\begin{aligned} \text{ATE} &\equiv E[\tau_i] \\ &= E[Y_{1i} - Y_{0i}] \\ &= E[Y_{1i}] - E[Y_{0i}] \end{aligned}$$

If we randomize treatment and control status, we observe Y_{1i} for a **representative sample** of the population, which gives us an unbiased estimate of $E[Y_{1i}]$.

Same for Y_{0i}.

Another way to explain why randomization works

Our target:

$$\begin{aligned} \text{ATE} &\equiv E[\tau_i] \\ &= E[Y_{1i} - Y_{0i}] \\ &= E[Y_{1i}] - E[Y_{0i}] \end{aligned}$$

If we randomize treatment and control status, we observe Y_{1i} for a **representative sample** of the population, which gives us an unbiased estimate of $E[Y_{1i}]$.

Same for Y_{0i}.

So DIGM gives an unbiased estimate of ATE.

Randomization, selection bias, and balance tests

Selection bias can't be tested (why not?), but **balance tests** are often used to check for differences in observable **pre-treatment covariates** between the treatment and control groups:

Randomization, selection bias, and balance tests

Selection bias can't be tested (why not?), but **balance tests** are often used to check for differences in observable **pre-treatment covariates** between the treatment and control groups:

- covariate-by-covariate comparison of means (e.g. via t-tests)
- multivariate regression of treatment status (DV) on all covariates

Randomization, selection bias, and balance tests

Selection bias can't be tested (why not?), but **balance tests** are often used to check for differences in observable **pre-treatment covariates** between the treatment and control groups:

- covariate-by-covariate comparison of means (e.g. via t-tests)
- multivariate regression of treatment status (DV) on all covariates

Important: randomization ensures that **all** pre-treatment covariates (observable, unobservable) are balanced **in expectation** (across randomizations).

Randomization, selection bias, and balance tests

Selection bias can't be tested (why not?), but **balance tests** are often used to check for differences in observable **pre-treatment covariates** between the treatment and control groups:

- covariate-by-covariate comparison of means (e.g. via t-tests)
- multivariate regression of treatment status (DV) on all covariates

Important: randomization ensures that **all** pre-treatment covariates (observable, unobservable) are balanced **in expectation** (across randomizations).

Balance tests useful to detect botched randomization and MM recommend them.

Randomization, selection bias, and balance tests

Selection bias can't be tested (why not?), but **balance tests** are often used to check for differences in observable **pre-treatment covariates** between the treatment and control groups:

- covariate-by-covariate comparison of means (e.g. via t-tests)
- multivariate regression of treatment status (DV) on all covariates

Important: randomization ensures that **all** pre-treatment covariates (observable, unobservable) are balanced **in expectation** (across randomizations).

Balance tests useful to detect botched randomization and MM recommend them.

But a little controversial: can never say anything about balance in unobserved covariates, so "philosophically unsound, of no practical value, and potentially misleading" (Senn 1994, quoted in Imai, King, and Stuart 2008).

Internal and external validity

Internal validity is high when estimate of ATE is expected to be close to the true ATE.

Internal and external validity

Internal validity is high when estimate of ATE is expected to be close to the true ATE.

Depends on bias, sample size.

Internal and external validity

Internal validity is high when estimate of ATE is expected to be close to the true ATE.

Depends on bias, sample size.

External validity is "the extent to which the conclusions can be generalized beyond a particular study" (Imai *QSS* 2017).

Internal and external validity

Internal validity is high when estimate of ATE is expected to be close to the true ATE.

Depends on bias, sample size.

External validity is "the extent to which the conclusions can be generalized beyond a particular study" (Imai *QSS* 2017).

Usual point is: randomized experiments have high internal validity, but usually low external validity:

Internal and external validity

Internal validity is high when estimate of ATE is expected to be close to the true ATE.

Depends on bias, sample size.

External validity is "the extent to which the conclusions can be generalized beyond a particular study" (Imai *QSS* 2017).

Usual point is: randomized experiments have high internal validity, but usually low external validity:

- Sample not representative (convenience sample)
- Treatment or outcome not really what we are interested in

Internal and external validity

Internal validity is high when estimate of ATE is expected to be close to the true ATE.

Depends on bias, sample size.

External validity is "the extent to which the conclusions can be generalized beyond a particular study" (Imai *QSS* 2017).

Usual point is: randomized experiments have high internal validity, but usually low external validity:

- Sample not representative (convenience sample)
- Treatment or outcome not really what we are interested in

This critique is sometimes simplistic, unfair.

Internal and external validity (2)

You have a study, characterized by a sample that represents a population $\mathcal{P}$, a treatment D_i, an outcome Y_i. There is a true ATE for this study $E[\tau_i]$ **(your target)**, and you have an estimate of it $E[\hat{\tau}_i]$.

Internal and external validity (2)

You have a study, characterized by a sample that represents a population $\mathcal{P}$, a treatment D_i, an outcome Y_i. There is a true ATE for this study $E[\tau_i]$ **(your target)**, and you have an estimate of it $E\hat{[}\tau_i]$.

I can imagine a different population of interest $\tilde{\mathcal{P}}$, treatment $\tilde{D}_i$, outcome $\tilde{Y}_i$, for which there would be a different ATE $E[\tilde{\tau}_i]$ **(my target)**.

Internal and external validity (2)

You have a study, characterized by a sample that represents a population $\mathcal{P}$, a treatment D_i, an outcome Y_i. There is a true ATE for this study $E[\tau_i]$ **(your target)**, and you have an estimate of it $E[\hat{\tau}_i]$.

I can imagine a different population of interest $\tilde{\mathcal{P}}$, treatment $\tilde{D}_i$, outcome $\tilde{Y}_i$, for which there would be a different ATE $E[\tilde{\tau}_i]$ **(my target)**.

Internal validity: how close is your estimate to your target?

Internal and external validity (2)

You have a study, characterized by a sample that represents a population $\mathcal{P}$, a treatment D_i, an outcome Y_i. There is a true ATE for this study $E[\tau_i]$ **(your target)**, and you have an estimate of it $E[\hat{\tau}_i]$.

I can imagine a different population of interest $\tilde{\mathcal{P}}$, treatment $\tilde{D}_i$, outcome $\tilde{Y}_i$, for which there would be a different ATE $E[\tilde{\tau}_i]$ **(my target)**.

Internal validity: how close is your estimate to your target?

External validity: how close is your estimate to **my** target?

Internal and external validity (2)

You have a study, characterized by a sample that represents a population $\mathcal{P}$, a treatment D_i, an outcome Y_i. There is a true ATE for this study $E[\tau_i]$ **(your target)**, and you have an estimate of it $E[\hat{\tau}_i]$.

I can imagine a different population of interest $\tilde{\mathcal{P}}$, treatment $\tilde{D}_i$, outcome $\tilde{Y}_i$, for which there would be a different ATE $E[\tilde{\tau}_i]$ **(my target)**.

Internal validity: how close is your estimate to your target?

External validity: how close is your estimate to **my** target?

- External validity depends on (i) internal validity and (ii) how close your target and my target are

Internal and external validity (2)

You have a study, characterized by a sample that represents a population $\mathcal{P}$, a treatment D_i, an outcome Y_i. There is a true ATE for this study $E[\tau_i]$ **(your target)**, and you have an estimate of it $E[\hat{\tau}_i]$.

I can imagine a different population of interest $\tilde{\mathcal{P}}$, treatment $\tilde{D}_i$, outcome $\tilde{Y}_i$, for which there would be a different ATE $E[\tilde{\tau}_i]$ **(my target)**.

Internal validity: how close is your estimate to your target?

External validity: how close is your estimate to **my** target?

- External validity depends on (i) internal validity and (ii) how close your target and my target are
- (ii) is generally not known; relative importance of (i) and (ii) will depend on targets, nature of reality. Becomes ideological issue.

Internal and external validity (2)

You have a study, characterized by a sample that represents a population $\mathcal{P}$, a treatment D_i, an outcome Y_i. There is a true ATE for this study $E[\tau_i]$ **(your target)**, and you have an estimate of it $E[\hat{\tau}_i]$.

I can imagine a different population of interest $\tilde{\mathcal{P}}$, treatment $\tilde{D}_i$, outcome $\tilde{Y}_i$, for which there would be a different ATE $E[\tilde{\tau}_i]$ **(my target)**.

Internal validity: how close is your estimate to your target?

External validity: how close is your estimate to **my** target?

- External validity depends on (i) internal validity and (ii) how close your target and my target are
- (ii) is generally not known; relative importance of (i) and (ii) will depend on targets, nature of reality. Becomes ideological issue.
- I prefer an external validity critique like "Did you answer your question?" rather than "Did you answer my question?" Otherwise, just say you're not interested.

Types of randomized experiments and validity

- Field experiments: natural setting, real interventions, real outcomes.

Types of randomized experiments and validity

- Field experiments: natural setting, real interventions, real outcomes.
- Survey experiments: survey setting, "weak" interventions (information, priming; forced choice among different alternatives), survey outcomes.

Types of randomized experiments and validity

- Field experiments: natural setting, real interventions, real outcomes.
- Survey experiments: survey setting, "weak" interventions (information, priming; forced choice among different alternatives), survey outcomes.
- Lab experiments: lab setting, contrived interventions and outcomes (but can be incentivized with money)

Types of randomized experiments and validity

- Field experiments: natural setting, real interventions, real outcomes.
- Survey experiments: survey setting, "weak" interventions (information, priming; forced choice among different alternatives), survey outcomes.
- Lab experiments: lab setting, contrived interventions and outcomes (but can be incentivized with money)

Activity: Assess the internal and external validity of these types of experiments.

Design of experiments: considerations

(1) Designing an experiment one often picks n (sample size). How to choose?

Design of experiments: considerations

(1) Designing an experiment one often picks n (sample size). How to choose?

- **Reality** If you have £X and the cost per subject is c, $n = X/c$

Design of experiments: considerations

(1) Designing an experiment one often picks n (sample size). How to choose?

- **Reality** If you have £X and the cost per subject is c, $n = X/c$
- **Ideal** Power analysis: choose n such that in 80% of randomizations you would reject the null hypothesis at 95% confidence level, given assumptions about size of ATE and variance of outcomes in treatment and control group

Design of experiments: considerations

(1) Designing an experiment one often picks n (sample size). How to choose?

- **Reality** If you have £X and the cost per subject is c, $n = X/c$
- **Ideal** Power analysis: choose n such that in 80% of randomizations you would reject the null hypothesis at 95% confidence level, given assumptions about size of ATE and variance of outcomes in treatment and control group

(2) Types of treatments:

- Baseline case: $k \geq 2$ distinct alternatives, one per unit/subject (**between-subjects** design)

Design of experiments: considerations

(1) Designing an experiment one often picks n (sample size). How to choose?

- **Reality** If you have £X and the cost per subject is c, $n = X/c$
- **Ideal** Power analysis: choose n such that in 80% of randomizations you would reject the null hypothesis at 95% confidence level, given assumptions about size of ATE and variance of outcomes in treatment and control group

(2) Types of treatments:

- Baseline case: $k \geq 2$ distinct alternatives, one per unit/subject (**between-subjects** design)
- In some survey and lab experiments, multiple treatments assigned in sequence (**within-subjects** comparisons)

Design of experiments: considerations

(1) Designing an experiment one often picks n (sample size). How to choose?

- **Reality** If you have £X and the cost per subject is c, $n = X/c$
- **Ideal** Power analysis: choose n such that in 80% of randomizations you would reject the null hypothesis at 95% confidence level, given assumptions about size of ATE and variance of outcomes in treatment and control group

(2) Types of treatments:

- Baseline case: $k \geq 2$ distinct alternatives, one per unit/subject (**between-subjects** design)
- In some survey and lab experiments, multiple treatments assigned in sequence (**within-subjects** comparisons)
- Treatments can also be composite of many features: e.g. factorial design, forced-choice conjoint