6. Population inference from samples

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

8 Sep 2022

Comments on Problem Set 1

Definition of potential outcomes

- $\{Y_i^1, Y_i^0\}$ are potential outcomes.
  When $A_i = 1$, then Y_i^1 is factual and Y_i^0 is counterfactual.
  When $A_i = 0$, then these are reversed.
  This is why potential, not necessarily counterfactual.
- Y^a is the outcome of a randomly sampled unit assigned to treatment value a. In itself, it is not an average over a group—that would be $\mathrm{E}(Y^a)$.

Comments on Problem Set 1

$\mathrm{E}(Y \mid A = 1) > \mathrm{E}(Y \mid A = 0)$

- Descriptive
- Outcomes were higher, on average, for those who got the treatment

$\mathrm{E}(Y^1) > \mathrm{E}(Y^0)$

- Causal
- The treatment (1 vs 0) increases outcomes, on average

$Y_i^1 > Y_i^0$ for all i

- Causal
- The treatment (1 vs 0) increases the outcome for every unit

Comments on Problem Set 1

Observational Claims	Causal Claims
Observational Evidence	Causal Evidence

Comments on Problem Set 1

Observational Claims	Causal Claims
Observational Evidence	~~Causal Evidence~~

Comments on Problem Set 1

Observational Claims	Causal Claims
Observational Evidence	~~Causal Evidence~~

"...all causal inference is based on assumptions that cannot be derived from observations alone," (Greenland, Pearl, & Robins 1999, p. 47)

Comments on Problem Set 1

Observational Claims	Causal Claims
Observational Evidence	~~Causal Evidence~~

"...all causal inference is based on assumptions that cannot be derived from observations alone," (Greenland, Pearl, & Robins 1999, p. 47)

There is no causal evidence.
There is only observational evidence,
which speaks to causal claims under assumptions.

Learning goals for today

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

1. Understand DAGs more fully
   - DAGs are nonparametric
   - DAGs are hard to learn from data
2. Generalize from a sample to a population
   - Encode sampling assumptions in DAGs

DAGs are nonparametric

DAGs are nonparametric

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

DAGs are nonparametric

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

This does **not** mean

$$Y = \beta_0 + \beta_1 X + \beta_2 A + \epsilon$$

DAGs are nonparametric

$$X \longrightarrow A \longrightarrow Y$$

This does **not** mean

$$Y = \beta_0 + \beta_1 X + \beta_2 A + \epsilon$$

This **does** mean

- $A = f(X, U_A)$ for some function $f()$
- $Y = g(X, A, U_Y)$ for some function $g()$

DAGs are nonparametric

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \\ U_X \arrow[u] & U_A \arrow[u] & U_Y \arrow[u] \end{tikzcd}$$

This does **not** mean

$$Y = \beta_0 + \beta_1 X + \beta_2 A + \epsilon$$

This does **does** mean

- $A = f(X, U_A)$ for some function $f()$
- $Y = g(X, A, U_Y)$ for some function $g()$

DAGs are nonparametric

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \\ U_X \arrow[u] & U_A \arrow[u] & U_Y \arrow[u] \end{tikzcd}$$

This does **not** mean

$$Y = \beta_0 + \beta_1 X + \beta_2 A + \epsilon$$

This **does** mean

- $A = f(X, U_A)$ for some function $f()$
- $Y = g(X, A, U_Y)$ for some function $g()$

which allows that

- The effect of A may depend on X (heterogeneity)
- $\mathrm{E}(Y \mid X, A)$ may be a nonlinear function of each input

DAGs are nonparametric: Why this is really great

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

DAGs are nonparametric: Why this is really great

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

This tells us:

$$\underbrace{\mathrm{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}$$

DAGs are nonparametric: Why this is really great

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

This tells us:

$$\underbrace{\mathrm{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}$$

Left statement:

- Among everyone with $X = x$,
- the average outcome if we set $A = a$

DAGs are nonparametric: Why this is really great

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

This tells us:

$$\underbrace{\mathrm{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}$$

Left statement:

- Among everyone with $X = x$,
- the average outcome if we set $A = a$

Right statement:

- Among everyone with $X = x$ and $A = a$,
- the average outcome

DAGs are nonparametric: Why this is really great

$$X \longrightarrow A \longrightarrow Y$$

This tells us:

$$\underbrace{\mathrm{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}$$

Left statement:

- Among everyone with $X = x$,
- the average outcome if we set $A = a$

Right statement:

- Among everyone with $X = x$ and $A = a$,
- the average outcome

These are two **different sets** of people

DAGs are nonparametric: Why this is really great

$$X \longrightarrow A \longrightarrow Y$$

This tells us:

$$\underbrace{\mathsf{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathsf{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}$$

DAGs are nonparametric: Why this is really great

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

This tells us:

$$\underbrace{\mathsf{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathsf{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}$$

Once the DAG gives us the above,

DAGs are nonparametric: Why this is really great

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

This tells us:

$$\underbrace{\mathrm{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}$$

Once the DAG gives us the above,
we can use **any** prediction function

DAGs are nonparametric: Why this is really great

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

This tells us:

$$\underbrace{\mathrm{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}$$

Once the DAG gives us the above,
we can use **any** prediction function
for the statistical part.

DAGs are nonparametric: Why this is really great

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

This contrasts with standard econometrics

DAGs are nonparametric: Why this is really great

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

This contrasts with standard econometrics

$$Y = \alpha + \beta_1 X + \gamma A + \eta XA + \epsilon$$

DAGs are nonparametric: Why this is really great

$$
\begin{tikzcd}
X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y
\end{tikzcd}
$$

This contrasts with standard econometrics

$$Y = \alpha + \beta_1 X + \gamma A + \eta XA + \epsilon$$

- $\{\beta, \gamma\}$ are "main effects"

DAGs are nonparametric: Why this is really great

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

This contrasts with standard econometrics

$$Y = \alpha + \beta_1 X + \gamma A + \eta XA + \epsilon$$

- $\{\beta, \gamma\}$ are "main effects"
- η is an "interaction": the effect of A varies by X

DAGs are nonparametric: Why this is really great

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}$$

This contrasts with standard econometrics

$$Y = \alpha + \beta_1 X + \gamma A + \eta XA + \epsilon$$

- $\{\beta, \gamma\}$ are "main effects"
- η is an "interaction": the effect of A varies by X
- Key assumption: $A \perp\!\!\!\perp \epsilon$, or A is "exogenous"

That requires us to do **both** causal reasoning **and** statistical reasoning simultaneously.

DAGs support causal reasoning **before** statistical reasoning

DAGs are nonparametric: Why this is really great

$$
\begin{tikzcd} X \arrow[r] \arrow[rr, bend left] & A \arrow[r] & Y \end{tikzcd}
$$

$$
\underbrace{\mathrm{E}(Y^a \mid X = x)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid A = a, X = x)}_{\text{Statistical Quantity}}
$$

Let's pause to discuss this.

Learning goals for today

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

1. Understand DAGs more fully
   - DAGs are nonparametric
   - DAGs are hard to learn from data
2. Generalize from a sample to a population
   - Encode sampling assumptions in DAGs

Causal Discovery[1]: DAGs are hard to learn from data

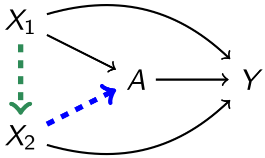
[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[1]: DAGs are hard to learn from data

$$
\begin{tikzcd}
X_1 \arrow[dr] \arrow[drr, bend left] & & \\
& A \arrow[r] & Y \\
X_2 \arrow[urr, bend right] & &
\end{tikzcd}
$$

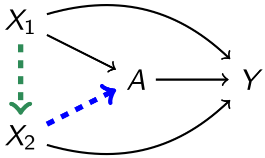
[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[1]: DAGs are hard to learn from data



[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

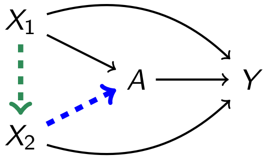
Causal Discovery[1]: DAGs are hard to learn from data



Can data tell us whether the dashed edges exist?

[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[1]: DAGs are hard to learn from data

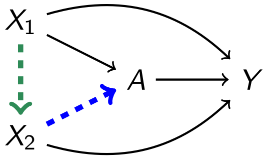


Can data tell us whether the dashed edges exist?

- In the absence of both edges,

[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[1]: DAGs are hard to learn from data

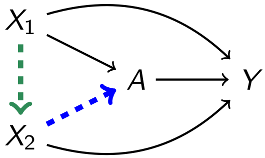


Can data tell us whether the dashed edges exist?

- In the absence of both edges, $X_1 \perp\!\!\!\perp X_2$ and $X_2 \perp\!\!\!\perp A$

[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[1]: DAGs are hard to learn from data

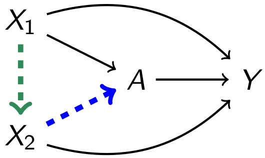


Can data tell us whether the dashed edges exist?

- In the absence of both edges, $X_1 \perp\!\!\!\perp X_2$ and $X_2 \perp\!\!\!\perp A$
- In the absence of the blue edge,

[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[1]: DAGs are hard to learn from data

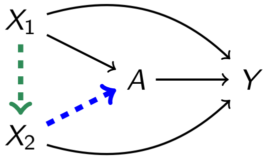


Can data tell us whether the dashed edges exist?

- In the absence of both edges, $X_1 \perp\!\!\!\perp X_2$ and $X_2 \perp\!\!\!\perp A$
- In the absence of the blue edge, $X_2 \perp\!\!\!\perp A \mid X_1$

[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[1]: DAGs are hard to learn from data

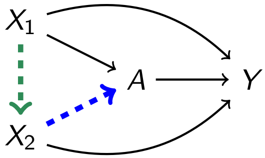


Can data tell us whether the dashed edges exist?

- In the absence of both edges, $X_1 \perp\!\!\!\perp X_2$ and $X_2 \perp\!\!\!\perp A$
- In the absence of the blue edge, $X_2 \perp\!\!\!\perp A \mid X_1$
- In the absence of the green edge,

[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[1]: DAGs are hard to learn from data



Can data tell us whether the dashed edges exist?

- In the absence of both edges, $X_1 \perp\!\!\!\perp X_2$ and $X_2 \perp\!\!\!\perp A$
- In the absence of the blue edge, $X_2 \perp\!\!\!\perp A \mid X_1$
- In the absence of the green edge, $X_1 \perp\!\!\!\perp X_2$ and $X_1 \not\!\perp\!\!\!\perp X_2 \mid A$

[1] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[2]: DAGs are hard to learn from data

Will data replace human researchers?

[2] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[2]: DAGs are hard to learn from data

Will data replace human researchers?

I think not.

[2] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[2]: DAGs are hard to learn from data

Will data replace human researchers?

I think not.

Often, what we want to know cannot be answered by the data.

[2] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

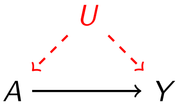
Causal Discovery[2]: DAGs are hard to learn from data

Will data replace human researchers?

I think not.

Often, what we want to know cannot be answered by the data.

Example: Does the unobserved U confound treatment assignment?



[2] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

Causal Discovery[2]: DAGs are hard to learn from data

Will data replace human researchers?

I think not.

Often, what we want to know cannot be answered by the data.

Example: Does the unobserved U confound treatment assignment?

$$\begin{tikzcd}
& \color{red}{U} \arrow[dl, dashed, red] \arrow[dr, dashed, red] & \\
A \arrow[rr] & & Y
\end{tikzcd}$$

A and Y are associated either way.

[2] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

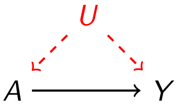
Causal Discovery[2]: DAGs are hard to learn from data

Will data replace human researchers?

I think not.

Often, what we want to know cannot be answered by the data.

Example: Does the unobserved U confound treatment assignment?



A and Y are associated either way.
The absence of U is a completely untestable assumption.

[2] See Spirtes, P., Glymour, C. N., Scheines, R., & Heckerman, D. (2000). Causation, Prediction, and Search. MIT Press.

As a general rule, the DAG encodes
substantive theory (made by a human)

rather than **data** (crunched by a computer)

Some academic history of DAGs

- Historical roots in path models in the 1920s
  - Wright, S. (1921). Correlation and causation. Part I: Method of path coefficients. Journal of Agricultural Research, 20(7), 557-585.
- Linear path models in the 1960s
  - Duncan, O. D. (1966). Path analysis: Sociological examples. American Journal of Sociology, 72(1), 1-16.
- Landmark contributions: Pearl, Greenland, Robins
  - **(assigned)** Greenland, S., Pearl, J., & Robins, J. M. (1999). Causal diagrams for epidemiologic research. Epidemiology, 37-48.
  - Pearl, J. (2000). Causality. Cambridge University Press.
  - Pearl, J., & Mackenzie, D. (2018). The Book of Why: The New Science of Cause and Effect. Basic Books.
- More accessible introduction for social scientists
  - Morgan, S. L., & Winship, C. (2015). Counterfactuals and Causal Inference. Cambridge University Press.

Learning goals for today

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

1. Understand DAGs more fully
   - DAGs are nonparametric
   - DAGs are hard to learn from data
2. Generalize from a sample to a population
   - Encode sampling assumptions in DAGs

Sample $\rightarrow$ Population

Fun example

Wang, Rothschild, Goel, & Gelman

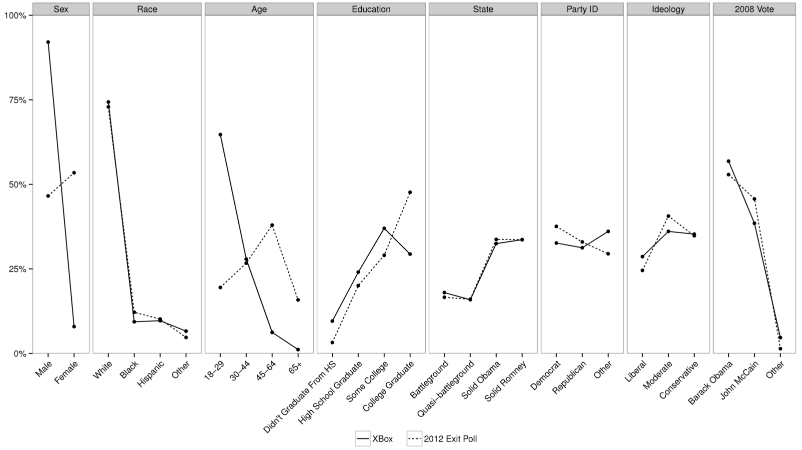
Survey of **Xbox users** to forecast the 2012 election![3]

---

[3]Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. International Journal of Forecasting, 31(3), 980-991.

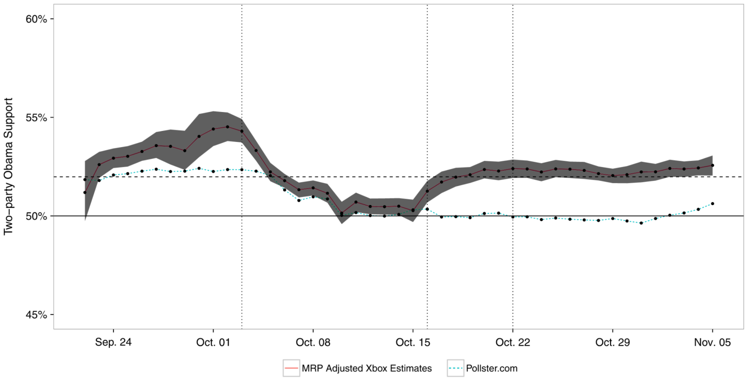
Fun example

W. Wang et al. / International Journal of Forecasting 31 (2015) 980–991



Fun example

W. Wang et al. / International Journal of Forecasting 31 (2015) 980–991



Today we will formalize the conditions under which this works

Imagine a study:

Imagine a study:

- We randomly sample 1,000 voters from the U.S. voter file.

Imagine a study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **whatever it takes**

Imagine a study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **whatever it takes**
- **100%** respond.

Imagine a study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **whatever it takes**
- **100%** respond.
- We ask them a question:

Imagine a study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **whatever it takes**
- **100%** respond.
- We ask them a question:
  - Was Barack Obama the best president of the past 20 years?

Imagine a study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **whatever it takes**
- **100%** respond.
- We ask them a question:
  - Was Barack Obama the best president of the past 20 years?

Can we draw conclusions about the population of U.S. voters?

Imagine a study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **whatever it takes**
- **100%** respond.
- We ask them a question:
  - Was Barack Obama the best president of the past 20 years?

Can we draw conclusions about the population of U.S. voters?

Yes! A probability sample

Imagine another study:

Imagine another study:

- We randomly sample 1,000 voters from the U.S. voter file.

Imagine another study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **$500** to participate.
- Only **90%** respond.

Imagine another study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **$500** to participate.
- Only **90%** respond.
- We ask them a question:
  - Was Barack Obama the best president of the past 20 years?

Imagine another study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **\$500** to participate.
- Only **90%** respond.
- We ask them a question:
  - Was Barack Obama the best president of the past 20 years?

Can we draw conclusions about the population of U.S. voters?

Imagine another study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We pay them **\$500** to participate.
- Only **90%** respond.
- We ask them a question:
    - Was Barack Obama the best president of the past 20 years?

Can we draw conclusions about the population of U.S. voters?

Iffy. Almost a probability sample

Imagine another study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We can only pay them **$5** to participate.
- Only **10%** respond.
- We ask them a question:
  - Was Barack Obama the best president of the past 20 years?

Can we draw conclusions about the population of U.S. voters?

Imagine another study:

- We randomly sample 1,000 voters from the U.S. voter file.
- We can only pay them **\$5** to participate.
- Only **10%** respond.
- We ask them a question:
  - Was Barack Obama the best president of the past 20 years?

Can we draw conclusions about the population of U.S. voters?

Big worry:

Do we believe that selection into the sample is independent of Obama support?

$$S = 1$$

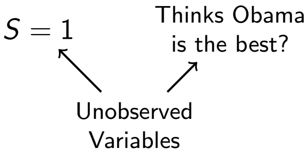
$S = 1$

Thinks Obama
is the best?

$S = 1$

Thinks Obama
is the best?

Unobserved
Variables

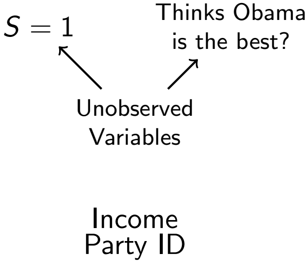

$S = 1$
Thinks Obama is the best?
Unobserved Variables

$S = 1$

Thinks Obama is the best?

Unobserved Variables

Income


$S = 1$
Thinks Obama is the best?
Unobserved Variables
Income
Party ID

$S = 1$

Thinks Obama is the best?

Unobserved Variables

Income
Party ID
Age

$S = 1$

Thinks Obama is the best?

Unobserved Variables

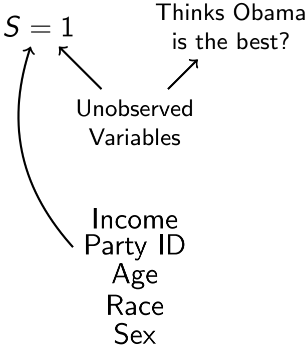
Income
Party ID
Age
Race

$S = 1$

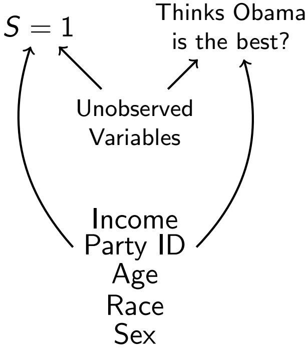
Thinks Obama is the best?

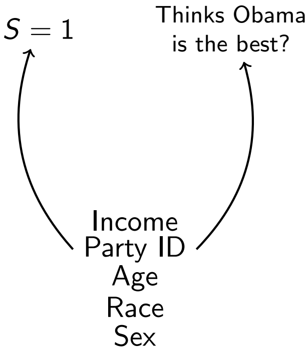
Unobserved Variables

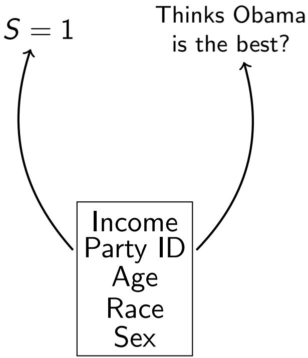
Income
Party ID
Age
Race
Sex



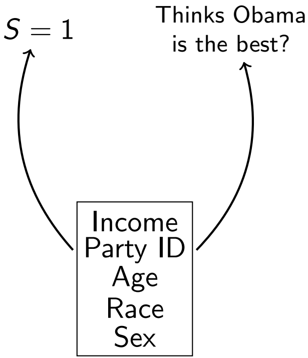
$S = 1$
Thinks Obama is the best?
Unobserved Variables
Income
Party ID
Age
Race
Sex


$S = 1$
Thinks Obama is the best?
Unobserved Variables
Income
Party ID
Age
Race
Sex

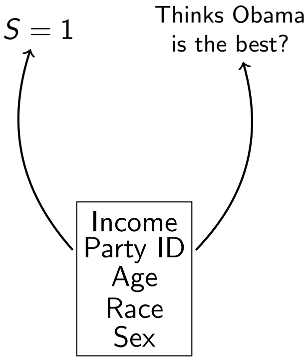

$S = 1$
Thinks Obama
is the best?
Income
Party ID
Age
Race
Sex



$S = 1$
Thinks Obama is the best?
Income
Party ID
Age
Race
Sex

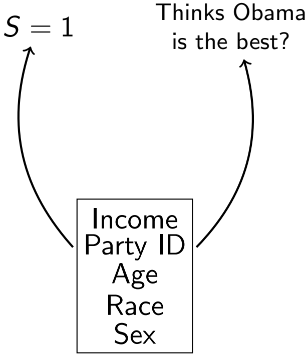


If this is the case:



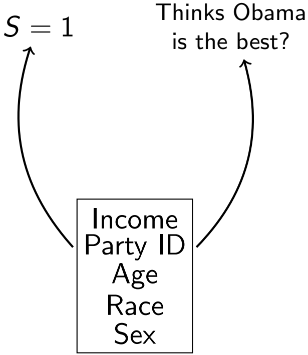
If this is the case:

1. Split into sample subgroups (in sample)



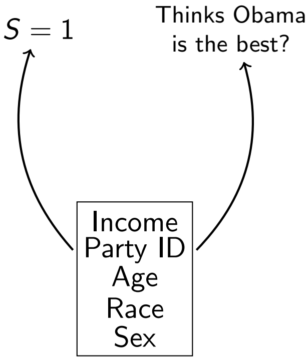
If this is the case:

1. Split into sample subgroups (in sample)
2. Take the mean Obama support (in sample)



If this is the case:

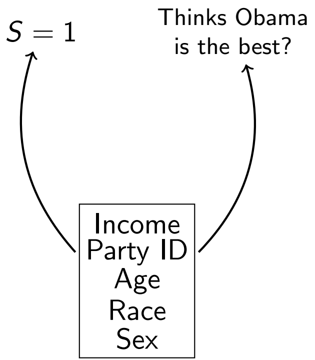
1. Split into sample subgroups (in sample)
2. Take the mean Obama support (in sample)
3. Find each subgroup size in all voter records (in population)



If this is the case:

1. Split into sample subgroups (in sample)
2. Take the mean Obama support (in sample)
3. Find each subgroup size in all voter records (in population)
4. Average over subgroups, weighted by the population size (population estimate!)

Post-Stratification



If this is the case:

1. Split into sample subgroups (in sample)
2. Take the mean Obama support (in sample)
3. Find each subgroup size in all voter records (in population)
4. Average over subgroups, weighted by the population size (population estimate!)

A step further: No probability sample.

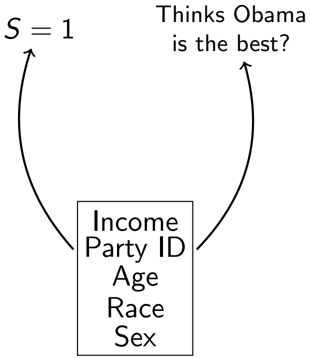
A step further: No probability sample.

We sample random passers-by on the streets of Chicago.

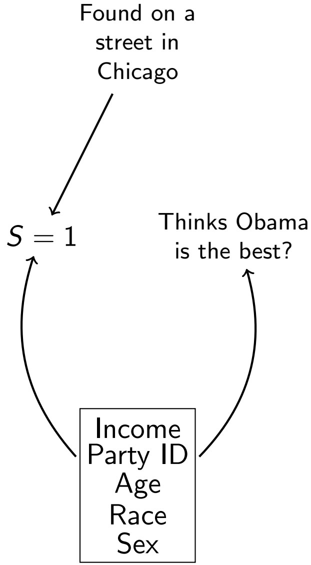
A step further: No probability sample.

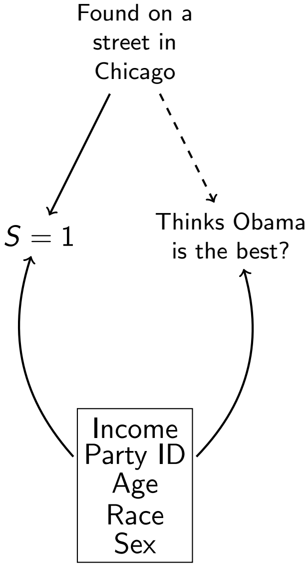
We sample random passers-by on the streets of Chicago.

- Was Barack Obama the best president
  of the past 20 years?



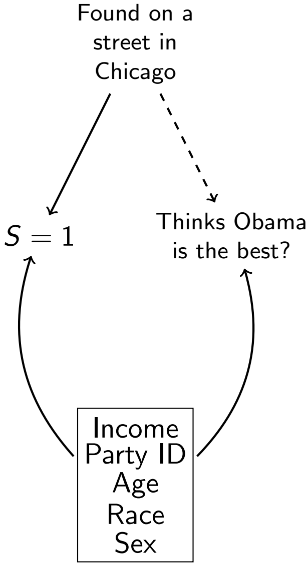
$S = 1$
Thinks Obama is the best?
Income
Party ID
Age
Race
Sex


Found on a street in Chicago
$S = 1$
Thinks Obama is the best?
Income
Party ID
Age
Race
Sex



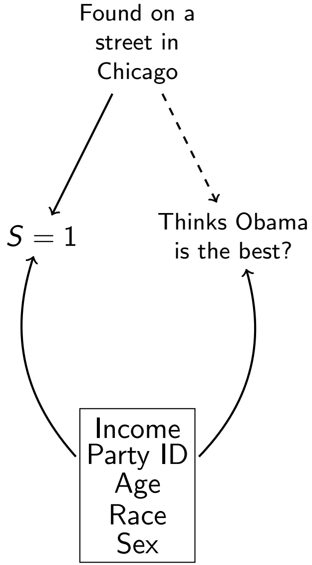
Found on a street in Chicago
$S = 1$
Thinks Obama is the best?
Income
Party ID
Age
Race
Sex

Post-stratification is not a cure-all



Post-stratification is not a cure-all

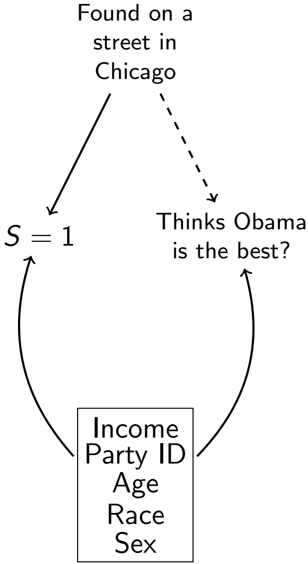
Credibility depends on causal assumptions
— what causes sample inclusion?
— what causes the outcome?
Need conditional independence.



Post-stratification is not a cure-all

Credibility depends on causal assumptions
— what causes sample inclusion?
— what causes the outcome?
Need conditional independence.

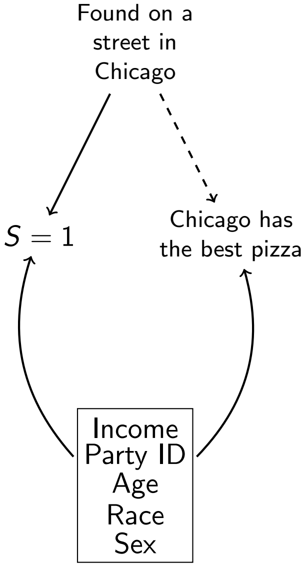
These assumptions belong in a DAG



Post-stratification is not a cure-all

Credibility depends on causal assumptions
— what causes sample inclusion?
— what causes the outcome?
Need conditional independence.

These assumptions belong in a DAG

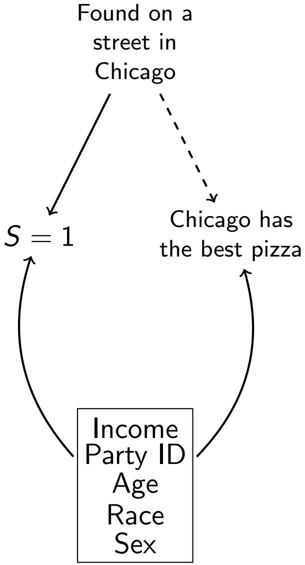


Post-stratification is not a cure-all

Credibility depends on causal assumptions
— what causes sample inclusion?
— what causes the outcome?
Need conditional independence.

These assumptions belong in a DAG

The DAG requires theory
about the particular question



Westreich et al. 2019

Westreich et al. 2019

We often care about **internal validity**

Westreich et al. 2019

We often care about **internal validity**

- Have I identified the causal effect well in my sample?

Westreich et al. 2019

We often care about **internal validity**

- Have I identified the causal effect well in my sample?

and also about **external validity**

Westreich et al. 2019

We often care about **internal validity**

- Have I identified the causal effect well in my sample?

and also about **external validity**

- Does my sample speak to the population of interest?

Westreich et al. 2019

We often care about **internal validity**

- Have I identified the causal effect well in my sample?

and also about **external validity**

- Does my sample speak to the population of interest?

The authors combine these to discuss **target validity**

Westreich et al. 2019, Fig 1 (modified)

Nonexchangeability
for **internal** validity
due to **confounding**

$$
\begin{tikzcd}
 & Z \arrow[dl] \arrow[dr] & \\
A \arrow[rr] & & Y
\end{tikzcd}
$$

Nonexchangeability
for **external** validity
due to **sampling bias**

$$
\begin{tikzcd}
\boxed{S = 1} & & Z \arrow[ll] \arrow[dr] & \\
 & A \arrow[rr] & & Y
\end{tikzcd}
$$

Learning goals for today

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

1. Understand DAGs more fully
   - DAGs are nonparametric
   - DAGs are hard to learn from data
2. Generalize from a sample to a population
   - Encode sampling assumptions in DAGs

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!