

What is Your Estimand?

Defining the Target Quantity
Connects Statistical Evidence
to Theory



Ian Lundberg

UCLA
Sociology
ianlundberg.org

Rebecca Johnson

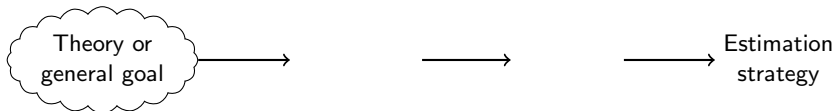
Dartmouth Quantitative
Social Science
rebeccajohnson.io

Brandon M. Stewart

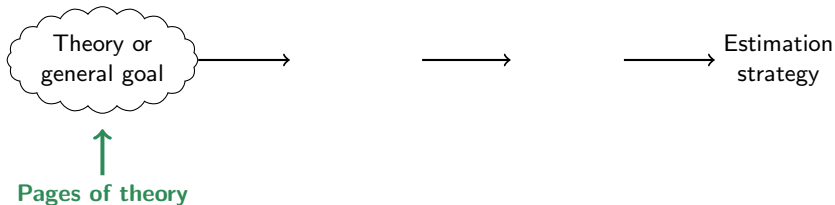
Princeton
Sociology
brandonstewart.org

6 October 2021. University of Wisconsin CDE Training Workshop.
Paper in *American Sociological Review*. Preprint on [SocArxiv](https://www.socarxiv.org/). Replication code on [Dataverse](https://dataverse.org/). Research reported in this publication was supported by The Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under Award Number P2CHD047879.

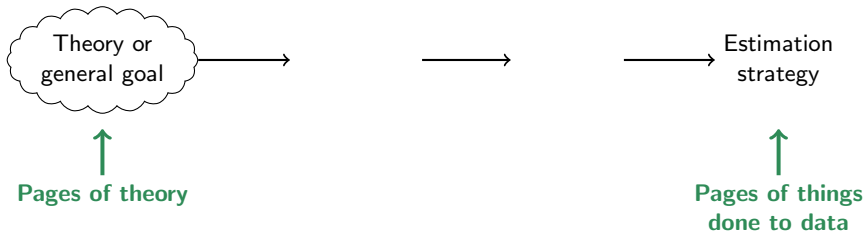
Research framework: Estimands connect theory to evidence



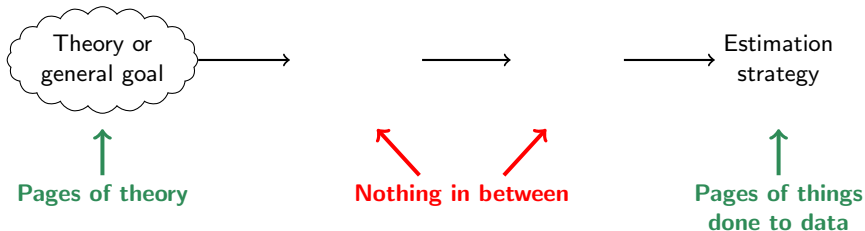
Research framework: Estimands connect theory to evidence



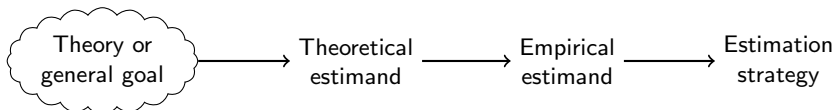
Research framework: Estimands connect theory to evidence



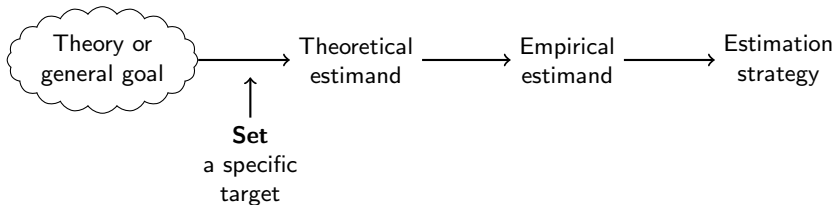
Research framework: Estimands connect theory to evidence



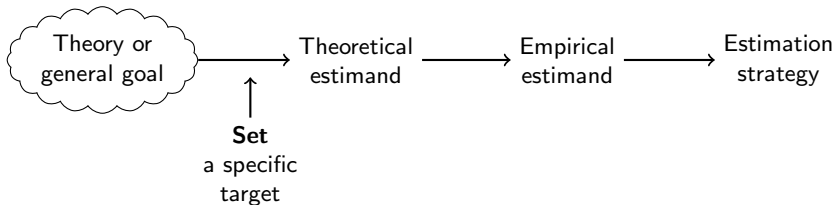
Research framework: Estimands connect theory to evidence



Research framework: Estimands connect theory to evidence



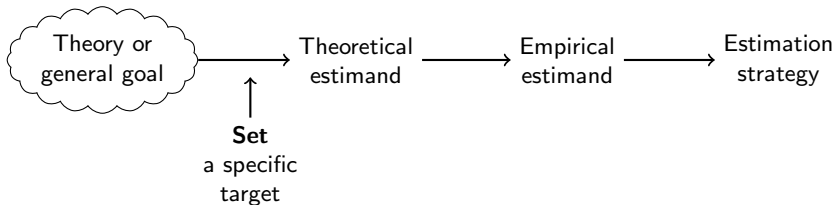
Research framework: Estimands connect theory to evidence



Definition

A **unit-specific quantity**
aggregated over a
target population

Research framework: Estimands connect theory to evidence



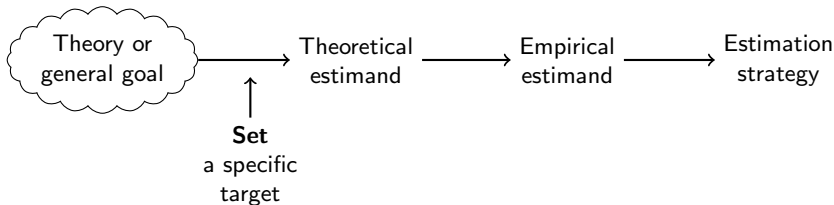
Definition

A **unit-specific quantity**
aggregated over a
target population

Example

$$\frac{1}{\text{Size of U.S. adult population}} \sum_{i \text{ in U.S. adult population}} \left(\text{Employed}_i \right)$$

Research framework: Estimands connect theory to evidence



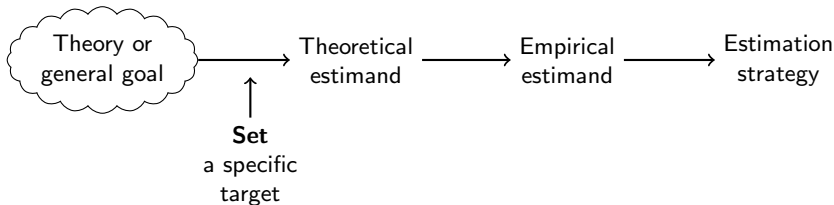
Definition

A **unit-specific quantity**
aggregated over a
target population

Example

$$\frac{1}{\text{Size of U.S. adult population}} \sum_{i \text{ in U.S. adult population}} \left(\underbrace{\text{Employed}_i(\text{Job training})}_{\text{Employment if received job training}} - \underbrace{\text{Employed}_i(\text{No job training})}_{\text{Employment if did not receive job training}} \right)$$

Research framework: Estimands connect theory to evidence



Definition

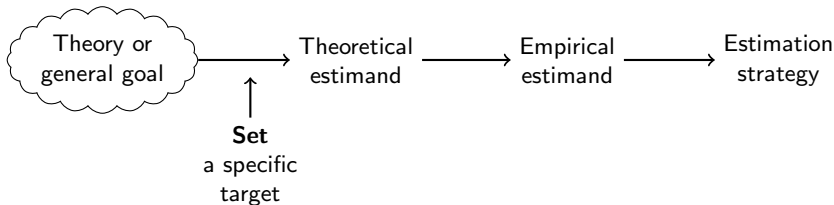
A **unit-specific quantity**
aggregated over a
target population

Example

$$\frac{1}{\text{Size of U.S. adult population}} \sum_{i \text{ in U.S. adult population}} \left(\underbrace{\text{Employed}_i(\text{Job training})}_{\text{Employment if received job training}} - \underbrace{\text{Employed}_i(\text{No job training})}_{\text{Employment if did not receive job training}} \right)$$

Liebersen 1987, Abbott 1988, Freedman 1991, Xie 2013, Hernán 2018

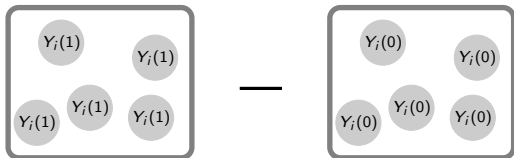
Research framework: Estimands connect theory to evidence



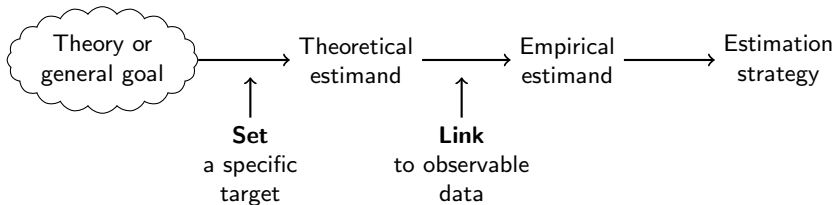
Definition

A **unit-specific quantity**
aggregated over a
target population

Example



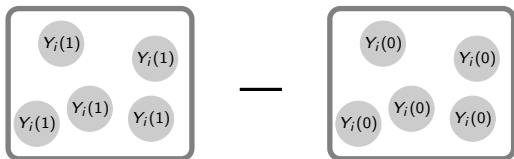
Research framework: Estimands connect theory to evidence



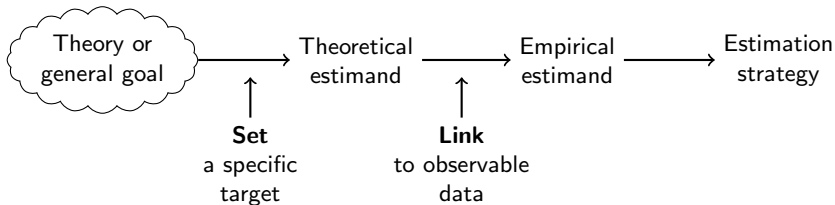
Definition

A quantity involving
observable data

Example



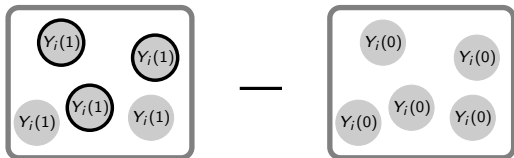
Research framework: Estimands connect theory to evidence



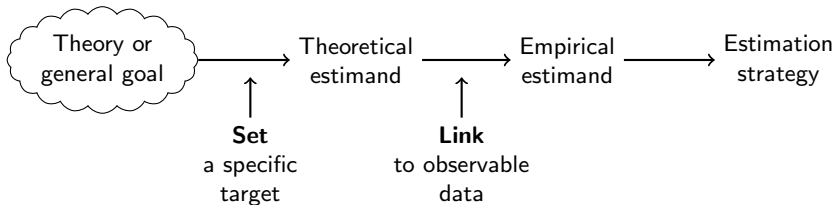
Definition

A quantity involving
observable data

Example



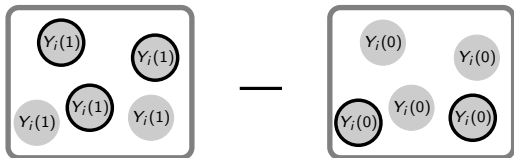
Research framework: Estimands connect theory to evidence



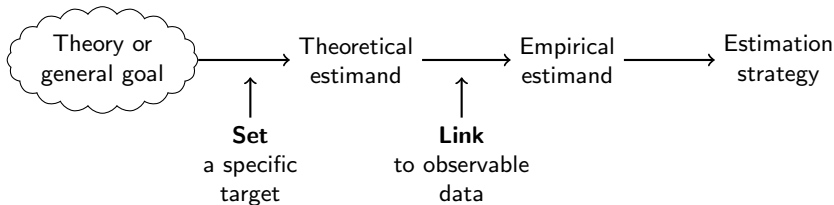
Definition

A quantity involving
observable data

Example



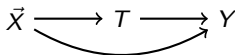
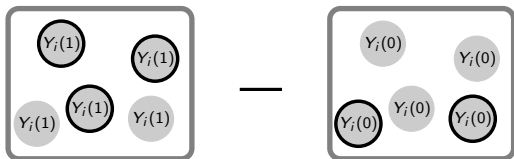
Research framework: Estimands connect theory to evidence



Definition

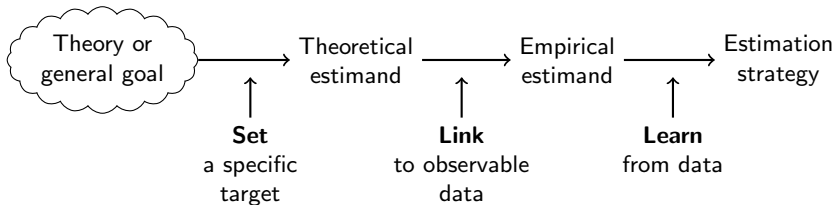
A quantity involving
observable data

Example



Pearl 2009, Imbens and Rubin 2015,
Morgan and Winship 2015, Elwert and Winship 2014

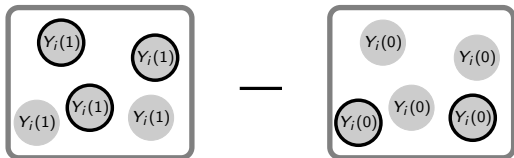
Research framework: Estimands connect theory to evidence



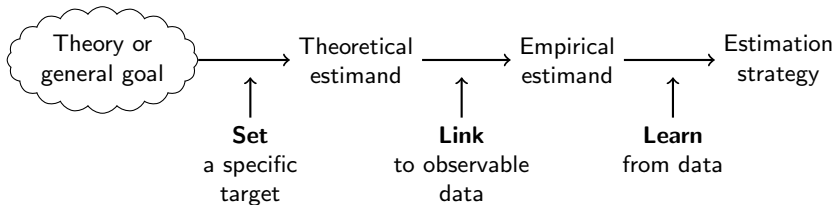
Definition

An algorithm applied to data

Example



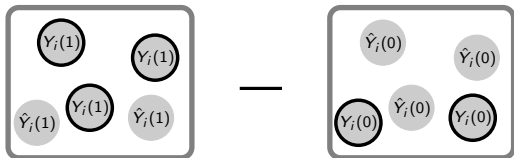
Research framework: Estimands connect theory to evidence



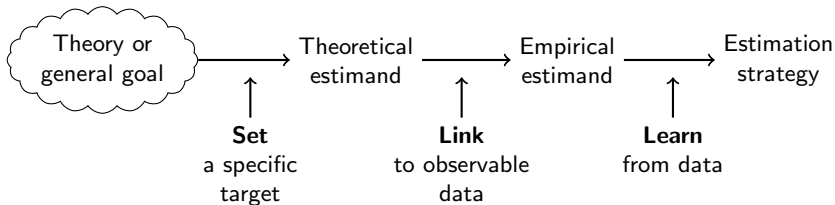
Definition

An algorithm applied to data

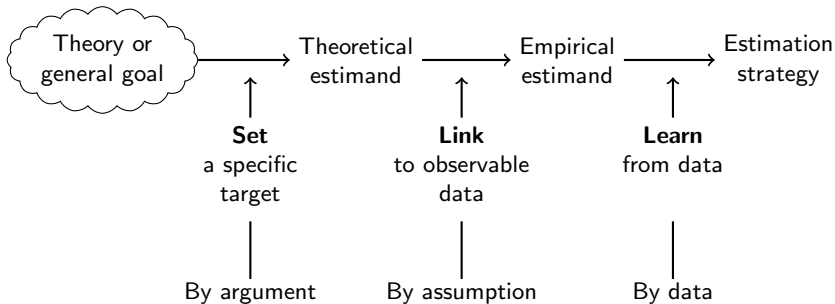
Example



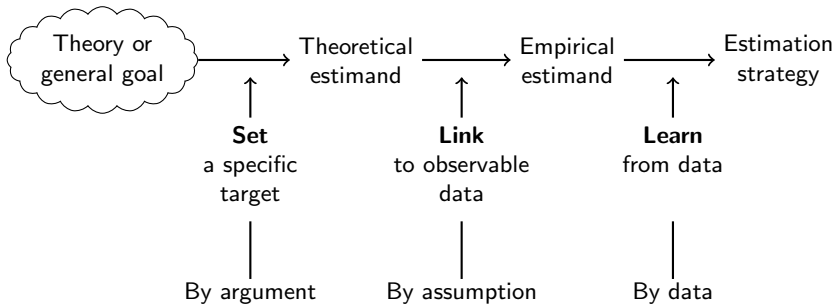
Research framework: Estimands connect theory to evidence



Research framework: Estimands connect theory to evidence



Research framework: Estimands connect theory to evidence



Plan for today

- Intro: What this looks like in practice
- Example: Revisiting a published paper
- Group exercises: Together we will state precise research goals



1. Set the target quantity.



Describe a population

What is the proportion employed
among U.S. resident women ages 21–35?



Describe a population

What is the proportion employed
among U.S. resident women ages 21–35?

Woman 1

Woman 2

Woman 3

Woman 4



Describe a population

What is the proportion employed
among U.S. resident women ages 21–35?

	<u>Employed?</u>
Woman 1	1
Woman 2	0
Woman 3	1
Woman 4	1



Describe population subgroups

What is the proportion employed among U.S. resident women ages 21–35, comparing mothers to non-mothers?



Describe population subgroups

What is the proportion employed among U.S. resident women ages 21–35, comparing mothers to non-mothers?

	<u>Employed?</u>		<u>Employed?</u>
Mother 1	0	Non-Mother 1	1
Mother 2	0	Non-Mother 2	0
Mother 3	0	Non-Mother 3	1
Mother 4	1	Non-Mother 4	1



Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?



Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

Woman 1

Woman 2

Woman 3

Woman 4



Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$
Woman 1	0
Woman 2	0
Woman 3	0
Woman 4	1



Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$
Woman 1	0	1
Woman 2	0	0
Woman 3	0	1
Woman 4	1	1



Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1	-1
Woman 2	0	0	0
Woman 3	0	1	-1
Woman 4	1	1	0



Describe population subgroups

What is the proportion employed among U.S. resident women ages 21–35, comparing mothers to non-mothers?

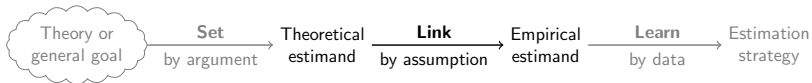
	Employed?		Employed?
Mother 1	0	Non-Mother 1	1
Mother 2	0	Non-Mother 2	0
Mother 3	0	Non-Mother 3	1
Mother 4	1	Non-Mother 4	1

Causal effect in a population

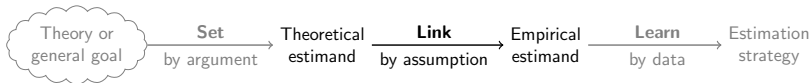
What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1	-1
Woman 2	0	0	0
Woman 3	0	1	-1
Woman 4	1	1	0

Very
different
research
goals



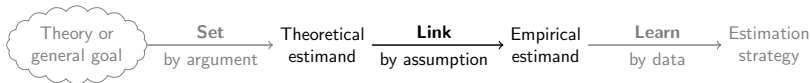
2. Link to observables



Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

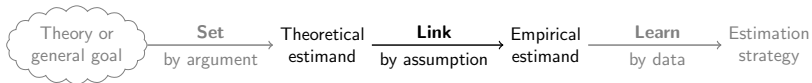
	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1	-1
Woman 2	0	0	0
Woman 3	0	1	-1
Woman 4	1	1	0



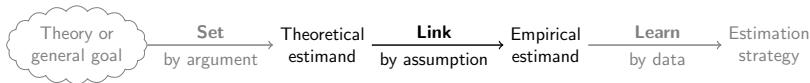
Causal effect in a population

What is the causal effect of motherhood on employment among U.S. resident women ages 21–35?

	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	?	?
Woman 2	0	?	?
Woman 3	?	1	?
Woman 4	?	1	?



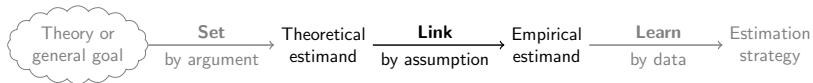
	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
	<hr/>	<hr/>	<hr/>
Woman 1	0	?	?
Woman 2	0	?	?
Woman 3	?	1	?
Woman 4	?	1	?



Subpopulation

- Age 30
- College-educated
- Employed last year

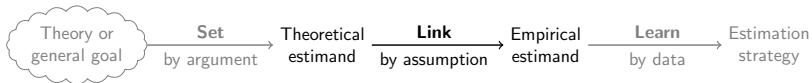
	Would be employed if a mother? <u>Y(1)</u>	Would be employed if a non-mother? <u>Y(0)</u>	Causal effect <u>Y(1) – Y(0)</u>
Woman 1	0	?	?
Woman 2	0	?	?
Woman 3	?	1	?
Woman 4	?	1	?



Subpopulation

- Age 30
- College-educated
- Employed last year

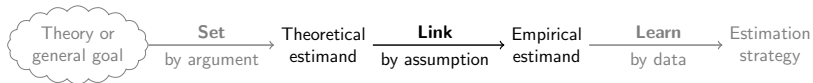
	Would be employed if a mother? <u>Y(1)</u>	Would be employed if a non-mother? <u>Y(0)</u>	Causal effect <u>Y(1) – Y(0)</u>
Woman 1	0	?	?
Woman 2	0	?	?
Woman 3	0?	1	?
Woman 4	0?	1	?



Subpopulation

- Age 30
- College-educated
- Employed last year

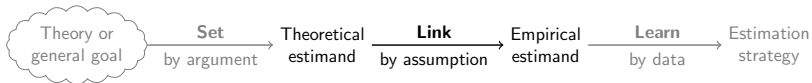
	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1?	?
Woman 2	0	1?	?
Woman 3	0?	1	?
Woman 4	0?	1	?



Subpopulation

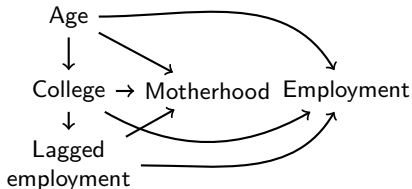
- Age 30
- College-educated
- Employed last year

	Would be employed if a mother? <u>$Y(1)$</u>	Would be employed if a non-mother? <u>$Y(0)$</u>	Causal effect <u>$Y(1) - Y(0)$</u>
Woman 1	0	1?	-1?
Woman 2	0	1?	-1?
Woman 3	0?	1	-1?
Woman 4	0?	1	-1?

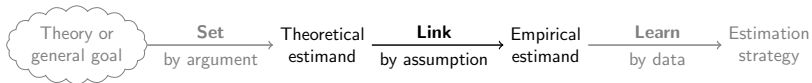


Subpopulation

- Age 30
- College-educated
- Employed last year

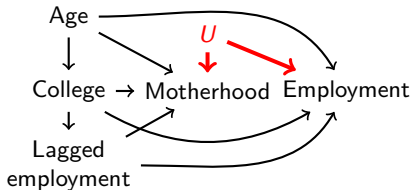


	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1?	-1?
Woman 2	0	1?	-1?
Woman 3	0?	1	-1?
Woman 4	0?	1	-1?

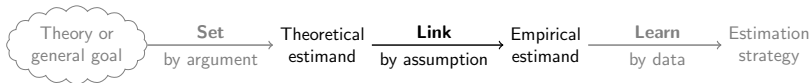


Subpopulation

- Age 30
- College-educated
- Employed last year

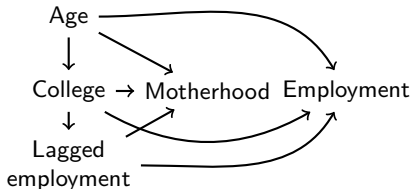


	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1?	-1?
Woman 2	0	1?	-1?
Woman 3	0?	1	-1?
Woman 4	0?	1	-1?

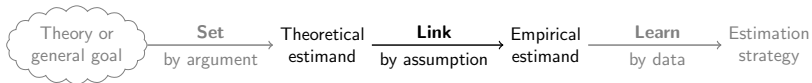


Subpopulation

- Age 30
- College-educated
- Employed last year



	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1?	-1?
Woman 2	0	1?	-1?
Woman 3	0?	1	-1?
Woman 4	0?	1	-1?



Subpopulation

- Age 30
- College-educated
- Employed last year

Getting the population

- 1) Do this in every subpopulation
- 2) Aggregate over subpopulations

	Would be employed if a mother? $Y(1)$	Would be employed if a non-mother? $Y(0)$	Causal effect $Y(1) - Y(0)$
Woman 1	0	1?	-1?
Woman 2	0	1?	-1?
Woman 3	0?	1	-1?
Woman 4	0?	1	-1?



3. Learn from data



Subpopulation

- Age 30
- College-educated
- Employed last year

	Would be employed if a mother? $Y(1)$
Woman 1	0
Woman 2	0
Woman 3	0?
Woman 4	0?



Subpopulation

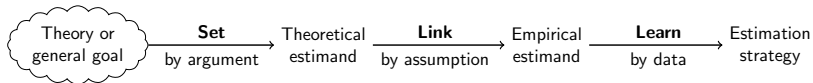
- Age 30
- College-educated
- Employed last year

	Would be employed if a mother? $Y(1)$
Woman 1	0
Woman 2	0
Woman 3	0?
Woman 4	0?



Learn

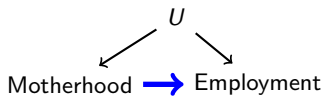
Do this better
with machine
learning

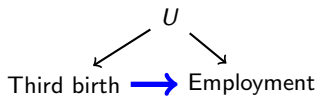
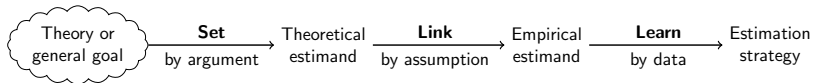


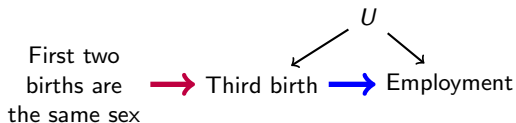
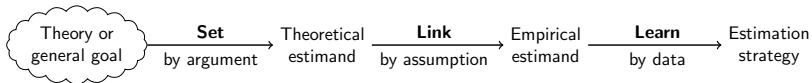
Example

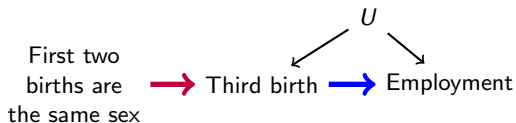


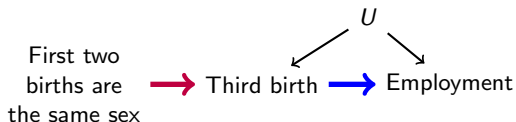
Motherhood → Employment











Mechanics of instrumental variables

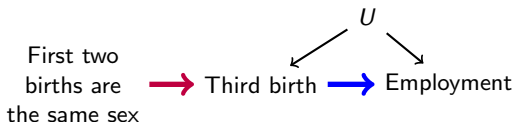
$$\text{blue arrow} = \frac{\text{green arrow}}{\text{red arrow}}$$

(Note: This can be made more precise)



But wait!

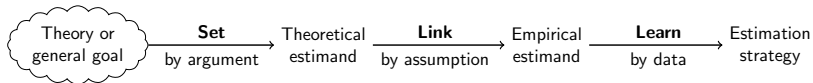
Not everyone is affected by this instrument



Mechanics of instrumental variables

$$\xrightarrow{\text{blue}} = \frac{\xrightarrow{\text{green}}}{\xrightarrow{\text{red}}}$$

(Note: This can be made more precise)



Four Subpopulations

(Angrist, Imbens, & Rubin 1996)



Four Subpopulations

(Angrist, Imbens, & Rubin 1996)

Always-Takers. Third kid no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\quad}$ Employment

(Note: A red dashed double-headed arrow is positioned above the 'Third birth' text in the original image.)



Four Subpopulations

(Angrist, Imbens, & Rubin 1996)

Always-Takers. Third kid no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\quad}$ Employment

Never-Takers. Sticking with two kids, no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\quad}$ Employment



Four Subpopulations

(Angrist, Imbens, & Rubin 1996)

Always-Takers. Third kid no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

Never-Takers. Sticking with two kids, no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

Compliers. Third kid if and only if the first two are the same sex.

First two births same sex $\overset{+1}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment



Four Subpopulations

(Angrist, Imbens, & Rubin 1996)

Always-Takers. Third kid no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

Never-Takers. Sticking with two kids, no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

Compliers. Third kid if and only if the first two are the same sex.

First two births same sex $\overset{+1}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

Defiers. Third kid if and only if the first two are NOT the same sex.

First two births same sex $\overset{-1}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment



Four Subpopulations

(Angrist, Imbens, & Rubin 1996)

Always-Takers. Third kid no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

Never-Takers. Sticking with two kids, no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

Compliers. Third kid if and only if the first two are the same sex.

First two births same sex $\overset{+1}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

~~**Defiers.**~~ (assumed away) Third kid if and only if the first two are NOT the same sex.

First two births same sex $\overset{-1}{- - -} \rightarrow$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment



Four Subpopulations

(Angrist, Imbens, & Rubin 1996)

Always-Takers. Third kid no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\quad}$ Employment

Never-Takers. Sticking with two kids, no matter what.

First two births same sex $\overset{0}{- - -} \rightarrow$ Third birth $\xrightarrow{\quad}$ Employment

Compliers. Third kid if and only if the first two are the same sex.

First two births same sex $\overset{+1}{- - -} \rightarrow$ Third birth $\xrightarrow{\quad}$ Employment

~~**Defiers.**~~ ^(assumed away) Third kid if and only if the first two are NOT the same sex.

First two births same sex $\overset{-1}{- - -} \rightarrow$ Third birth $\xrightarrow{\quad}$ Employment



Takeaway

Instrumental variables estimates the average causal effect among a particular **target population**: compliers

Compliers. Third kid if and only if the first two are the same sex.

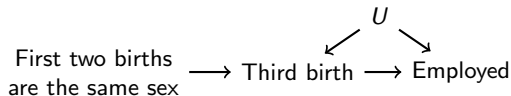
First two births same sex $\overset{+1}{- - - \rightarrow}$ Third birth $\xrightarrow{\hspace{1cm}}$ Employment

$\xleftarrow{\hspace{1cm}} \xrightarrow{\hspace{1cm}}$



Vague estimand

Effect of motherhood
on employment

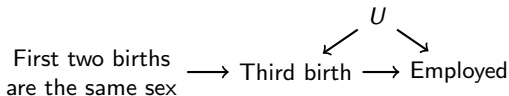




Vague estimand

Effect of motherhood
on employment

Precise estimand





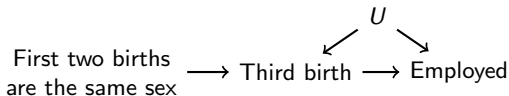
Vague estimand

Effect of motherhood
on employment

Precise estimand

Effect of having **3 vs. 2 children**

**unit-specific
quantity**





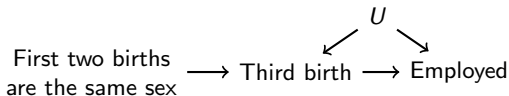
Vague estimand

Effect of motherhood
on employment

target population

Precise estimand

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex





Precise estimand

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex

$\approx 4\%$ of all mothers



Precise estimand

Effect of having
1 vs. 0 children
among all women

Precise estimand

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex

$\approx 4\%$ of all mothers



Precise estimand A

Effect of having
1 vs. 0 children
among all women

Precise estimand B

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex

$\approx 4\%$ of all mothers



Precise estimand A

Effect of having
1 vs. 0 children
among all women

Precise estimand B

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex

$\approx 4\%$ of all mothers

The estimand you choose affects your argument.



Precise estimand A

Effect of having
1 vs. 0 children
among all women

Precise estimand B

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex

$\approx 4\%$ of all mothers

The estimand you choose affects your argument.

Choice A) By what assumptions can we learn that?



Precise estimand A

Effect of having
1 vs. 0 children
among all women

Precise estimand B

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex

$\approx 4\%$ of all mothers

The estimand you choose affects your argument.

Choice A) By what assumptions can we learn that? (**link** step is hard)



Precise estimand A

Effect of having
1 vs. 0 children
among all women

Precise estimand B

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex

$\approx 4\%$ of all mothers

The estimand you choose affects your argument.

Choice A) By what assumptions can we learn that? (**link** step is hard)

Choice B) Why does this quantity matter for theory?



Precise estimand A

Effect of having
1 vs. 0 children
among all women

Precise estimand B

Effect of having 3 vs. 2 children
among those with at least two children who
would have a third birth if and only if the
first two were of the same sex

$\approx 4\%$ of all mothers

The estimand you choose affects your argument.

Choice A) By what assumptions can we learn that? (**link** step is hard)

Choice B) Why does this quantity matter for theory? (**set** step is hard)

Group exercises

Group exercises: Plan

1. We will give you a vague research question
2. In groups, you will choose some way to make it precise
3. Our answers may differ. That is ok
4. We will come together and discuss

Exercise 1

“Marriage has a large effect on the economic well-being of poor women.”

— A policymaker

What is the estimand or target quantity?

1. What is the unit-specific quantity?

- Define the unit of analysis
- Define the outcome
- Define the treatment variable (if any)
- Define the values to which the treatment is set (if any)

2. What is/are the target population(s) of units?

Note: There is no single right answer.

Exercise 2

“Those whose parents attended college are more likely to complete college themselves”

— A sociologist

What is the estimand or target quantity?

1. What is the unit-specific quantity?
 - Define the unit of analysis
 - Define the outcome
 - Define the treatment variable (if any)
 - Define the values to which the treatment is set (if any)
2. What is/are the target population(s) of units?

Note: There is no single right answer.