21. Principal Stratification (Part 2)

Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

3 Nov 2022

Learning goals for today

At the end of class, you will be able to:

1. Finish the class exercise we started on Tuesday [solutions]
2. See principal stratification in action:
   quantifying racial bias in policing

Wrapping up the class exercise: Question 6

Size of the stratum who have a child regardless of college

Wrapping up the class exercise: Question 6

Size of the stratum who have a child regardless of college

Among those who finish college ($A = 1$),
having a child ($M = 1$) is the same as
being in the always-child stratum ($S = 1$)

Wrapping up the class exercise: Question 6

Size of the stratum who have a child regardless of college

Among those who finish college ($A = 1$),
having a child ($M = 1$) is the same as
being in the always-child stratum ($S = 1$)

$$\mathrm{P}(M = 1 \mid A = 1) = \mathrm{P}(S = 1 \mid A = 1)$$

Wrapping up the class exercise: Question 6

Size of the stratum who have a child regardless of college

Among those who finish college ($A = 1$),
having a child ($M = 1$) is the same as
being in the always-child stratum ($S = 1$)

$$\mathrm{P}(M = 1 \mid A = 1) = \mathrm{P}(S = 1 \mid A = 1)$$

We are assuming A is randomly assigned, so

$$\mathrm{P}(S = 1 \mid A = 1) = \mathrm{P}(S = 1)$$

Wrapping up the class exercise: Question 6

Size of the stratum who have a child regardless of college

Among those who finish college ($A = 1$),
having a child ($M = 1$) is the same as
being in the always-child stratum ($S = 1$)

$$\mathrm{P}(M = 1 \mid A = 1) = \mathrm{P}(S = 1 \mid A = 1)$$

We are assuming A is randomly assigned, so

$$\mathrm{P}(S = 1 \mid A = 1) = \mathrm{P}(S = 1)$$

So $\mathrm{P}(S = 1)$ equals the motherhood rate
among college women: 0.71

Wrapping up the class exercise: Question 7

Size of the stratum who have a child only if no college

Wrapping up the class exercise: Question 7

Size of the stratum who have a child only if no college

Among those who do not finish college ($A = 0$),
having a child ($M = 1$) is equivalent to being in either
the always-child stratum ($S = 1$) or
the child-if-no-college stratum ($S = 3$)

$$\text{P}(M = 1 \mid A = 0) = \text{P}(S \in \{1, 3\} \mid A = 1)$$

Wrapping up the class exercise: Question 7

Size of the stratum who have a child only if no college

Among those who do not finish college ($A = 0$),
having a child ($M = 1$) is equivalent to being in either
the always-child stratum ($S = 1$) or
the child-if-no-college stratum ($S = 3$)

$$\text{P}(M = 1 \mid A = 0) = \text{P}(S \in \{1, 3\} \mid A = 1)$$

We are assuming A is randomly assigned, so

$$\text{P}(S = \in \{1, 3\} \mid A = 0) = \text{P}(S = \in \{1, 3\})$$

Wrapping up the class exercise: Question 7

Size of the stratum who have a child only if no college

Among those who do not finish college ($A = 0$),
having a child ($M = 1$) is equivalent to being in either
the always-child stratum ($S = 1$) or
the child-if-no-college stratum ($S = 3$)

$$\mathrm{P}(M = 1 \mid A = 0) = \mathrm{P}(S \in \{1, 3\} \mid A = 1)$$

We are assuming A is randomly assigned, so

$$\mathrm{P}(S = \in \{1, 3\} \mid A = 0) = \mathrm{P}(S = \in \{1, 3\})$$

So $\mathrm{P}(S \in \{1, 3\})$ equals the motherhood rate
among non-college women: 0.83

Wrapping up the class exercise: Question 7

Size of the stratum who have a child only if no college

Among those who do not finish college ($A = 0$),
having a child ($M = 1$) is equivalent to being in either
the always-child stratum ($S = 1$) or
the child-if-no-college stratum ($S = 3$)

$$\text{P}(M = 1 \mid A = 0) = \text{P}(S \in \{1, 3\} \mid A = 1)$$

We are assuming A is randomly assigned, so

$$\text{P}(S =\in \{1, 3\} \mid A = 0) = \text{P}(S =\in \{1, 3\})$$

So $\text{P}(S \in \{1, 3\})$ equals the motherhood rate
among non-college women: 0.83

Estimate $\text{P}(S = 3) = \text{P}(S \in \{1, 3\}) - \text{P}(S = 1) = .83 - .71 = .12$

Wrapping up the class exercise: Question 8

Child outcome under mother college, in the always-child stratum

Wrapping up the class exercise: Question 8

Child outcome under mother college, in the always-child stratum

$$\mu_1^1 = \mathrm{E}(Y^1 \mid S = 1)$$

Wrapping up the class exercise: Question 8

Child outcome under mother college, in the always-child stratum

$$
\begin{aligned}
\mu_1^1 &= \mathsf{E}(Y^1 \mid S = 1) \\
&= \mathsf{E}(Y \mid A = 1, S = 1) \quad \text{exchangeability}
\end{aligned}
$$

Wrapping up the class exercise: Question 8

Child outcome under mother college, in the always-child stratum

$$\begin{aligned}
\mu_1^1 &= \mathsf{E}(Y^1 \mid S = 1) \\
&= \mathsf{E}(Y \mid A = 1, S = 1) && \text{exchangeability} \\
&= \mathsf{E}(Y \mid A = 1) && \text{monotonicity}
\end{aligned}$$

Wrapping up the class exercise: Question 8

Child outcome under mother college, in the always-child stratum

$$\begin{aligned}
\mu_1^1 &= \mathsf{E}(Y^1 \mid S = 1) \\
&= \mathsf{E}(Y \mid A = 1, S = 1) \quad \text{exchangeability} \\
&= \mathsf{E}(Y \mid A = 1) \quad \text{monotonicity} \\
&= 0.40
\end{aligned}$$

Wrapping up the class exercise: Question 9

Child outcome under mother no college, in the always-child stratum

Wrapping up the class exercise: Question 9

Child outcome under mother no college, in the always-child stratum

To identify μ_1^0, first write a formula for $\bar{y}^0$: the mean outcome among non-college women.

Wrapping up the class exercise: Question 9

Child outcome under mother no college, in the always-child stratum

To identify μ_1^0, first write a formula for $\bar{y}^0$: the mean outcome among non-college women.

That is a weighted average of outcomes among strata 1 and 3.

$$\overbrace{\frac{\pi_1}{\pi_1+\pi_3}}^{\substack{\text{Proportion} \\ \text{in } S=1}} \mu_1^0 + \overbrace{\frac{\pi_3}{\pi_1+\pi_3}}^{\substack{\text{Proportion} \\ \text{in } S=3}} \mu_3^0 = \overbrace{\bar{y}^0}^{\substack{\text{Mean outcome among} \\ \text{non-college mothers}}}$$

Wrapping up the class exercise: Question 9

Child outcome under mother no college, in the always-child stratum

To identify μ_1^0, first write a formula for $\bar{y}^0$: the mean outcome among non-college women.

That is a weighted average of outcomes among strata 1 and 3.

$$\overbrace{\frac{\pi_1}{\pi_1+\pi_3}}^{\substack{\text{Proportion} \\ \text{in } S=1}} \mu_1^0 + \overbrace{\frac{\pi_3}{\pi_1+\pi_3}}^{\substack{\text{Proportion} \\ \text{in } S=3}} \mu_3^0 = \overbrace{\bar{y}^0}^{\substack{\text{Mean outcome among} \\ \text{non-college mothers}}}$$

Rearrange to solve for μ_1^0

Wrapping up the class exercise: Question 9

Child outcome under mother no college, in the always-child stratum

To identify μ_1^0, first write a formula for $\bar{y}^0$: the mean outcome among non-college women.

That is a weighted average of outcomes among strata 1 and 3.

$$\overbrace{\frac{\pi_1}{\pi_1+\pi_3}}^{\substack{\text{Proportion}\\ \text{in } S=1}}\mu_1^0 + \overbrace{\frac{\pi_3}{\pi_1+\pi_3}}^{\substack{\text{Proportion}\\ \text{in } S=3}}\mu_3^0 = \overbrace{\bar{y}^0}^{\substack{\text{Mean outcome among}\\ \text{non-college mothers}}}$$

Rearrange to solve for μ_1^0

$$\frac{\pi_1}{\pi_1+\pi_3}\mu_1^0 = \bar{y}^0 - \frac{\pi_3}{\pi_1+\pi_3}\mu_3^0$$

Wrapping up the class exercise: Question 9

Child outcome under mother no college, in the always-child stratum

To identify μ_1^0, first write a formula for $\bar{y}^0$: the mean outcome among non-college women.

That is a weighted average of outcomes among strata 1 and 3.

$$\overbrace{\frac{\pi_1}{\pi_1+\pi_3}}^{\substack{\text{Proportion} \\ \text{in } S=1}} \mu_1^0 + \overbrace{\frac{\pi_3}{\pi_1+\pi_3}}^{\substack{\text{Proportion} \\ \text{in } S=3}} \mu_3^0 = \overbrace{\bar{y}^0}^{\substack{\text{Mean outcome among} \\ \text{non-college mothers}}}$$

Rearrange to solve for μ_1^0

$$\frac{\pi_1}{\pi_1+\pi_3}\mu_1^0 = \bar{y}^0 - \frac{\pi_3}{\pi_1+\pi_3}\mu_3^0$$

$$\mu_1^0 = \frac{\pi_1+\pi_3}{\pi_1}\left(\bar{y}^0 - \frac{\pi_3}{\pi_1+\pi_3}\mu_3^0\right)$$

Wrapping up the class exercise: Question 10

Child outcome under mother no college, in the always-child stratum

$$\mu_1^0 = \frac{\pi_1 + \pi_3}{\pi_1}\left(\bar{y}^0 - \frac{\pi_3}{\pi_1 + \pi_3}\mu_3^0\right)$$

Wrapping up the class exercise: Question 10

Child outcome under mother no college, in the always-child stratum

$$\mu_1^0 = \frac{\pi_1 + \pi_3}{\pi_1}\left(\bar{y}^0 - \frac{\pi_3}{\pi_1 + \pi_3}\mu_3^0\right)$$

The only unknown term is μ_3^0. We can bound it.

Wrapping up the class exercise: Question 10

Child outcome under mother no college, in the always-child stratum

$$\mu_1^0 = \frac{\pi_1 + \pi_3}{\pi_1}\left(\bar{y}^0 - \frac{\pi_3}{\pi_1 + \pi_3}\mu_3^0\right)$$

The only unknown term is μ_3^0. We can bound it.

$$\hat{\mu}_1^{0,\text{Upper}} = \frac{.71 + .12}{.71}\left(.18 - \frac{.12}{.71 + .12}1\right) = .04$$

Wrapping up the class exercise: Question 10

Child outcome under mother no college, in the always-child stratum

$$\mu_1^0 = \frac{\pi_1 + \pi_3}{\pi_1}\left(\bar{y}^0 - \frac{\pi_3}{\pi_1 + \pi_3}\mu_3^0\right)$$

The only unknown term is μ_3^0. We can bound it.

$$\hat{\mu}_1^{0,\text{Upper}} = \frac{.71 + .12}{.71}\left(.18 - \frac{.12}{.71 + .12}1\right) = .04$$

$$\hat{\mu}_1^{0,\text{Lower}} = \frac{.71 + .12}{.71}\left(.18 - \frac{.12}{.71 + .12}0\right) = .21$$

Wrapping up the class exercise: Question 10

Set identification for the causal effect

[1] (This is of course subject to the doubtful assumptions made in the problem, but even under those strong assumptions the interval is very wide.)

Wrapping up the class exercise: Question 10

Set identification for the causal effect

The causal effect estimate is set-identified by

$$\hat{\tau}_1^{\text{Lower}} = \hat{\mu}_1^1 - \hat{\mu}_1^{0,\text{Upper}} \quad = .40 - .21 = .19$$
$$\hat{\tau}_1^{\text{Upper}} = \hat{\mu}_1^1 - \hat{\mu}_1^{0,\text{Lower}} \quad = .40 - .04 = .36$$

[1](This is of course subject to the doubtful assumptions made in the problem, but even under those strong assumptions the interval is very wide.)

Wrapping up the class exercise: Question 10

Set identification for the causal effect

The causal effect estimate is set-identified by

$$\begin{aligned}\hat{\tau}_1^{\text{Lower}} &= \hat{\mu}_1^1 - \hat{\mu}_1^{0,\text{Upper}} &&= .40 - .21 = .19 \\ \hat{\tau}_1^{\text{Upper}} &= \hat{\mu}_1^1 - \hat{\mu}_1^{0,\text{Lower}} &&= .40 - .04 = .36\end{aligned}$$

Among women who would have a child regardless of their own education,

stratum-specific

[1] (This is of course subject to the doubtful assumptions made in the problem, but even under those strong assumptions the interval is very wide.)

Wrapping up the class exercise: Question 10

Set identification for the causal effect

The causal effect estimate is set-identified by

$$\begin{aligned}
\hat{\tau}_1^{\text{Lower}} &= \hat{\mu}_1^1 - \hat{\mu}_1^{0,\text{Upper}} &&= .40 - .21 = .19 \\
\hat{\tau}_1^{\text{Upper}} &= \hat{\mu}_1^1 - \hat{\mu}_1^{0,\text{Lower}} &&= .40 - .04 = .36
\end{aligned}$$

Among women who would have a child regardless of their own education,	stratum- specific
the effect of a mother finishing college on the probability that her child finishes college	causal effect

[1](This is of course subject to the doubtful assumptions made in the problem, but even under those strong assumptions the interval is very wide.)

Wrapping up the class exercise: Question 10

Set identification for the causal effect

The causal effect estimate is set-identified by

$$\begin{aligned}\hat{\tau}_1^{\text{Lower}} &= \hat{\mu}_1^1 - \hat{\mu}_1^{0,\text{Upper}} &&= .40 - .21 = .19\\ \hat{\tau}_1^{\text{Upper}} &= \hat{\mu}_1^1 - \hat{\mu}_1^{0,\text{Lower}} &&= .40 - .04 = .36\end{aligned}$$

Among women who would have a child regardless of their own education, — stratum-specific

the effect of a mother finishing college on the probability that her child finishes college — causal effect

is somewhere between 0.19 and 0.36.[1] — set-identified

[1](This is of course subject to the doubtful assumptions made in the problem, but even under those strong assumptions the interval is very wide.)

Big picture ideas in this exercise

Big picture ideas in this exercise

- Principal stratification solved a hard problem
  - A mediator that can render an outcome undefined

Big picture ideas in this exercise

- Principal stratification solved a hard problem
  - A mediator that can render an outcome undefined
- You can bound estimates by assuming monotonicity (and applying lots of algebra)

Learning goals for today

At the end of class, you will be able to:

1. Finish the class exercise we started on Tuesday [solutions]
2. See principal stratification in action:
   quantifying racial bias in policing

American Political Science Review (2020) 114, 3, 619–637
doi:10.1017/S0003055420000039

Administrative Records Mask Racially Biased Policing

DEAN KNOX *Princeton University*
WILL LOWE *Hertie School of Governance*
JONATHAN MUMMOLO *Princeton University*

A police officer encounters a person

A police officer encounters a person

1. Stop them? Or not?

A police officer encounters a person

1. Stop them? Or not?
2. Use force? Or not?

A police officer encounters a person

1. Stop them? Or not?
2. Use force? Or not?

Effect of race:
Would the outcome of this encounter differ
if the civilian were of a different race

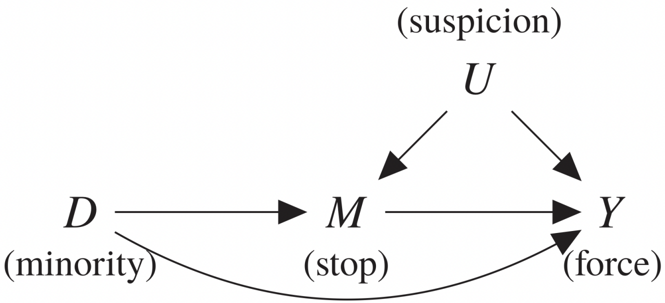
A police officer encounters a person

1. Stop them? Or not?
2. Use force? Or not?

Effect of race:
Would the outcome of this encounter differ
if the civilian were of a different race

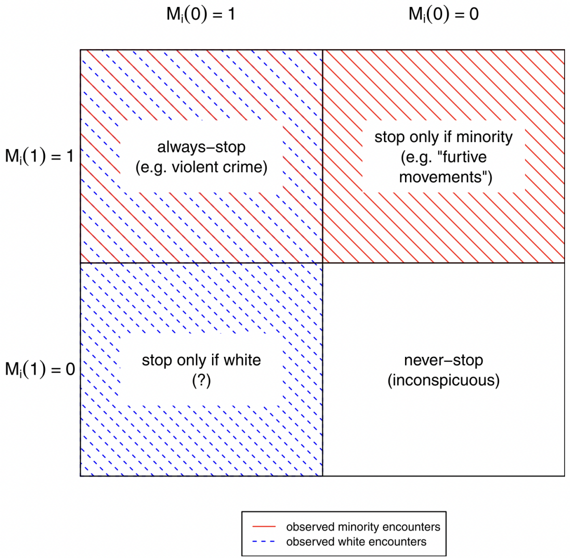
Unit of analysis is an **encounter** not a **person**

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police



Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

FIGURE 2. Principal Strata and Observed Police–Civilian Encounters



We would want the ATE

$$E(Y^{1M^1} - Y^{0M^0})$$

To estimate that, the authors say we need two things

1. Count of minority encounters[2]
2. Count of white encounters

within strata of X

[2] (including all four strata)

Point estimates

Note: All steps are within X. Notation dropped.

Point estimates

Note: All steps are within X. Notation dropped.

Important caveat:

The following is my reconstruction of one of the simplest of many results in Knox, Lowe, & Mummolon 2020.

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$E(Y^1)$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$E(Y^1) = E(Y^1 \mid D = 1)$ Exchangeability

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \mathrm{E}(Y^1 \mid D = 1) \qquad \text{Exchangeability}$$

$$= \mathrm{E}(Y \mid D = 1) \qquad \text{Consistency}$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \mathrm{E}(Y^1 \mid D = 1) \qquad \text{Exchangeability}$$

$$= \mathrm{E}(Y \mid D = 1) \qquad \text{Consistency}$$

$$= \overbrace{\mathrm{P}(M = 1 \mid D = 1)}\,\overbrace{\mathrm{E}(Y \mid D = 1, M = 1)} + \underbrace{\mathrm{P}(M = 0 \mid D = 1)}\quad\underbrace{\mathrm{E}(Y \mid D = 1, M = 0)} \qquad \text{Law of Total Probability}$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$
\begin{aligned}
\mathrm{E}(Y^1) &= \mathrm{E}(Y^1 \mid D = 1) && \text{Exchangeability} \\
&= \mathrm{E}(Y \mid D = 1) && \text{Consistency} \\
&= \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)} && \text{Law of Total} \\
&\quad + \underbrace{\mathrm{P}(M = 0 \mid D = 1)} \quad \underbrace{\mathrm{E}(Y \mid D = 1, M = 0)} && \text{Probability}
\end{aligned}
$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1) \qquad \text{Exchangeability}$$

$$= E(Y \mid D = 1) \qquad \text{Consistency}$$

$$= \overbrace{P(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{E(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}} + \underbrace{P(M = 0 \mid D = 1)} \quad \underbrace{E(Y \mid D = 1, M = 0)} \qquad \text{Law of Total Probability}$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$E(Y^1) = E(Y^1 \mid D = 1) \qquad \text{Exchangeability}$$

$$= E(Y \mid D = 1) \qquad \text{Consistency}$$

$$\begin{aligned} &= \overbrace{P(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{E(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}} \\ &\quad + \underbrace{P(M = 0 \mid D = 1)}_{\substack{\text{Non-stop rate among} \\ \text{minority encounters}}} \underbrace{E(Y \mid D = 1, M = 0)} \end{aligned} \qquad \text{Law of Total Probability}$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$
\begin{aligned}
\mathrm{E}(Y^1) &= \mathrm{E}(Y^1 \mid D = 1) && \text{Exchangeability} \\
&= \mathrm{E}(Y \mid D = 1) && \text{Consistency} \\
&= \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}} && \text{Law of Total} \\
&\quad + \underbrace{\mathrm{P}(M = 0 \mid D = 1)}_{\substack{\text{Non-stop rate among} \\ \text{minority encounters}}} \underbrace{\mathrm{E}(Y \mid D = 1, M = 0)}_{\substack{\text{Use of force among} \\ \text{non-stopped minority encounters} \\ (=0)}} && \text{Probability}
\end{aligned}
$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \mathrm{E}(Y^1 \mid D = 1) \qquad \text{Exchangeability}$$

$$= \mathrm{E}(Y \mid D = 1) \qquad \text{Consistency}$$

$$= \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}} + \underbrace{\mathrm{P}(M = 0 \mid D = 1)}_{\substack{\text{Non-stop rate among} \\ \text{minority encounters}}} \underbrace{\mathrm{E}(Y \mid D = 1, M = 0)}_{\substack{\text{Use of force among} \\ \text{non-stopped minority encounters} \\ (=0)}} \qquad \text{Law of Total Probability}$$

$$= \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}}$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathsf{E}(Y^1) = \overbrace{\mathsf{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathsf{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}}$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}}$$

vs if they involved a non-minority civilian?

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \overbrace{\mathrm{P}(M=1 \mid D=1)}^{\substack{\text{Stop rate among}\\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D=1, M=1)}^{\substack{\text{Use of force among}\\ \text{stopped minority encounters}}}$$

vs if they involved a non-minority civilian?

$$\mathrm{E}(Y^0) = \overbrace{\mathrm{P}(M=1 \mid D=0)}^{\substack{\text{Stop rate among}\\ \text{non-minority encounters}}} \overbrace{\mathrm{E}(Y \mid D=0, M=1)}^{\substack{\text{Use of force among}\\ \text{stopped non-minority encounters}}}$$

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}}$$

vs if they involved a non-minority civilian?

$$\mathrm{E}(Y^0) = \overbrace{\mathrm{P}(M = 1 \mid D = 0)}^{\substack{\text{Stop rate among} \\ \text{non-minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 0, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped non-minority encounters}}}$$

Difference is the ATE.

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}}$$

vs if they involved a non-minority civilian?

$$\mathrm{E}(Y^0) = \overbrace{\mathrm{P}(M = 1 \mid D = 0)}^{\substack{\text{Stop rate among} \\ \text{non-minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 0, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped non-minority encounters}}}$$

Difference is the ATE.

You just needed to augment the data with stop rates!

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}}$$

vs if they involved a non-minority civilian?

$$\mathrm{E}(Y^0) = \overbrace{\mathrm{P}(M = 1 \mid D = 0)}^{\substack{\text{Stop rate among} \\ \text{non-minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 0, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped non-minority encounters}}}$$

Difference is the ATE.

You just needed to augment the data with stop rates!

Works because of two key factors:

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \overbrace{\mathrm{P}(M=1 \mid D=1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D=1, M=1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}}$$

vs if they involved a non-minority civilian?

$$\mathrm{E}(Y^0) = \overbrace{\mathrm{P}(M=1 \mid D=0)}^{\substack{\text{Stop rate among} \\ \text{non-minority encounters}}} \overbrace{\mathrm{E}(Y \mid D=0, M=1)}^{\substack{\text{Use of force among} \\ \text{stopped non-minority encounters}}}$$

Difference is the ATE.

You just needed to augment the data with stop rates!

Works because of two key factors:

- Race is assumed exchangeable given X

Point estimates

Note: All steps are within X. Notation dropped.

What proportion of encounters would involve force if they involved a minority civilian?

$$\mathrm{E}(Y^1) = \overbrace{\mathrm{P}(M = 1 \mid D = 1)}^{\substack{\text{Stop rate among} \\ \text{minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 1, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped minority encounters}}}$$

vs if they involved a non-minority civilian?

$$\mathrm{E}(Y^0) = \overbrace{\mathrm{P}(M = 1 \mid D = 0)}^{\substack{\text{Stop rate among} \\ \text{non-minority encounters}}} \overbrace{\mathrm{E}(Y \mid D = 0, M = 1)}^{\substack{\text{Use of force among} \\ \text{stopped non-minority encounters}}}$$

Difference is the ATE.

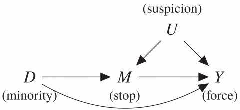
You just needed to augment the data with stop rates!

Works because of two key factors:

- Race is assumed exchangeable given X
- When $M = 0$ (no stop), then $Y = 0$ (no force)

Many possible estimands

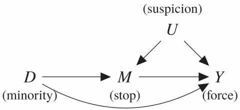
FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police



Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

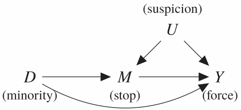


Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

- ATE: $\text{E}(Y^{1M^1} - Y^{0M^0})$

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

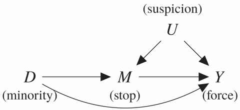


Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

- ATE: $\mathrm{E}(Y^{1M^1} - Y^{0M^0})$
  - Racial bias, where non-stops are coded $Y = 0$

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

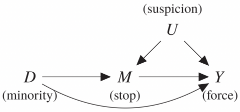


Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

- ATE: $\mathrm{E}(Y^{1M^1} - Y^{0M^0})$
  - Racial bias, where non-stops are coded $Y = 0$
- CDE: $\mathrm{E}(Y^{11} - Y^{01})$

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

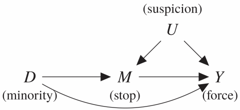


Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

- ATE: $\mathrm{E}(Y^{1M^1} - Y^{0M^0})$
  - Racial bias, where non-stops are coded $Y = 0$
- CDE: $\mathrm{E}(Y^{11} - Y^{01})$
  - Racial bias if we stopped everyone

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

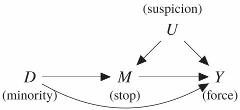


Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

- ATE: $\mathrm{E}(Y^{1M^1} - Y^{0M^0})$
  - Racial bias, where non-stops are coded $Y = 0$
- CDE: $\mathrm{E}(Y^{11} - Y^{01})$
  - Racial bias if we stopped everyone
- ATE among the stopped

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

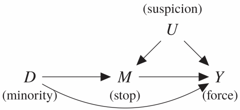


Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

- ATE: $\mathrm{E}(Y^{1M^1} - Y^{0M^0})$
  - Racial bias, where non-stops are coded $Y = 0$
- CDE: $\mathrm{E}(Y^{11} - Y^{01})$
  - Racial bias if we stopped everyone
- ATE among the stopped
  - $\mathrm{ATE}_{M=1} = \mathrm{E}(Y^{1M^1} \mid M = 1) - \mathrm{E}(Y^{0M^0} \mid M = 1)$

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police

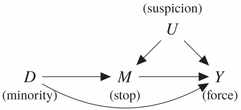


Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

- ATE: $\mathrm{E}(Y^{1M^1} - Y^{0M^0})$
  - Racial bias, where non-stops are coded $Y = 0$
- CDE: $\mathrm{E}(Y^{11} - Y^{01})$
  - Racial bias if we stopped everyone
- ATE among the stopped
  - $\text{ATE}_{M=1} = \mathrm{E}(Y^{1M^1} \mid M = 1) - \mathrm{E}(Y^{0M^0} \mid M = 1)$
- Proportion of minority stops due to race

Many possible estimands

FIGURE 1. Directed Acyclic Graph of Racial Discrimination in the Use of Force by Police



Notes: Observed *X* is left implicit; these covariates may be causally prior to any subset of *D*, *M*, and *Y*.

- ATE: $E(Y^{1M^1} - Y^{0M^0})$
  - Racial bias, where non-stops are coded $Y = 0$
- CDE: $E(Y^{11} - Y^{01})$
  - Racial bias if we stopped everyone
- ATE among the stopped
  - $\text{ATE}_{M=1} = E(Y^{1M^1} \mid M = 1) - E(Y^{0M^0} \mid M = 1)$
- Proportion of minority stops due to race
  - $E(Y^{1M^1} - Y^{0M^0} \mid D = 1, M = 1)$

Many estimands: Necessary Assumptions

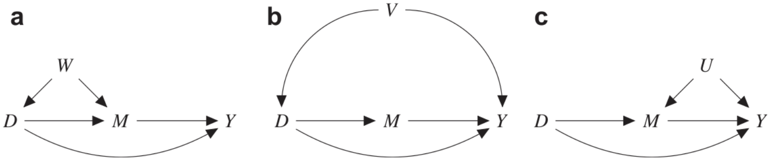
1. Mandatory reporting: $Y_i^{d0} = 0$ for all i and d
2. Mediator monotonicity: $M_i^1 \geq M_i^0$
3. Relative nonseverity of racial stops:

$$\mathrm{E}(Y^{dm} \mid D = d', \overbrace{M^1 = 1, M^0 = 1}^{\text{Always Stop Stratum}}, X)$$
$$\geq \mathrm{E}(Y^{dm} \mid D = d', \underbrace{M^1 = 1, M^0 = 0}_{\text{Racial Stop Stratum}}, X)$$

4. Treatment ignorability
   - $M^d \perp\!\!\!\perp D \mid X$
   - $Y^{dm} \perp\!\!\!\perp D \mid M^0, M^1, X$

Many Estimands: Necessary Assumptions

Assume absence of W and V. Ok to have U.



Many Estimands: Strong (As In Doubtful) Assumptions

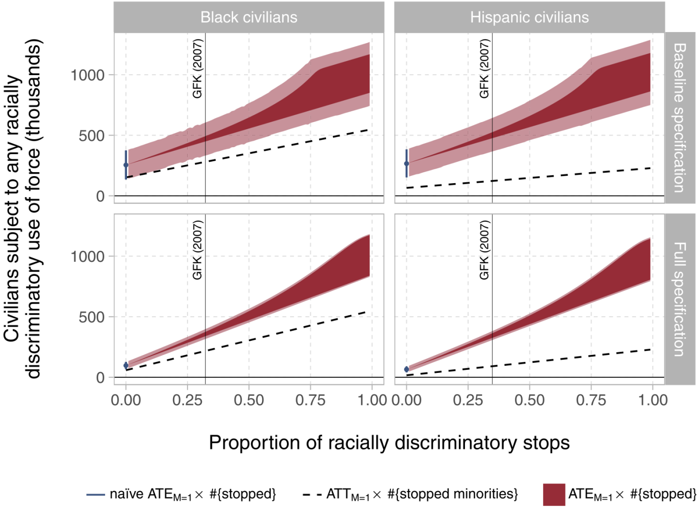
Studies about the effect of race conditional
on an interaction implicitly assume these things:

1. Mediator ignorability: $Y^{dm} \perp\!\!\!\perp M^0 \mid D = d, M^1 = 1, X$
   - "violence rates in always-stop encounters must be identical to those in observationally equivalent racial stops"
2. No racial stops: $M^0 = M^1 \mid M = 1$
   - "all reported encounters were of the always-stop kind"

Knox, Lowe, & Mummolo argue that the above are implausible assumptions in the context of policing.

Without the strong assumptions, things can be learned

FIGURE 4. Bounds for Racially Discriminatory Use of Force, any Severity



Learning goals for today

At the end of class, you will be able to:

1. Finish the class exercise we started on Tuesday [solutions]
2. See principal stratification in action:
   quantifying racial bias in policing

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!