Week 8: Regression in the Social Sciences

Brandon Stewart[1]

Princeton

November 7 and 9, 2016

[1]These slides are heavily influenced by Matt Blackwell, Justin Grimmer, Jens Hainmueller, Erin Hartman, Kosuke Imai and Gary King.

Where We've Been and Where We're Going...

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences
  - Wednesday:
    - causal inference

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences
  - Wednesday:
    - causal inference
- Next Week

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences
  - Wednesday:
    - causal inference
- Next Week
  - regression diagnostics

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences
  - Wednesday:
    - causal inference
- Next Week
  - regression diagnostics
- Long Run
  - regression $\rightarrow$ diagnostics $\rightarrow$ causal inference

Questions?

Why Are We Doing All of This Again?

Why Are We Doing All of This Again?

- We are all here because we are trying to do some social science, that is, we are in the business of knowledge production.

Why Are We Doing All of This Again?

- We are all here because we are trying to do some social science, that is, we are in the business of knowledge production.
- Quantitative methods are an increasingly big part of that so whether you are reading or actively doing quantitative analysis it is going to be there.

Why Are We Doing All of This Again?

- We are all here because we are trying to do some social science, that is, we are in the business of knowledge production.
- Quantitative methods are an increasingly big part of that so whether you are reading or actively doing quantitative analysis it is going to be there.
- So why all the math?

Why Are We Doing All of This Again?

- We are all here because we are trying to do some social science, that is, we are in the business of knowledge production.
- Quantitative methods are an increasingly big part of that so whether you are reading or actively doing quantitative analysis it is going to be there.
- So why all the math? We are taking a future-oriented approach. We want to prepare you for the next big thing

Why Are We Doing All of This Again?

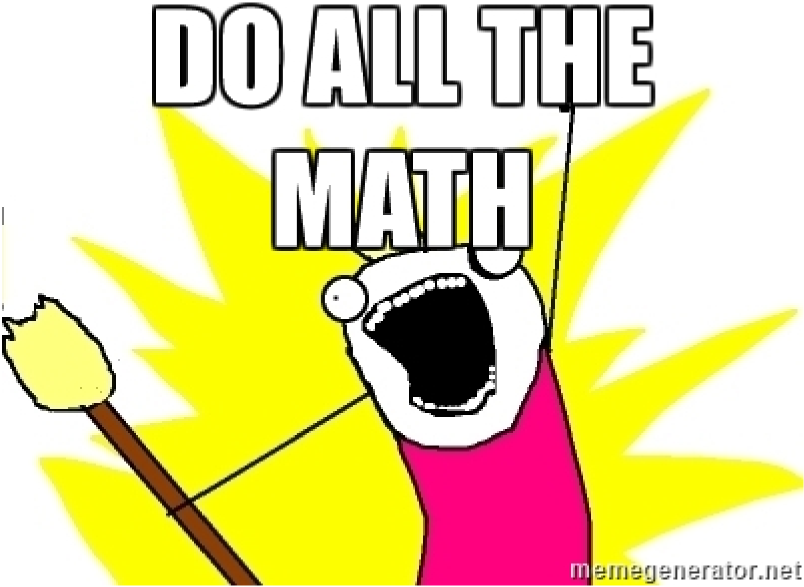
- We are all here because we are trying to do some social science, that is, we are in the business of knowledge production.
- Quantitative methods are an increasingly big part of that so whether you are reading or actively doing quantitative analysis it is going to be there.
- So why all the math? We are taking a future-oriented approach. We want to prepare you for the next big thing
- Methods that became popular in the social sciences since I took the equivalent of this class: machine learning, text-as-data, Bayesian nonparametrics, design-based inference, DAG-based causal inference

Why Are We Doing All of This Again?

- We are all here because we are trying to do some social science, that is, we are in the business of knowledge production.
- Quantitative methods are an increasingly big part of that so whether you are reading or actively doing quantitative analysis it is going to be there.
- So why all the math? We are taking a future-oriented approach. We want to prepare you for the next big thing
- Methods that became popular in the social sciences since I took the equivalent of this class: machine learning, text-as-data, Bayesian nonparametrics, design-based inference, DAG-based causal inference
- A technical foundation prepares you to learn new methods for the rest of your career. Trust me now is the time to invest.

Why Are We Doing All of This Again?

- We are all here because we are trying to do some social science, that is, we are in the business of knowledge production.
- Quantitative methods are an increasingly big part of that so whether you are reading or actively doing quantitative analysis it is going to be there.
- So why all the math? We are taking a future-oriented approach. We want to prepare you for the next big thing
- Methods that became popular in the social sciences since I took the equivalent of this class: machine learning, text-as-data, Bayesian nonparametrics, design-based inference, DAG-based causal inference
- A technical foundation prepares you to learn new methods for the rest of your career. Trust me now is the time to invest.
- Knowing how methods work also makes you a better reader of work.


DO ALL THE
MATH
memegenerator.net

Quantitative Social Science

Quantitative Social Science

- Three components of quantitative social science:

Quantitative Social Science

- Three components of quantitative social science:
  1. Argument
  2. Research Design
  3. Presentation

Quantitative Social Science

- Three components of quantitative social science:
  1. Argument
  2. Research Design
  3. Presentation
- This week we will cover a few issues in these areas:

Quantitative Social Science

- Three components of quantitative social science:
  1. Argument
  2. Research Design
  3. Presentation
- This week we will cover a few issues in these areas:
  - power (research design)

Quantitative Social Science

- Three components of quantitative social science:
  1. Argument
  2. Research Design
  3. Presentation
- This week we will cover a few issues in these areas:
  - power (research design)
  - problems with p-values (argument, design, presentation)

Quantitative Social Science

- Three components of quantitative social science:
  1. Argument
  2. Research Design
  3. Presentation
- This week we will cover a few issues in these areas:
  - power (research design)
  - problems with p-values (argument, design, presentation)
  - visualization and quantities of interest (argument, presentation)

Quantitative Social Science

- Three components of quantitative social science:
  1. Argument
  2. Research Design
  3. Presentation
- This week we will cover a few issues in these areas:
  - power (research design)
  - problems with p-values (argument, design, presentation)
  - visualization and quantities of interest (argument, presentation)
  - identification and causal inference (argument, design)

Quantitative Social Science

- Three components of quantitative social science:
  1. Argument
  2. Research Design
  3. Presentation
- This week we will cover a few issues in these areas:
  - power (research design)
  - problems with p-values (argument, design, presentation)
  - visualization and quantities of interest (argument, presentation)
  - identification and causal inference (argument, design)

We will mostly talk about statistical methods here (it is a statistics class!) but the best work is a combination of substantive and statistical theory.

Are Most Published Findings False?

Are Most Published Findings False?

The backdrop for this is numerous studies with a dim view of the veracity of the average academic article

Are Most Published Findings False?

The backdrop for this is numerous studies with a dim view of the veracity of the average academic article

- Ioannidis (2005) "Why most published research findings are false" *PLOS Medicine*
- Begley and Ellis (2012) "Drug development: raise standard for preclinical cancer research" *Nature*
- Johnson (2013) "Revised standards for statistical evidence" *PNAS*
- Franco, Malhotra, Simonovits (2014) "Publication Bias in the Social Sciences: Unlocking the File Drawer" *Science*
- Nosek et al (2015) "Estimating the reproducibility of psychological science" *Science*
- Leek and Jager (2017) "Is Most Published Research Really False?" *Annual Review of Statistics and Its Applications*

An Example Gerber, Green and Larimer (2008)

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

- Matt Salganik in his book *Bit by Bit* argues that it is an ethical imperative to only use as many subjects as we need in an experiment

An Example Gerber, Green and Larimer (2008)

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

- Matt Salganik in his book *Bit by Bit* argues that it is an ethical imperative to only use as many subjects as we need in an experiment
- Why did GGL use sample sizes of 38,000? Could they have used fewer? How would we know?

An Example Gerber, Green and Larimer (2008)

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

- Matt Salganik in his book *Bit by Bit* argues that it is an ethical imperative to only use as many subjects as we need in an experiment
- Why did GGL use sample sizes of 38,000? Could they have used fewer? How would we know?
- We choose the sample size that can ensure that we can detect what we think is the true treatment effect (i.e. reject the null of no effect).

An Example Gerber, Green and Larimer (2008)

TABLE 2. Effects of Four Mail Treatments on Voter Turnout in the August 2006 Primary Election

	Experimental Group				
	Control	Civic Duty	Hawthorne	Self	Neighbors
Percentage Voting	29.7%	31.5%	32.2%	34.5%	37.8%
N of Individuals	191,243	38,218	38,204	38,218	38,201

- Matt Salganik in his book *Bit by Bit* argues that it is an ethical imperative to only use as many subjects as we need in an experiment
- Why did GGL use sample sizes of 38,000? Could they have used fewer? How would we know?
- We choose the sample size that can ensure that we can detect what we think is the true treatment effect (i.e. reject the null of no effect).
- Small effects will require more observations than large effects. But how many more?

Statistical Power

Type 1 errors–false discovery

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

- Experimental design: what effect can I detect?

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

- Experimental design: what effect can I detect?
- What is the power of my study?

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

- Experimental design: what effect can I detect?
- What is the power of my study?
- Power: 1 - Prob. type 2 error.

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

- Experimental design: what effect can I detect?
- What is the power of my study?
- Power: 1 - Prob. type 2 error.

Power will (in general) depend on four factors (particularly for t/normally distributed test statistics):

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

- Experimental design: what effect can I detect?
- What is the power of my study?
- Power: 1 - Prob. type 2 error.

Power will (in general) depend on four factors (particularly for t/normally distributed test statistics):

1) Type I error rate (α)

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

- Experimental design: what effect can I detect?
- What is the power of my study?
- Power: 1 - Prob. type 2 error.

Power will (in general) depend on four factors (particularly for t/normally distributed test statistics):

1) Type I error rate (α)
2) Effect size

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

- Experimental design: what effect can I detect?
- What is the power of my study?
- Power: 1 - Prob. type 2 error.

Power will (in general) depend on four factors (particularly for t/normally distributed test statistics):

1) Type I error rate (α)
2) Effect size
3) Variance

Statistical Power

Type 1 errors–false discovery
α identifies tolerance for type 1 errors.
Type 2 errors–failed discovery.

- Experimental design: what effect can I detect?
- What is the power of my study?
- Power: 1 - Prob. type 2 error.

Power will (in general) depend on four factors (particularly for t/normally distributed test statistics):

1) Type I error rate (α)
2) Effect size
3) Variance
4) Number of observations

Power and Hypothesis Tests

Suppose (again) we're running an experiment, sampling from two normal distributions (treatment and control).

Power and Hypothesis Tests

Suppose (again) we're running an experiment, sampling from two normal distributions (treatment and control).
Testing hypotheses:

Power and Hypothesis Tests

Suppose (again) we're running an experiment, sampling from two normal distributions (treatment and control).
Testing hypotheses:

H_0: $\mu_t - \mu_c \equiv \mu_{diff} = 0$

Power and Hypothesis Tests

Suppose (again) we're running an experiment, sampling from two normal distributions (treatment and control).
Testing hypotheses:

H_0: $\mu_t - \mu_c \equiv \mu_{diff} = 0$

H_1: $\mu_{diff} \neq 0$

Power and Hypothesis Tests

Suppose (again) we're running an experiment, sampling from two normal distributions (treatment and control).

Testing hypotheses:

H_0: $\mu_t - \mu_c \equiv \mu_{diff} = 0$

H_1: $\mu_{diff} \neq 0$

Test statistic:

Power and Hypothesis Tests

Suppose (again) we're running an experiment, sampling from two normal distributions (treatment and control).
Testing hypotheses:

H_0: $\mu_t - \mu_c \equiv \mu_{diff} = 0$

H_1: $\mu_{diff} \neq 0$

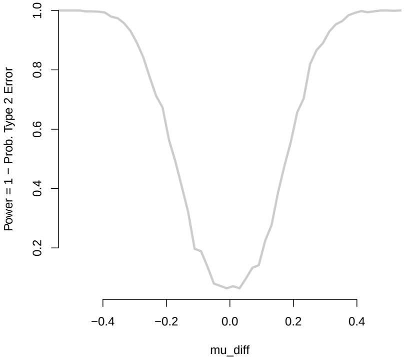
Test statistic:

$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}_t^2}{n_t} + \frac{\widehat{\sigma_c^2}}{n_c}}}$$

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$

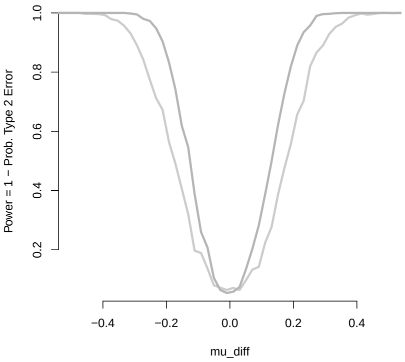


$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}_t^2}{n_t} + \frac{\widehat{\sigma_c^2}}{n_c}}}$$

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



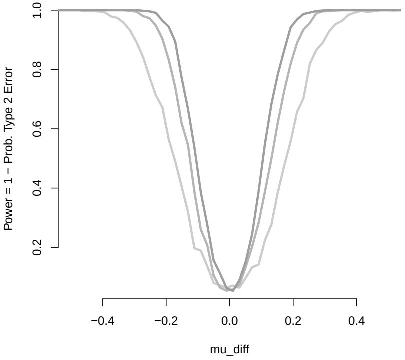
$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}_t^2}{n_t} + \frac{\widehat{\sigma_c^2}}{n_c}}}$$

Increasing n_t and n_c $\rightarrow$ facilitates detection of smaller effects

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



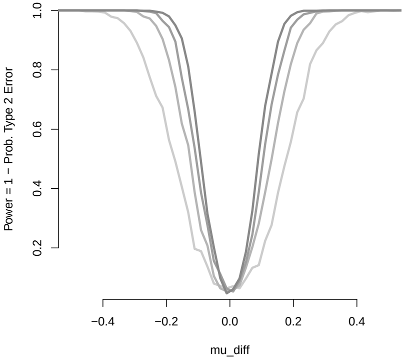
$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}_t^2}{n_t} + \frac{\widehat{\sigma_c^2}}{n_c}}}$$

Increasing n_t and n_c $\rightarrow$ facilitates detection of smaller effects

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



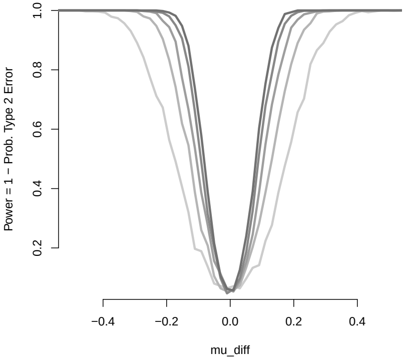
$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}_t^2}{n_t} + \frac{\widehat{\sigma_c^2}}{n_c}}}$$

Increasing n_t and n_c $\rightarrow$ facilitates detection of smaller effects

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



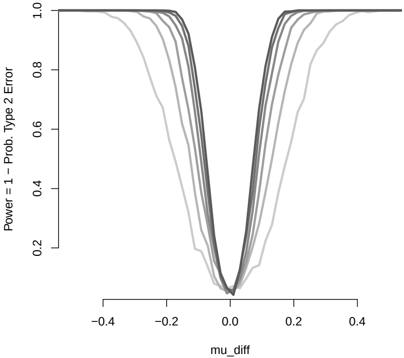
$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}^2_t}{n_t} + \frac{\widehat{\sigma^2_c}}{n_c}}}$$

Increasing n_t and n_c $\rightarrow$ facilitates detection of smaller effects

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



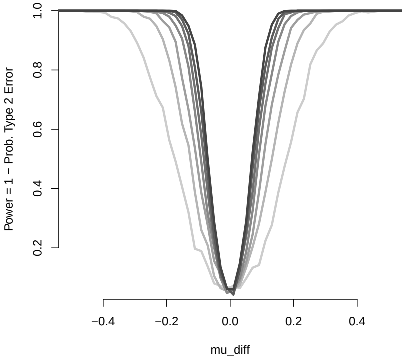
$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}_t^2}{n_t} + \frac{\widehat{\sigma_c^2}}{n_c}}}$$

Increasing n_t and n_c $\rightarrow$ facilitates detection of smaller effects

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



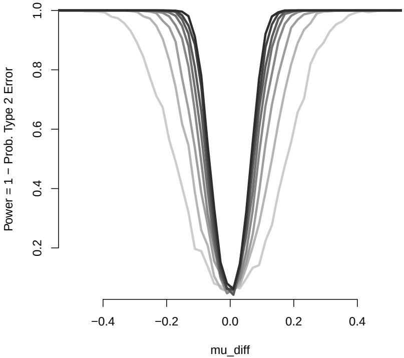
$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}^2_t}{n_t} + \frac{\widehat{\sigma^2_c}}{n_c}}}$$

Increasing n_t and n_c $\rightarrow$ facilitates detection of smaller effects

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



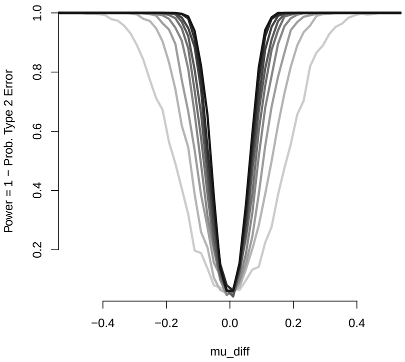
$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}_t^2}{n_t} + \frac{\widehat{\sigma_c^2}}{n_c}}}$$

Increasing n_t and n_c $\rightarrow$ facilitates detection of smaller effects

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



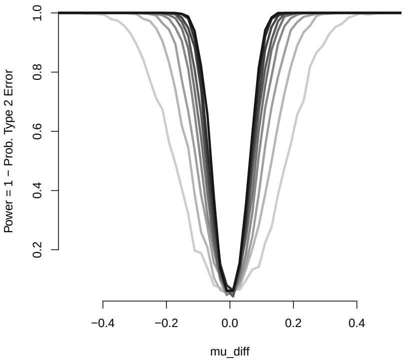
$$t = \frac{\overline{T} - \overline{C}}{\sqrt{\frac{\widehat{\sigma}_t^2}{n_t} + \frac{\widehat{\sigma_c^2}}{n_c}}}$$

Increasing n_t and n_c $\rightarrow$ facilitates detection of smaller effects

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$

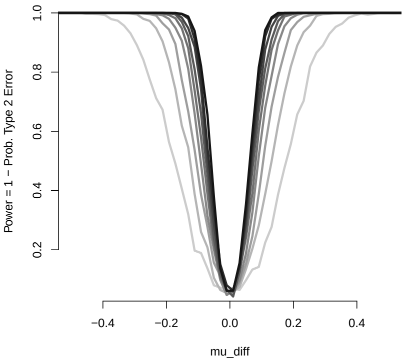


- Power analysis:

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned}\Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96)\\ \text{Power} &= 1 - \Pr(\text{Type 2 error})\end{aligned}$$

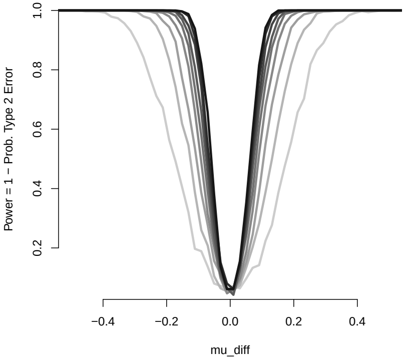


- Power analysis:
- a priori How many participants in study

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$

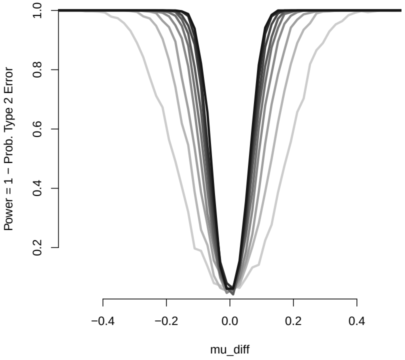


- Power analysis:
- a priori How many participants in study
- post-hoc Largest effect size likely detected, given study characteristics

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$

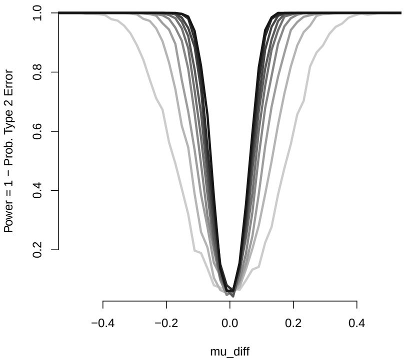


- Power analysis:
- a priori How many participants in study
- post-hoc Largest effect size likely detected, given study characteristics
- Precise zeros: under powered (small n_t and n_c relative to variance) make failed rejection likely.

Power and Hypothesis Tests

Key question: given true value of $\mu^*_{\text{diff}} \neq 0$ what is the probability t falls in "fail to reject" region?

$$\begin{aligned} \Pr(\text{Type 2 error}) &= P(-1.96 < t < 1.96) \\ \text{Power} &= 1 - \Pr(\text{Type 2 error}) \end{aligned}$$



- Power analysis:
- a priori How many participants in study
- post-hoc Largest effect size likely detected, given study characteristics
- Precise zeros: under powered (small n_t and n_c relative to variance) make failed rejection likely.
- not evidence of no effect!

Back to Gerber, Green and Larimer

- Before starting they did a power analysis

Back to Gerber, Green and Larimer

- Before starting they did a power analysis
- Steps to a power analysis

Back to Gerber, Green and Larimer

- Before starting they did a power analysis
- Steps to a power analysis
  1. Pick some hypothetical effect size $\mu_y - \mu_x = .05$
  2. Calculate the distribution of test statistic T under that effect size
  3. Calculate the probability of rejecting the null under that distribution
  4. Repeat for different effect sizes

Back to Gerber, Green and Larimer

- Before starting they did a power analysis
- Steps to a power analysis
  1. Pick some hypothetical effect size $\mu_y - \mu_x = .05$
  2. Calculate the distribution of test statistic T under that effect size
  3. Calculate the probability of rejecting the null under that distribution
  4. Repeat for different effect sizes
- Let's say we want to run another experiment where we believe the true effect is $\mu_y - \mu_x = 0.05$ and the variances are $\sigma_y^2 = \sigma_x^2 = .2$

Back to Gerber, Green and Larimer

- Before starting they did a power analysis
- Steps to a power analysis
  1. Pick some hypothetical effect size $\mu_y - \mu_x = .05$
  2. Calculate the distribution of test statistic T under that effect size
  3. Calculate the probability of rejecting the null under that distribution
  4. Repeat for different effect sizes
- Let's say we want to run another experiment where we believe the true effect is $\mu_y - \mu_x = 0.05$ and the variances are $\sigma_y^2 = \sigma_x^2 = .2$
- We can only afford to send out 500 mailers. Should we run the experiment?

Back to Gerber, Green and Larimer

- Before starting they did a power analysis
- Steps to a power analysis
  1. Pick some hypothetical effect size $\mu_y - \mu_x = .05$
  2. Calculate the distribution of test statistic T under that effect size
  3. Calculate the probability of rejecting the null under that distribution
  4. Repeat for different effect sizes
- Let's say we want to run another experiment where we believe the true effect is $\mu_y - \mu_x = 0.05$ and the variances are $\sigma_y^2 = \sigma_x^2 = .2$
- We can only afford to send out 500 mailers. Should we run the experiment?
- To the board!

A Short Case Study of Retrospective Power Analysis

A Short Case Study of Retrospective Power Analysis

- Durante, Arsena and Griskevicius (2013) publish a study in *Psychological Science* on menstrual cycles and political attitudes

A Short Case Study of Retrospective Power Analysis

- Durante, Arsena and Griskevicius (2013) publish a study in *Psychological Science* on menstrual cycles and political attitudes
- They report a 17 point swing in voting preferences in the 2012 election.

A Short Case Study of Retrospective Power Analysis

- Durante, Arsena and Griskevicius (2013) publish a study in *Psychological Science* on menstrual cycles and political attitudes
- They report a 17 point swing in voting preferences in the 2012 election.
  - for context polling showed Obama's support varying by 7 points during the entire general campaign (Gallup 2012)

A Short Case Study of Retrospective Power Analysis

- Durante, Arsena and Griskevicius (2013) publish a study in *Psychological Science* on menstrual cycles and political attitudes
- They report a 17 point swing in voting preferences in the 2012 election.
  - for context polling showed Obama's support varying by 7 points during the entire general campaign (Gallup 2012)
  - the implied standard error was 8.1 percentage points with a p-value of 0.035.

A Short Case Study of Retrospective Power Analysis

- Durante, Arsena and Griskevicius (2013) publish a study in *Psychological Science* on menstrual cycles and political attitudes
- They report a 17 point swing in voting preferences in the 2012 election.
  - for context polling showed Obama's support varying by 7 points during the entire general campaign (Gallup 2012)
  - the implied standard error was 8.1 percentage points with a p-value of 0.035.
- Gelman and Carlin (2014) show how the study design should cause us to question the finding.

A Short Case Study of Retrospective Power Analysis

- Durante, Arsena and Griskevicius (2013) publish a study in *Psychological Science* on menstrual cycles and political attitudes
- They report a 17 point swing in voting preferences in the 2012 election.
  - for context polling showed Obama's support varying by 7 points during the entire general campaign (Gallup 2012)
  - the implied standard error was 8.1 percentage points with a p-value of 0.035.
- Gelman and Carlin (2014) show how the study design should cause us to question the finding.
  - they assume a true effect size of 2 points (upper end of plausible)

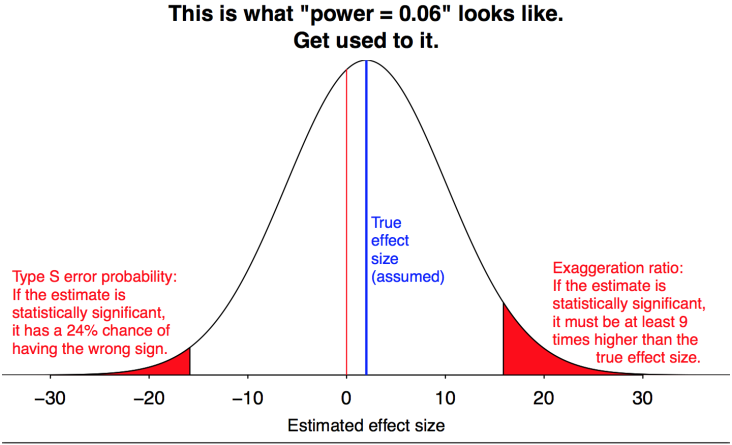
A Short Case Study of Retrospective Power Analysis

- Durante, Arsena and Griskevicius (2013) publish a study in *Psychological Science* on menstrual cycles and political attitudes
- They report a 17 point swing in voting preferences in the 2012 election.
  - for context polling showed Obama's support varying by 7 points during the entire general campaign (Gallup 2012)
  - the implied standard error was 8.1 percentage points with a *p*-value of 0.035.
- Gelman and Carlin (2014) show how the study design should cause us to question the finding.
  - they assume a true effect size of 2 points (upper end of plausible)
  - perform a power analysis under given effect size, observed standard error of measurement.

A Short Case Study of Retrospective Power Analysis

- Durante, Arsena and Griskevicius (2013) publish a study in *Psychological Science* on menstrual cycles and political attitudes
- They report a 17 point swing in voting preferences in the 2012 election.
  - for context polling showed Obama's support varying by 7 points during the entire general campaign (Gallup 2012)
  - the implied standard error was 8.1 percentage points with a p-value of 0.035.
- Gelman and Carlin (2014) show how the study design should cause us to question the finding.
  - they assume a true effect size of 2 points (upper end of plausible)
  - perform a power analysis under given effect size, observed standard error of measurement.
  - power comes out to 0.06

A Troubling Figure (via Gelman)



Problems with p-values

Problems with p-values

- p-values are extremely common in the social sciences and are often the standard by which the value of the finding is judged.

Problems with p-values

- p-values are extremely common in the social sciences and are often the standard by which the value of the finding is judged.
- p-values are not:

Problems with p-values

- p-values are extremely common in the social sciences and are often the standard by which the value of the finding is judged.
- p-values are not:
  - an indication of a large substantive effect
  - the probability that the null hypothesis is true
  - the probability that the alternative hypothesis is false

Problems with p-values

- p-values are extremely common in the social sciences and are often the standard by which the value of the finding is judged.
- p-values are not:
  - an indication of a large substantive effect
  - the probability that the null hypothesis is true
  - the probability that the alternative hypothesis is false
- a large p-value could mean either that we are in the null world OR that we had insufficient power

So what is the basic idea?

The idea was to run an experiment, then see if the results were consistent with what random chance might produce. Researchers would first set up a 'null hypothesis' that they wanted to disprove, such as there being no correlation or no difference between groups. Next, they would play the devil's advocate and, assuming that this null hypothesis was in fact true, calculate the chances of getting results at least as extreme as what was actually observed. This probability was the P value. The smaller it was, suggested Fisher, the greater the likelihood that the straw-man null hypothesis was false.
(Nunzo 2014, emphasis mine)

I've got 99 problems. . .

I've got 99 problems. . .

p-values are hard to interpret, but even in the best scenarios they have some key problems:

I've got 99 problems. . .

p-values are hard to interpret, but even in the best scenarios they have some key problems:

- they remove focus from data, measurement, theory and the substantive quantity of interest

I've got 99 problems. . .

p-values are hard to interpret, but even in the best scenarios they have some key problems:

- they remove focus from data, measurement, theory and the substantive quantity of interest
- they are often applied outside the dichotomous/decision-making framework where they make some sense

I've got 99 problems...

p-values are hard to interpret, but even in the best scenarios they have some key problems:

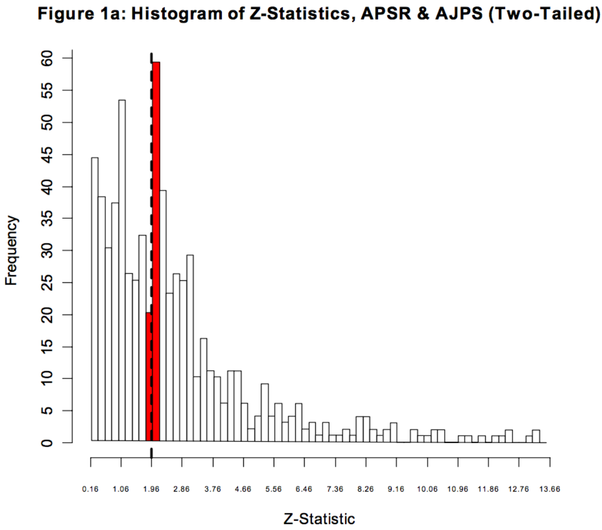
- they remove focus from data, measurement, theory and the substantive quantity of interest
- they are often applied outside the dichotomous/decision-making framework where they make some sense
- "significant covariates" aren't even necessarily good predictors (Ward et al 2010, Lo et al 2015)

I've got 99 problems. . .

p-values are hard to interpret, but even in the best scenarios they have some key problems:

- they remove focus from data, measurement, theory and the substantive quantity of interest
- they are often applied outside the dichotomous/decision-making framework where they make some sense
- "significant covariates" aren't even necessarily good predictors (Ward et al 2010, Lo et al 2015)
- they lead to publication filtering on arbitrary cutoffs.

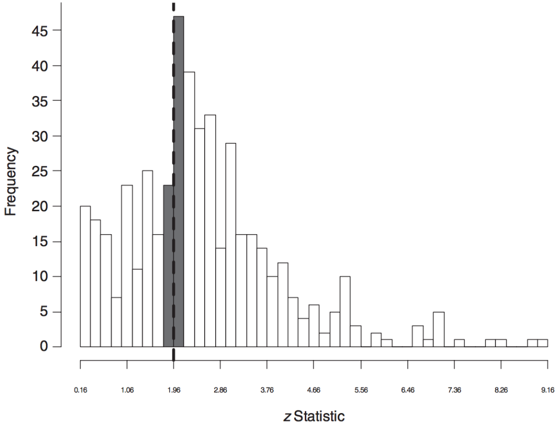
Arbitrary Cutoffs



Gerber and Malhotra (2006) Top Political Science Journals

Arbitrary Cutoffs

Figure 1
Histogram of *z* Statistics From the *American Sociological Review*, the *American Journal of Sociology*, and *The Sociological Quarterly* (Two-Tailed)



Gerber and Malhotra (2008) Top Sociology Journals

Arbitrary Cutoffs

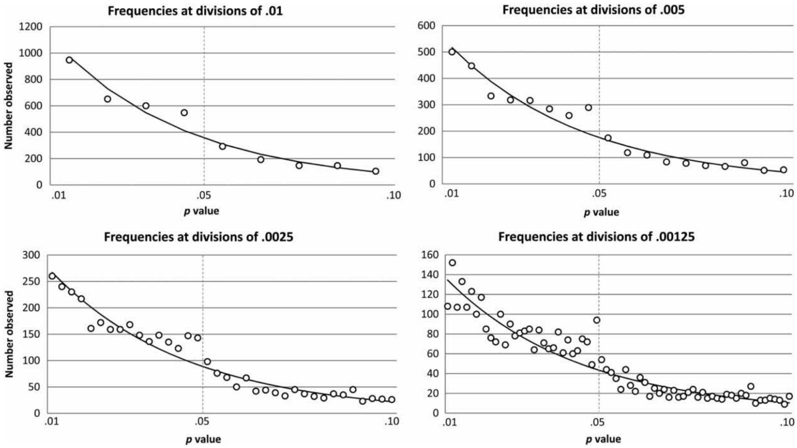


Figure 1.. *The graphs show the distribution of 3,627* p *values from three major psychology journals.*

Masicampo and Lalande (2012) Top Psychology Journals

Still Not Convinced?
The Real Harm of Misinterpreted p-values

ELSEVIER

Accident Analysis and Prevention 36 (2004) 495–500

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Viewpoint

The harm done by tests of significance

Ezra Hauer*

35 Merton Street, Apt. 1706, Toronto, Ont., Canada M4S 3G4

Abstract

Three historical episodes in which the application of null hypothesis significance testing (NHST) led to the mis-interpretation of data are described. It is argued that the pervasive use of this statistical ritual impedes the accumulation of knowledge and is unfit for use.

Keywords: Significance; Statistical hypothesis; Scientific method

Example from Hauer: Right-Turn-On-Red

Table 1
The Virginia RTOR study

	Before RTOR signing	After RTOR signing
Fatal crashes	0	0
Personal injury crashes	43	60
Persons injured	69	72
Property damage crashes	265	277
Property damage (US$)	161243	170807
Total crashes	308	337

The Point in Hauer

The Point in Hauer

- Two other interesting examples in Hauer (2004)

The Point in Hauer

- Two other interesting examples in Hauer (2004)
- Core issue is that lack of significance is not an indication of a zero effect, it could also be a lack of power (i.e. a small sample size relative to the difficulty of detecting the effect)

The Point in Hauer

- Two other interesting examples in Hauer (2004)
- Core issue is that lack of significance is not an indication of a zero effect, it could also be a lack of power (i.e. a small sample size relative to the difficulty of detecting the effect)
- On the opposite end, large tech companies essentially never use significance testing because they have huge samples which essentially always find some non-zero effect. But that doesn't make the effect significant in a colloquial sense of important.

P-values and Confidence Intervals

P-values one of most used tests in the social sciences–and you're telling me not to rely on them?

P-values and Confidence Intervals

P-values one of most used tests in the social sciences–and you're telling me not to rely on them?

- Basically, yes.

P-values and Confidence Intervals

What's the matter with you?

P-values and Confidence Intervals

What's the matter with you?

- Two reasons not to worship p-values [of many]

P-values and Confidence Intervals

What's the matter with you?

- Two reasons not to worship p-values [of many]
    1) Statistical: they represent a very specific quantity under a null distribution. If you don't really care about rejecting just that null, then you should focus on providing more information

P-values and Confidence Intervals

What's the matter with you?

- Two reasons not to worship p-values [of many]
  1) Statistical: they represent a very specific quantity under a null distribution. If you don't really care about rejecting just that null, then you should focus on providing more information
  2) Substantive: p-values are divorced from your quantity of interest–which almost always should relate to how much an intervention changes a quantity of social scientific interest (Grandparent rule)

P-values and Confidence Intervals

But I want to assess the probability that my hypothesis is true–why can't I use a p-value?

P-values and Confidence Intervals

But I want to assess the probability that my hypothesis is true–why can't I use a p-value?

1) Me too, good luck.

P-values and Confidence Intervals

But I want to assess the probability that my hypothesis is true–why can't I use a p-value?

1) Me too, good luck.
2) That's not what p-values measure

P-values and Confidence Intervals

But I want to assess the probability that my hypothesis is true–why can't I use a p-value?

1) Me too, good luck.
2) That's not what p-values measure
3) No one study is going to eliminate an entire hypothesis; even if that study generates a really small p-value, you'd probably want an entirely different infrastructure

P-values and Confidence Intervals

I like to estimate the certainty of my results using a p-value, do you find this problematic?

P-values and Confidence Intervals

I like to estimate the certainty of my results using a p-value, do you find this problematic?

- Yes

P-values and Confidence Intervals

I like to estimate the certainty of my results using a p-value, do you find this problematic?

- Yes
- What could be wrong with this?

P-values and Confidence Intervals

I like to estimate the certainty of my results using a p-value, do you find this problematic?

- Yes
- What could be wrong with this?

P-values and Confidence Intervals

OK–I get it.

P-values and Confidence Intervals

OK–I get it. No p-values

P-values and Confidence Intervals

OK–I get it. No p-values (but I'm still going to use them anyways.

P-values and Confidence Intervals

OK–I get it. No p-values (but I'm still going to use them anyways. I'm going to put an extra star in my table and name it after you).

P-values and Confidence Intervals

OK–I get it. No p-values (but I'm still going to use them anyways. I'm going to put an extra star in my table and name it after you).

But what should I do? (Other than talk about methodologists and how they are jerks?)

P-values and Confidence Intervals

OK–I get it. No p-values (but I'm still going to use them anyways. I'm going to put an extra star in my table and name it after you).

But what should I do? (Other than talk about methodologists and how they are jerks?)
Confidence Intervals, Graphical Presentation of a quantity of interest

P-values and Confidence Intervals

OK–I get it. No p-values (but I'm still going to use them anyways. I'm going to put an extra star in my table and name it after you).

But what should I do? (Other than talk about methodologists and how they are jerks?)
Confidence Intervals, Graphical Presentation of a quantity of interest
Why?

P-values and Confidence Intervals

OK–I get it. No p-values (but I'm still going to use them anyways. I'm going to put an extra star in my table and name it after you).

But what should I do? (Other than talk about methodologists and how they are jerks?)
Confidence Intervals, Graphical Presentation of a quantity of interest
Why?

1) Substantive significance and statistical significance simultaneously

P-values and Confidence Intervals

OK–I get it. No p-values (but I'm still going to use them anyways. I'm going to put an extra star in my table and name it after you).

But what should I do? (Other than talk about methodologists and how they are jerks?)
Confidence Intervals, Graphical Presentation of a quantity of interest
Why?

1) Substantive significance and statistical significance simultaneously
2) Make comparisons across factors approximately and accurately (though exercise caution)

P-values and Confidence Intervals

OK–I get it. No p-values (but I'm still going to use them anyways. I'm going to put an extra star in my table and name it after you).

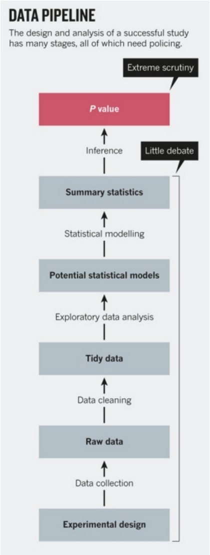
But what should I do? (Other than talk about methodologists and how they are jerks?)

Confidence Intervals, Graphical Presentation of a quantity of interest

Why?

1) Substantive significance and statistical significance simultaneously
2) Make comparisons across factors approximately and accurately (though exercise caution)
3) Harder to hide weird looking effects

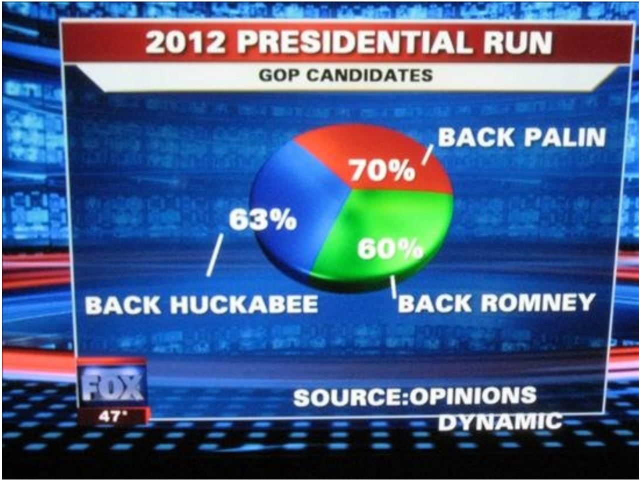
But Let's Not Obsess Too Much About p-values



From Leek and Peng (2015) "*P* values are just the tip of the iceberg" *Nature*.

An Intro Motivation

An Intro Motivation



Visualization

Visualization

- Visualization is hard but ultimately extremely important

Visualization

- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.

Visualization

- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes

Visualization

- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes
  - drawing people into a topic/dataset

Visualization

- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes
  - drawing people into a topic/dataset
  - presenting evidence

Visualization

- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes
  - drawing people into a topic/dataset
  - presenting evidence
  - exploration/model checking

Visualization

- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes
  - drawing people into a topic/dataset
  - presenting evidence
  - exploration/model checking
- Three steps involved

Visualization

- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes
  - drawing people into a topic/dataset
  - presenting evidence
  - exploration/model checking
- Three steps involved
  1. clearly define the goal

Visualization

- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes
  - drawing people into a topic/dataset
  - presenting evidence
  - exploration/model checking
- Three steps involved
  1. clearly define the goal
  2. estimate quantities of interest

Visualization

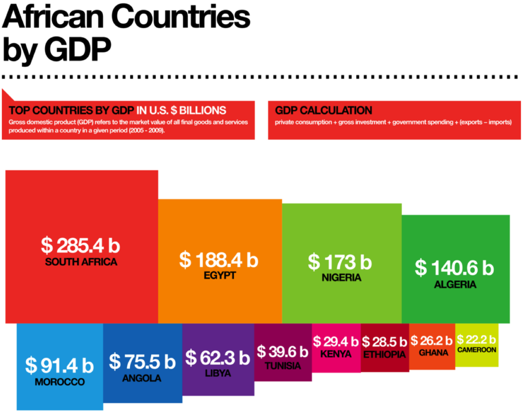
- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes
  - drawing people into a topic/dataset
  - presenting evidence
  - exploration/model checking
- Three steps involved
  1. clearly define the goal
  2. estimate quantities of interest
  3. present those quantities in a compelling way

Visualization

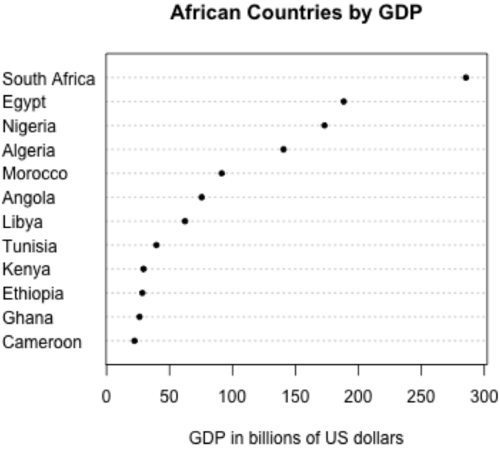
- Visualization is hard but ultimately extremely important
- It is absurd that we spend months collecting data, weeks analyzing it and five minutes slapping it into an unreadable table.
- Visualization can be used for many purposes
  - drawing people into a topic/dataset
  - presenting evidence
  - exploration/model checking
- Three steps involved
  1. clearly define the goal
  2. estimate quantities of interest
  3. present those quantities in a compelling way
- Good design involves thinking carefully about the audience

Examples via Gelman

Examples via Gelman

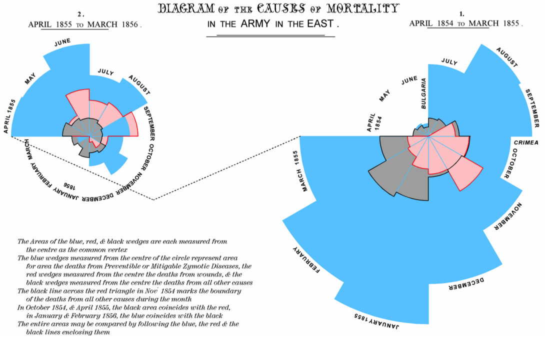


Examples via Gelman



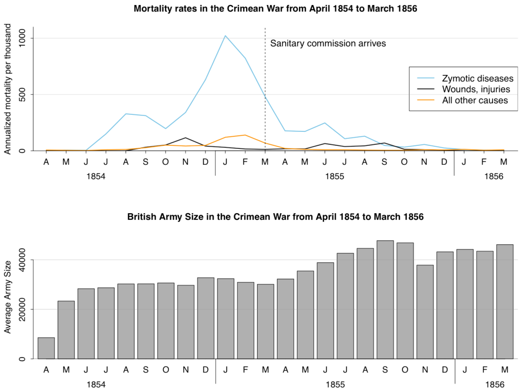
One may be better for drawing people in, the other for evidence.

Examples via Gelman



A classic infographic by Nightingale. Dramatizes the problem.

Examples via Gelman



A simpler version where it is easier to see the patterns.

How Not to Present Statistical Results

How Not to Present Statistical Results

How Not to Present Statistical Results

TABLE 1
Predicting Which Ethnic Group Conquered Most of Bosnia

Attention to Bosnia crisis	.609**
Age	.007**
Education	.289**
Family income	.151**
Race (non-White/White)	.695**
Gender (female/male)	.789**
Region (South/non-South)	.076
Network coverage	.000
Education × Time	–.003*
Time in months	.078**
Constant	–9.257**
Number	7,021
–2 log-likelihood	7,215.231
Goodness of fit	6,789.45
Cox & Snell R^2	.212
Nagelkerke R^2	.295
Overall correct classification (%)	73.96

SOURCE: *Times Mirror* polls from September 1992, January 1993, September 1993, January 1994, and June 1995.
NOTE: Unstandardized coefficients for logistic regression. Dependent variable is knowledge of which group conquered most of Bosnia.
$*p \leq .05$, two-tailed. $**p \leq .01$, two-tailed.

How Not to Present Statistical Results

TABLE 1
Predicting Which Ethnic Group Conquered Most of Bosnia

Attention to Bosnia crisis	.609**
Age	.007**
Education	.289**
Family income	.151**
Race (non-White/White)	.695**
Gender (female/male)	.789**
Region (South/non-South)	.076
Network coverage	.000
Education × Time	–.003*
Time in months	.078**
Constant	–9.257**
Number	7,021
–2 log-likelihood	7,215.231
Goodness of fit	6,789.45
Cox & Snell R^2	.212
Nagelkerke R^2	.295
Overall correct classification (%)	73.96

SOURCE: *Times Mirror* polls from September 1992, January 1993, September 1993, January 1994, and June 1995.
NOTE: Unstandardized coefficients for logistic regression. Dependent variable is knowledge of which group conquered most of Bosnia.
*$p \leq .05$, two-tailed. **$p \leq .01$, two-tailed.

- This one is typical of current practice, not especially bad.

How Not to Present Statistical Results

TABLE 1

Predicting Which Ethnic Group Conquered Most of Bosnia

Attention to Bosnia crisis	.609**
Age	.007**
Education	.289**
Family income	.151**
Race (non-White/White)	.695**
Gender (female/male)	.789**
Region (South/non-South)	.076
Network coverage	.000
Education × Time	–.003*
Time in months	.078**
Constant	–9.257**
Number	7,021
–2 log-likelihood	7,215.231
Goodness of fit	6,789.45
Cox & Snell R^2	.212
Nagelkerke R^2	.295
Overall correct classification (%)	73.96

SOURCE: *Times Mirror* polls from September 1992, January 1993, September 1993, January 1994, and June 1995.

NOTE: Unstandardized coefficients for logistic regression. Dependent variable is knowledge of which group conquered most of Bosnia.

$*p \leq .05$, two-tailed. $**p \leq .01$, two-tailed.

- This one is typical of current practice, not especially bad.
- What do these numbers mean?

How Not to Present Statistical Results

TABLE 1

Predicting Which Ethnic Group Conquered Most of Bosnia

Attention to Bosnia crisis	.609**
Age	.007**
Education	.289**
Family income	.151**
Race (non-White/White)	.695**
Gender (female/male)	.789**
Region (South/non-South)	.076
Network coverage	.000
Education × Time	–.003*
Time in months	.078**
Constant	–9.257**
Number	7,021
–2 log-likelihood	7,215.231
Goodness of fit	6,789.45
Cox & Snell R^2	.212
Nagelkerke R^2	.295
Overall correct classification (%)	73.96

SOURCE: *Times Mirror* polls from September 1992, January 1993, September 1993, January 1994, and June 1995.
NOTE: Unstandardized coefficients for logistic regression. Dependent variable is knowledge of which group conquered most of Bosnia.
*$p \leq .05$, two-tailed. **$p \leq .01$, two-tailed.

- This one is typical of current practice, not especially bad.
- What do these numbers mean?
- Why so much whitespace? Can you connect cols A and B?

How Not to Present Statistical Results: Continued

TABLE 1
Predicting Which Ethnic Group Conquered Most of Bosnia

Attention to Bosnia crisis	.609**
Age	.007**
Education	.289**
Family income	.151**
Race (non-White/White)	.695**
Gender (female/male)	.789**
Region (South/non-South)	.076
Network coverage	.000
Education × Time	–.003*
Time in months	.078**
Constant	–9.257**
Number	7,021
–2 log-likelihood	7,215.231
Goodness of fit	6,789.45
Cox & Snell R^2	.212
Nagelkerke R^2	.295
Overall correct classification (%)	73.96

SOURCE: *Times Mirror* polls from September 1992, January 1993, September 1993, January 1994, and June 1995.
NOTE: Unstandardized coefficients for logistic regression. Dependent variable is knowledge of which group conquered most of Bosnia.
*$p \leq .05$, two-tailed. **$p \leq .01$, two-tailed.

How Not to Present Statistical Results: Continued

TABLE 1

Predicting Which Ethnic Group Conquered Most of Bosnia

Attention to Bosnia crisis	.609**
Age	.007**
Education	.289**
Family income	.151**
Race (non-White/White)	.695**
Gender (female/male)	.789**
Region (South/non-South)	.076
Network coverage	.000
Education × Time	–.003*
Time in months	.078**
Constant	–9.257**
Number	7,021
–2 log-likelihood	7,215.231
Goodness of fit	6,789.45
Cox & Snell R^2	.212
Nagelkerke R^2	.295
Overall correct classification (%)	73.96

SOURCE: *Times Mirror* polls from September 1992, January 1993, September 1993, January 1994, and June 1995.

NOTE: Unstandardized coefficients for logistic regression. Dependent variable is knowledge of which group conquered most of Bosnia.

*$p \leq .05$, two-tailed. **$p \leq .01$, two-tailed.

- What does the star-gazing add?

How Not to Present Statistical Results: Continued

TABLE 1

Predicting Which Ethnic Group Conquered Most of Bosnia

Attention to Bosnia crisis	.609**
Age	.007**
Education	.289**
Family income	.151**
Race (non-White/White)	.695**
Gender (female/male)	.789**
Region (South/non-South)	.076
Network coverage	.000
Education × Time	–.003*
Time in months	.078**
Constant	–9.257**
Number	7,021
–2 log-likelihood	7,215.231
Goodness of fit	6,789.45
Cox & Snell R^2	.212
Nagelkerke R^2	.295
Overall correct classification (%)	73.96

SOURCE: *Times Mirror* polls from September 1992, January 1993, September 1993, January 1994, and June 1995.

NOTE: Unstandardized coefficients for logistic regression. Dependent variable is knowledge of which group conquered most of Bosnia.

*$p \leq .05$, two-tailed. **$p \leq .01$, two-tailed.

- What does the star-gazing add?
- Can any be interpreted as causal estimates?

How Not to Present Statistical Results: Continued

TABLE 1
Predicting Which Ethnic Group Conquered Most of Bosnia

Attention to Bosnia crisis	.609**
Age	.007**
Education	.289**
Family income	.151**
Race (non-White/White)	.695**
Gender (female/male)	.789**
Region (South/non-South)	.076
Network coverage	.000
Education × Time	–.003*
Time in months	.078**
Constant	–9.257**
Number	7,021
–2 log-likelihood	7,215.231
Goodness of fit	6,789.45
Cox & Snell R^2	.212
Nagelkerke R^2	.295
Overall correct classification (%)	73.96

SOURCE: *Times Mirror* polls from September 1992, January 1993, September 1993, January 1994, and June 1995.
NOTE: Unstandardized coefficients for logistic regression. Dependent variable is knowledge of which group conquered most of Bosnia.
*$p \leq .05$, two-tailed. **$p \leq .01$, two-tailed.

- What does the star-gazing add?
- Can any be interpreted as causal estimates?
- Can you compute a quantity of interest from these numbers?

Interpretation and Presentation

Interpretation and Presentation

1. Statistical presentations should

Interpretation and Presentation

1. Statistical presentations should
   (a) Convey numerically precise estimates of the quantities of greatest substantive interest,

Interpretation and Presentation

1. Statistical presentations should
   (a) Convey numerically precise estimates of the quantities of greatest substantive interest,
   (b) Include reasonable measures of uncertainty about those estimates, and

Interpretation and Presentation

1. Statistical presentations should
   (a) Convey numerically precise estimates of the quantities of greatest substantive interest,
   (b) Include reasonable measures of uncertainty about those estimates, and
   (c) Require little specialized knowledge to understand.

Interpretation and Presentation

1. Statistical presentations should
   (a) Convey numerically precise estimates of the quantities of greatest substantive interest,
   (b) Include reasonable measures of uncertainty about those estimates, and
   (c) Require little specialized knowledge to understand.
   (d) Include no superfluous information, no long lists of coefficients no one understands, no star gazing, etc.

Interpretation and Presentation

1. Statistical presentations should
   (a) Convey numerically precise estimates of the quantities of greatest substantive interest,
   (b) Include reasonable measures of uncertainty about those estimates, and
   (c) Require little specialized knowledge to understand.
   (d) Include no superfluous information, no long lists of coefficients no one understands, no star gazing, etc.
2. For example: Other things being equal, an additional year of education would increase your annual income by $1,500 on average, plus or minus about $500.

Interpretation and Presentation

1. Statistical presentations should
   (a) Convey numerically precise estimates of the quantities of greatest substantive interest,
   (b) Include reasonable measures of uncertainty about those estimates, and
   (c) Require little specialized knowledge to understand.
   (d) Include no superfluous information, no long lists of coefficients no one understands, no star gazing, etc.
2. For example: Other things being equal, an additional year of education would increase your annual income by $1,500 on average, plus or minus about $500.
3. Your work should satisfy the expert reviewer and the lay person.

Interpretation and Presentation

1. Statistical presentations should
   (a) Convey numerically precise estimates of the quantities of greatest substantive interest,
   (b) Include reasonable measures of uncertainty about those estimates, and
   (c) Require little specialized knowledge to understand.
   (d) Include no superfluous information, no long lists of coefficients no one understands, no star gazing, etc.
2. For example: Other things being equal, an additional year of education would increase your annual income by $1,500 on average, plus or minus about $500.
3. Your work should satisfy the expert reviewer and the lay person.
4. King, Tomz, Wittenberg, "Making the Most of Statistical Analyses: Improving Interpretation and Presentation" *American Journal of Political Science*, Vol. 44, No. 2 (March, 2000): 341-355.

How to Calculate Quantities of Interest

How to Calculate Quantities of Interest

- We will discuss more general recipes in Soc504 but essentially three approaches: analytic, parametric simulation, non-parametric simulation.

How to Calculate Quantities of Interest

- We will discuss more general recipes in Soc504 but essentially three approaches: analytic, parametric simulation, non-parametric simulation.
- Analytic method involve finding the quantity and using the properties of variance and covariances to calculate a confidence interval. Can be challenging for complicated quantities.

How to Calculate Quantities of Interest

- We will discuss more general recipes in Soc504 but essentially three approaches: analytic, parametric simulation, non-parametric simulation.
- Analytic method involve finding the quantity and using the properties of variance and covariances to calculate a confidence interval. Can be challenging for complicated quantities.
- Parametric simulation uses the estimated coefficients and variance-covariance matrix to simulate outcomes (we will cover this in Soc504)

How to Calculate Quantities of Interest

- We will discuss more general recipes in Soc504 but essentially three approaches: analytic, parametric simulation, non-parametric simulation.
- Analytic method involve finding the quantity and using the properties of variance and covariances to calculate a confidence interval. Can be challenging for complicated quantities.
- Parametric simulation uses the estimated coefficients and variance-covariance matrix to simulate outcomes (we will cover this in Soc504)
- Non-parametric simulation uses the bootstrap. Calculate the quantity of interest within each bootstrap sample and then aggregate to get an estimate of the variance.

Visualization

Looking at data

Visualization

Looking at data

- Most basic method of inference

Visualization

Looking at data

- Most basic method of inference
- Hardest method of inference

Visualization

Looking at data

- Most basic method of inference
- Hardest method of inference art

Visualization

Looking at data

- Most basic method of inference
- Hardest method of inference art
- Why visualize?

Visualization

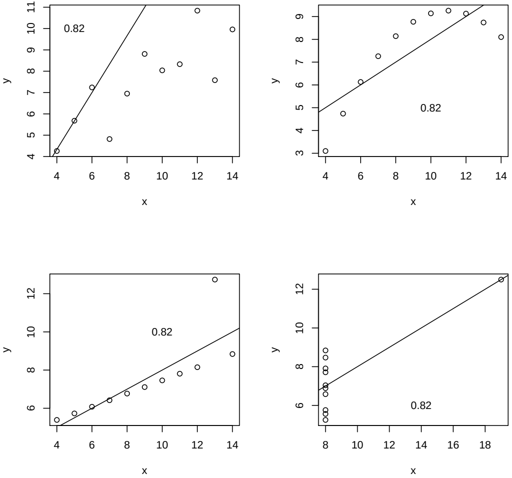
Looking at data

- Most basic method of inference
- Hardest method of inference art
- Why visualize?

Four (related) reasons

Reason 1: Models Lie

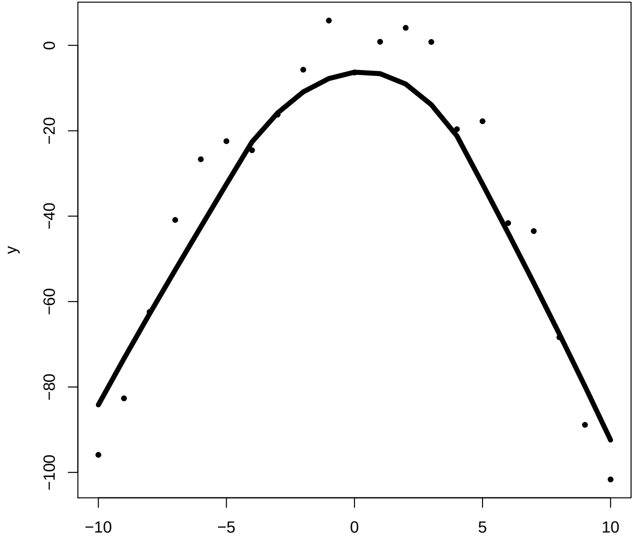
Remember anscombe's quartet?



Reason 2: Delivery of Information

	x	y
1	-10.00000	-95.89003
2	-9.00000	-82.65720
3	-8.00000	-62.42655
4	-7.00000	-40.87958
5	-6.00000	-26.67474
6	-5.00000	-22.43607
7	-4.00000	-24.55663
8	-3.00000	-16.21567
9	-2.00000	-5.69815
10	-1.00000	5.80266
11	0.00000	-6.36366
12	1.00000	0.83601
13	2.00000	4.10150
14	3.00000	0.79349
15	4.00000	-19.63152
16	5.00000	-17.76795
17	6.00000	-41.61587
18	7.00000	-43.49159
19	8.00000	-68.36981
20	9.00000	-88.86339
21	10.00000	-101.64692

Reason 2: Delivery of Information



Reason 3: Model Skepticism

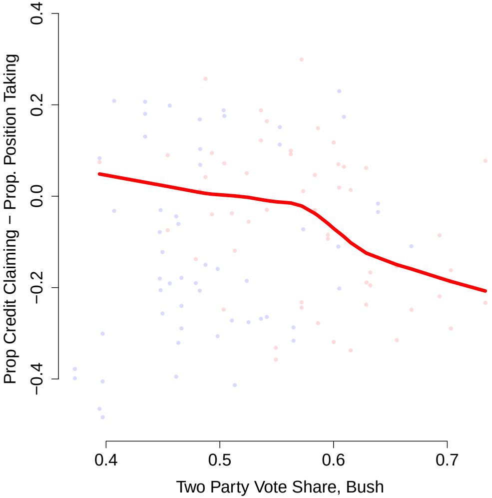
Reason 3: Model Skepticism and Checking

Reason 3: Model Skepticism

Reason 3: Model Skepticism and Checking

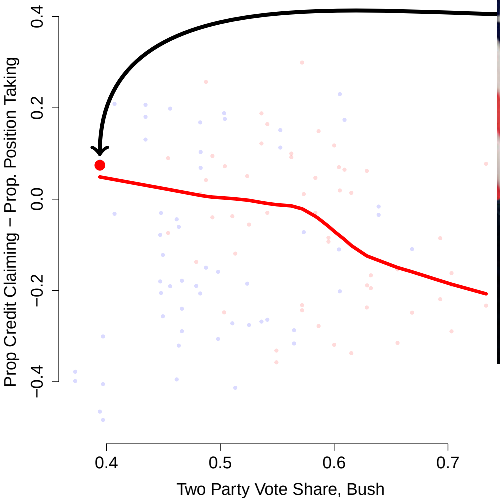
All inferences rest on assumption–visualization is a particularly reliable method for identifying obvious violations

Reason 4: Presentation and Persuasion



Example from Justin Grimmer's work.

Reason 4: Presentation and Persuasion



Lincoln Chafee, (R-RI)

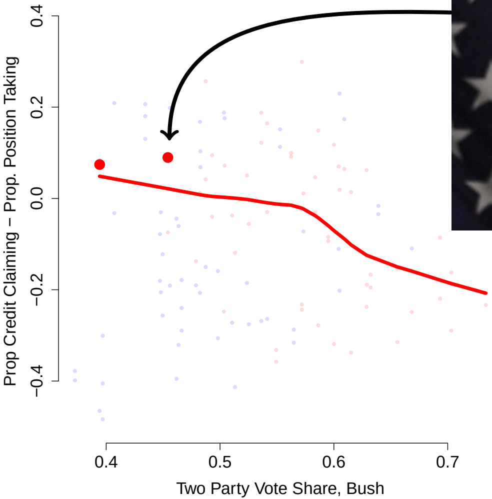
33% Credit Claiming

26% Position Taking

Airport Grants

Example from Justin Grimmer's work.

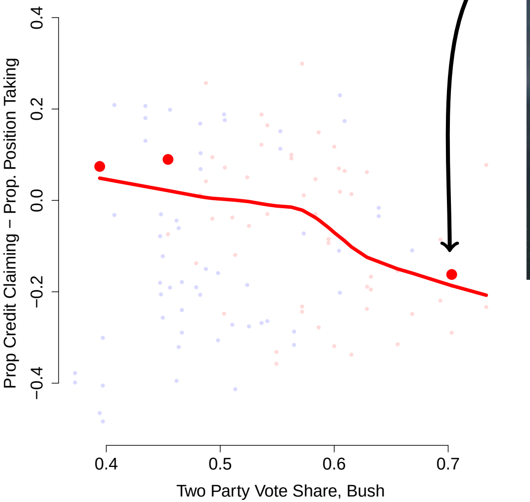
Reason 4: Presentation and Persuasion



Susan Collins, (R-RI)
29% Credit Claiming
20% Position Taking
Airport Grants

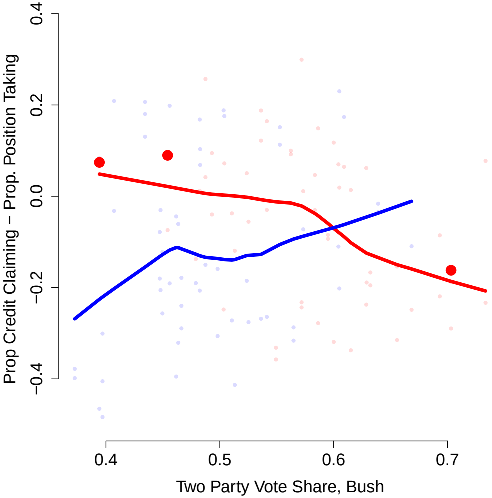
Example from Justin Grimmer's work.

Reason 4: Presentation and Persuasion



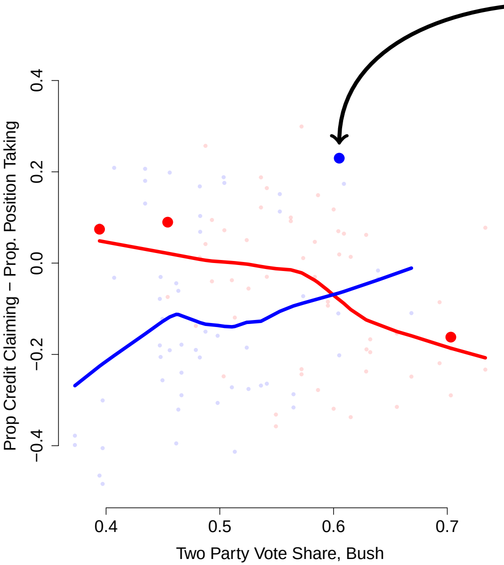
Mike Enzi, (R-WY)
18% Credit Claiming
35% Position Taking
Tax Reform

Reason 4: Presentation and Persuasion



Example from Justin Grimmer's work.

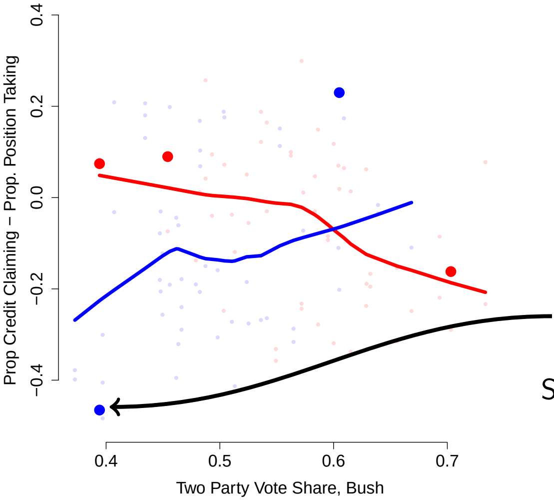
Reason 4: Presentation and Persuasion



John Tester, (D-MT)
43% Credit Claiming
20% Position Taking
Water Grants

Example from Justin Grimmer's work.

Reason 4: Presentation and Persuasion



Sheldon Whitehouse, (D-RI)
8% Credit Claiming
54% Position Taking
Iraq War

Example from Justin Grimmer's work.

Power, *p*-values and quantities of interest

Power, p-values and quantities of interest

- There are a number of problems with p-values and significance testing as currently applied

Power, p-values and quantities of interest

- There are a number of problems with p-values and significance testing as currently applied
- At the core of these problem is that p-values don't mean quite what we want them to mean

Power, p-values and quantities of interest

- There are a number of problems with p-values and significance testing as currently applied
- At the core of these problem is that p-values don't mean quite what we want them to mean
- A solution? Generate quantities of interest connected to your theory and present those.

Power, *p*-values and quantities of interest

- There are a number of problems with *p*-values and significance testing as currently applied
- At the core of these problem is that *p*-values don't mean quite what we want them to mean
- A solution? Generate quantities of interest connected to your theory and present those.
- Tools such as the bootstrap and simulation techniques described today can help get the right quantities. We will expand on this next semester.

Power, *p*-values and quantities of interest

- There are a number of problems with *p*-values and significance testing as currently applied
- At the core of these problem is that *p*-values don't mean quite what we want them to mean
- A solution? Generate quantities of interest connected to your theory and present those.
- Tools such as the bootstrap and simulation techniques described today can help get the right quantities. We will expand on this next semester.
- Visualizations can serve many purposes including a compelling way to present these quantities.

Power, *p*-values and quantities of interest

- There are a number of problems with *p*-values and significance testing as currently applied
- At the core of these problem is that *p*-values don't mean quite what we want them to mean
- A solution? Generate quantities of interest connected to your theory and present those.
- Tools such as the bootstrap and simulation techniques described today can help get the right quantities. We will expand on this next semester.
- Visualizations can serve many purposes including a compelling way to present these quantities.
- Many of these concerns have motivated a turn towards causal inference

What is Causal Inference?

What is Causal Inference?

- A causal inference is a statement about counterfactuals

What is Causal Inference?

- A causal inference is a statement about counterfactuals
- The difference between prediction and causal inference is the intervention on the system under study

What is Causal Inference?

- A causal inference is a statement about counterfactuals
- The difference between prediction and causal inference is the intervention on the system under study
- Like it or not, social science theories are almost always expressed as causal claims: e.g. "an increase in X causes an increase in Y"

What is Causal Inference?

- A causal inference is a statement about counterfactuals
- The difference between prediction and causal inference is the intervention on the system under study
- Like it or not, social science theories are almost always expressed as causal claims: e.g. "an increase in X causes an increase in Y"
- The study of causal inference helps us understand the assumptions we need to make this kind of claim.

Identification

Identification

- A quantity of interest is identified when access to infinite data would result in the estimate taking on only a single value

Identification

- A quantity of interest is identified when access to infinite data would result in the estimate taking on only a single value
- For example, having all dummy variables in a linear model is not statistically identified because they cannot be distinguished from the intercept.

Identification

- A quantity of interest is identified when access to infinite data would result in the estimate taking on only a single value
- For example, having all dummy variables in a linear model is not statistically identified because they cannot be distinguished from the intercept.
- Causal identification is what we can learn about a causal effect from available data.

Identification

- A quantity of interest is identified when access to infinite data would result in the estimate taking on only a single value
- For example, having all dummy variables in a linear model is not statistically identified because they cannot be distinguished from the intercept.
- Causal identification is what we can learn about a causal effect from available data.
- If an effect is not identified, no estimation method will recover it.

Identification

- A quantity of interest is identified when access to infinite data would result in the estimate taking on only a single value
- For example, having all dummy variables in a linear model is not statistically identified because they cannot be distinguished from the intercept.
- Causal identification is what we can learn about a causal effect from available data.
- If an effect is not identified, no estimation method will recover it.
- This means the relevant question is "what's your identification strategy?" or what are the set of assumptions that let you claim you've estimated a causal effect?

Identification

- A quantity of interest is identified when access to infinite data would result in the estimate taking on only a single value
- For example, having all dummy variables in a linear model is not statistically identified because they cannot be distinguished from the intercept.
- Causal identification is what we can learn about a causal effect from available data.
- If an effect is not identified, no estimation method will recover it.
- This means the relevant question is "what's your identification strategy?" or what are the set of assumptions that let you claim you've estimated a causal effect?
- As we will see this is not a conversation about estimation (in other words the answer cannot be "regression")

Identification vs. Estimation

- Identification: How much can you learn about the estimand if you have an infinite amount of data?
- Estimation: How much can you learn about the estimand from a finite sample?

Identification vs. Estimation

- Identification: How much can you learn about the estimand if you have an infinite amount of data?
- Estimation: How much can you learn about the estimand from a finite sample?
- Identification precedes estimation

Identification vs. Estimation

- Identification: How much can you learn about the estimand if you have an infinite amount of data?
- Estimation: How much can you learn about the estimand from a finite sample?
- Identification precedes estimation

The role of assumptions:

- Often identification requires (hopefully minimal) assumptions

Identification vs. Estimation

- Identification: How much can you learn about the estimand if you have an infinite amount of data?
- Estimation: How much can you learn about the estimand from a finite sample?
- Identification precedes estimation

The role of assumptions:

- Often identification requires (hopefully minimal) assumptions
- Even when identification is possible, estimation may impose additional assumptions (i.e. regression)

Identification vs. Estimation

- Identification: How much can you learn about the estimand if you have an infinite amount of data?
- Estimation: How much can you learn about the estimand from a finite sample?
- Identification precedes estimation

The role of assumptions:

- Often identification requires (hopefully minimal) assumptions
- Even when identification is possible, estimation may impose additional assumptions (i.e. regression)
- Law of Decreasing Credibility (Manski): The credibility of inference decreases with the strength of the assumptions maintained

Next Time

Next Time

- Next class we will talk about what kind of identification assumptions we need to estimate causal effects!

Next Time

- Next class we will talk about what kind of identification assumptions we need to estimate causal effects!
- Causal inference is tricky and I highly recommend you take a look at Morgan and Winship Chapter 1 before class.

Fun with Censorship

Fun with Censorship

- Often you don't need sophisticated methods to reveal interesting findings

Fun with Censorship

- Often you don't need sophisticated methods to reveal interesting findings
- "Ansolabahere's Law": real relationship is visible in a bivariate plot and remains in a more sophisticated in a statistical model

Fun with Censorship

- Often you don't need sophisticated methods to reveal interesting findings
- "Ansolabahere's Law": real relationship is visible in a bivariate plot and remains in a more sophisticated in a statistical model
- In other words: all inferences require both visual and mathematical evidence

Fun with Censorship

- Often you don't need sophisticated methods to reveal interesting findings
- "Ansolabahere's Law": real relationship is visible in a bivariate plot and remains in a more sophisticated in a statistical model
- In other words: all inferences require both visual and mathematical evidence
- Example: King, Pan and Roberts (2013) "How Censorship in China Allows Government Criticism but Silences Collective Expression" *American Political Science Review*

Fun with Censorship

- Often you don't need sophisticated methods to reveal interesting findings
- "Ansolabahere's Law": real relationship is visible in a bivariate plot and remains in a more sophisticated in a statistical model
- In other words: all inferences require both visual and mathematical evidence
- Example: King, Pan and Roberts (2013) "How Censorship in China Allows Government Criticism but Silences Collective Expression" *American Political Science Review*
- They use very simple (statistical) methods to great effect.

Fun with Censorship

- Often you don't need sophisticated methods to reveal interesting findings
- "Ansolabahere's Law": real relationship is visible in a bivariate plot and remains in a more sophisticated in a statistical model
- In other words: all inferences require both visual and mathematical evidence
- Example: King, Pan and Roberts (2013) "How Censorship in China Allows Government Criticism but Silences Collective Expression" *American Political Science Review*
- They use very simple (statistical) methods to great effect.
- This line of work is one of my favorites.

Sequence of slides that follow courtesy of King, Pan and Roberts

Chinese Censorship

The largest selective suppression of human expression in history

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. Stop criticism of the state

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. Stop criticism of the state
2. Stop collective action

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. Stop criticism of the state
2. Stop collective action

Either or both could be right or wrong.

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. ~~Stop criticism of the state~~ Wrong
2. Stop collective action

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship

1. ~~Stop criticism of the state~~ Wrong
2. Stop collective action Right

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship | Benefit

1. ~~Stop criticism of the state~~ Wrong
2. Stop collective action Right

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship	Benefit	Cost
1. ~~Stop criticism of the state~~ Wrong		
2. Stop collective action Right		

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship	Benefit	Cost
1. ~~Stop criticism of the state~~ Wrong	?	
2. Stop collective action Right		

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship	Benefit	Cost
1. ~~Stop criticism of the state~~ Wrong	?	Huge
2. Stop collective action Right		

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship	Benefit	Cost
1. ~~Stop criticism of the state~~ Wrong	?	Huge
2. Stop collective action Right	Huge	

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship	Benefit	Cost
1. ~~Stop criticism of the state~~ Wrong	?	Huge
2. Stop collective action Right	Huge	Small

Chinese Censorship

The largest selective suppression of human expression in history:

- implemented manually,
- by $\approx 200,000$ workers,
- located in government and inside social media firms

Theories of the Goal of Censorship	Benefit	Cost
1. ~~Stop criticism of the state~~ Wrong	?	Huge
2. Stop collective action Right	Huge	Small

(They also censor 2 other smaller categories)

Observational Study

Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months

Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months
- Random sample: 127,283

Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months
- Random sample: 127,283
- (Repeat design; Total analyzed: 11,382,221)

Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months
- Random sample: 127,283
- (Repeat design; Total analyzed: 11,382,221)
- $\rightsquigarrow$ For each post (on a timeline in one of 85 content areas):

Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months
- Random sample: 127,283
- (Repeat design; Total analyzed: 11,382,221)
- $\rightsquigarrow$ For each post (on a timeline in one of 85 content areas):
  - Download content the instant it appears

Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months
- Random sample: 127,283
- (Repeat design; Total analyzed: 11,382,221)
- $\rightsquigarrow$ For each post (on a timeline in one of 85 content areas):
  - Download content the instant it appears
  - (Carefully) revisit each later to determine if it was censored

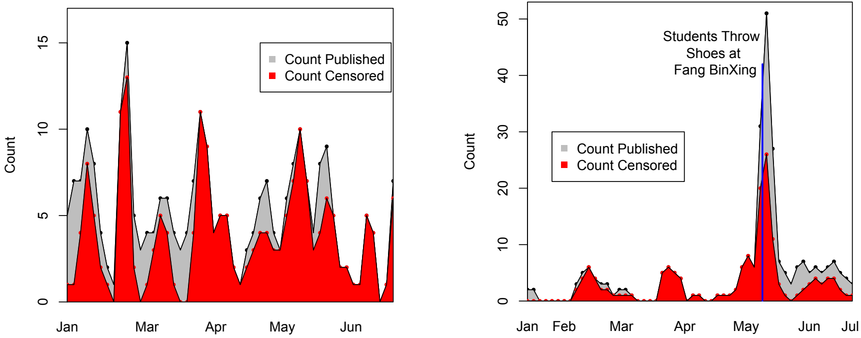
Observational Study

- Collect 3,674,698 social media posts in 85 topic areas over 6 months
- Random sample: 127,283
- (Repeat design; Total analyzed: 11,382,221)
- $\rightsquigarrow$ For each post (on a timeline in one of 85 content areas):
  - Download content the instant it appears
  - (Carefully) revisit each later to determine if it was censored
  - Use computer-assisted methods of text analysis (some existing, some new, all adapted to Chinese)

Censorship is not Ambiguous: BBS Error Page

For 2 Unusual Topics: Constant Censorship Effort

For 2 Unusual Topics: Constant Censorship Effort

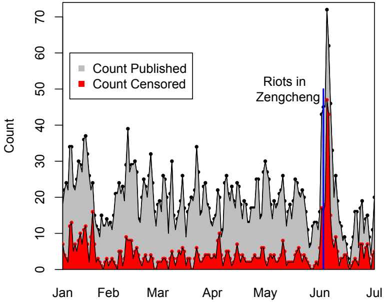


Pornography

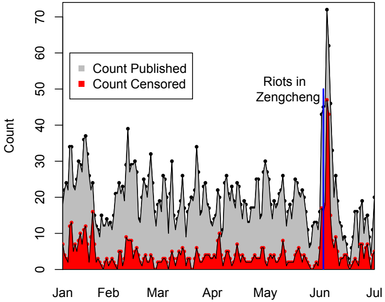
Criticism of the Censors

All other topics: Censorship & Post Volume are "Bursty"

All other topics: Censorship & Post Volume are "Bursty"

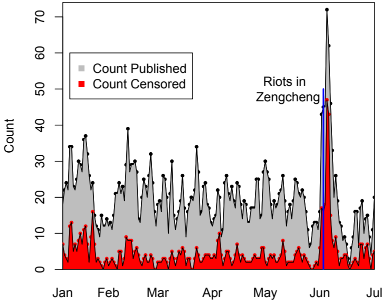


All other topics: Censorship & Post Volume are "Bursty"



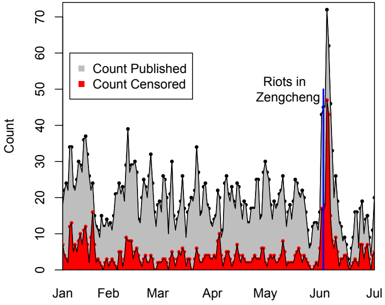
- Unit of analysis:

All other topics: Censorship & Post Volume are "Bursty"



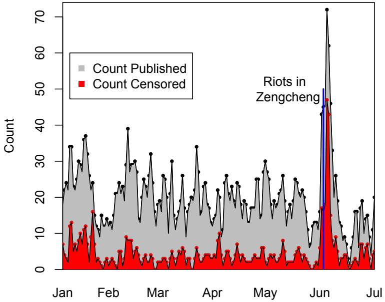
- Unit of analysis:
  - volume burst

All other topics: Censorship & Post Volume are "Bursty"



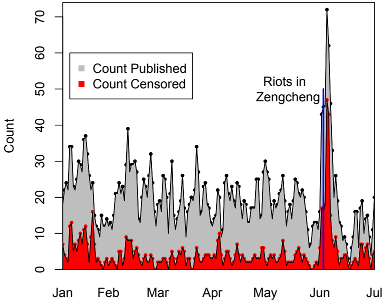
- Unit of analysis:
  - volume burst
  - (≈ 3 SDs greater than baseline volume)

All other topics: Censorship & Post Volume are “Bursty”



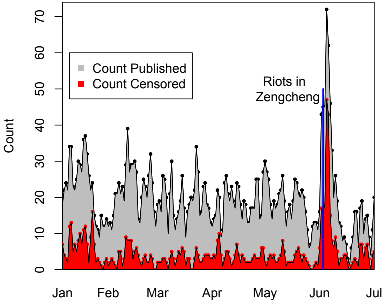
- Unit of analysis:
  - volume burst
  - ($\approx$ 3 SDs greater than baseline volume)
- They monitored 85 topic areas (Jan–July 2011)

All other topics: Censorship & Post Volume are "Bursty"



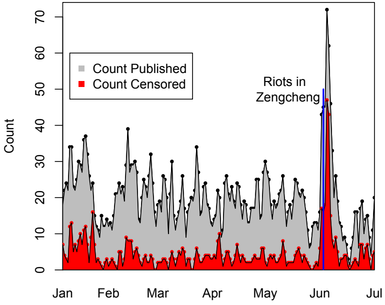
- Unit of analysis:
  - volume burst
  - ($\approx$ 3 SDs greater than baseline volume)
- They monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total

All other topics: Censorship & Post Volume are "Bursty"



- Unit of analysis:
  - volume burst
  - ($\approx$ 3 SDs greater than baseline volume)
- They monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
- Identified real world events associated with each burst

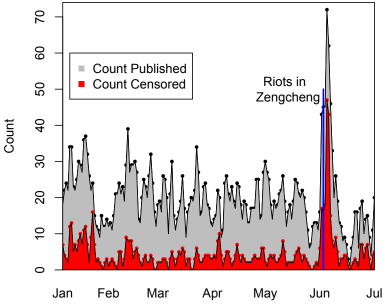
All other topics: Censorship & Post Volume are "Bursty"



- Unit of analysis:
  - volume burst
  - ($\approx$ 3 SDs greater than baseline volume)
- They monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
- Identified real world events associated with each burst

Their hypothesis: The government censors all posts in volume bursts associated with events with collective action (regardless of how critical or supportive of the state)

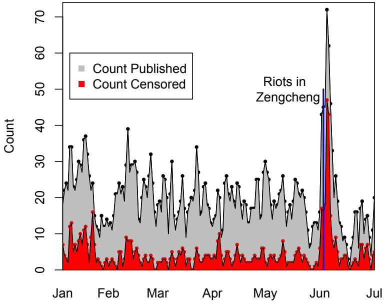
All other topics: Censorship & Post Volume are "Bursty"



- Unit of analysis:
  - volume burst
  - (≈ 3 SDs greater than baseline volume)
- They monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
- Identified real world events associated with each burst

Their hypothesis: The government censors all posts in volume bursts associated with events with collective action (regardless of how critical or supportive of the state)

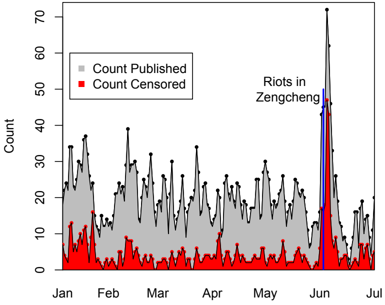
All other topics: Censorship & Post Volume are "Bursty"



- Unit of analysis:
  - volume burst
  - (≈ 3 SDs greater than baseline volume)
- They monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
- Identified real world events associated with each burst

Their hypothesis: The government censors all posts in volume bursts associated with events with collective action (regardless of how critical or supportive of the state)

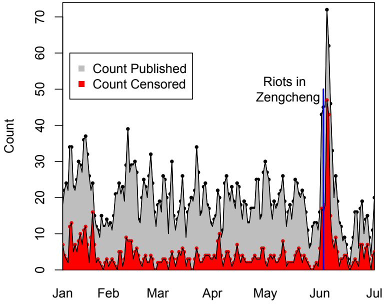
All other topics: Censorship & Post Volume are "Bursty"



- Unit of analysis:
  - volume burst
  - ($\approx$ 3 SDs greater than baseline volume)
- They monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
- Identified real world events associated with each burst

Their hypothesis: The government censors all posts in volume bursts associated with events with collective action (regardless of how critical or supportive of the state)

All other topics: Censorship & Post Volume are "Bursty"



- Unit of analysis:
  - volume burst
  - ($\approx$ 3 SDs greater than baseline volume)
- They monitored 85 topic areas (Jan–July 2011)
- Found 87 volume bursts in total
- Identified real world events associated with each burst

Their hypothesis: The government censors all posts in volume bursts associated with events with collective action (regardless of how critical or supportive of the state)

Observational Test 1: Post Volume

Observational Test 1: Post Volume

- Begin with 87 volume bursts in 85 topics areas

Observational Test 1: Post Volume

- Begin with 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude

Observational Test 1: Post Volume

- Begin with 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
- If goal of censorship is to stop collective action, they expect:

Observational Test 1: Post Volume

- Begin with 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
- If goal of censorship is to stop collective action, they expect:

1. On average, % censored should increase during volume bursts

Observational Test 1: Post Volume

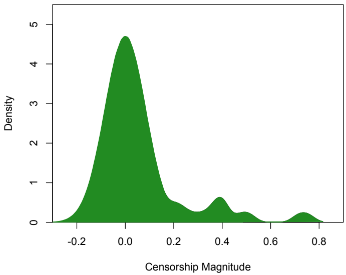
- Begin with 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
- If goal of censorship is to stop collective action, they expect:

1. On average, % censored should increase during volume bursts
2. Some bursts (associated with politically relevant events) should have much higher censorship

Observational Test 1: Post Volume

- Begin with 87 volume bursts in 85 topics areas
- For each burst, calculate change in % censorship inside to outside each volume burst within topic areas – censorship magnitude
- If goal of censorship is to stop collective action, they expect:

1. On average, % censored should increase during volume bursts
2. Some bursts (associated with politically relevant events) should have much higher censorship



Observational Test 2: The Event Generating Volume Bursts

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. Collective Action Potential

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. Criticism of censors

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. Criticism of censors
3. Pornography

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. Criticism of censors
3. Pornography
4. (Other) News

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. Collective Action Potential
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. Criticism of censors
3. Pornography
4. (Other) News
5. Government Policies

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. ~~Collective Action Potential~~ CENSORED
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. Criticism of censors
3. Pornography
4. (Other) News
5. Government Policies

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. ~~Collective Action Potential~~ CENSORED
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. ~~Criticism of censors~~ CENSORED
3. Pornography
4. (Other) News
5. Government Policies

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be +, −, or neutral comments about the state)

1. ~~Collective Action Potential~~ CENSORED
    - protest or organized crowd formation outside the Internet
    - individuals who have organized or incited collective action on the ground in the past;
    - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. ~~Criticism of censors~~ CENSORED
3. ~~Pornography~~ CENSORED
4. (Other) News
5. Government Policies

Observational Test 2: The Event Generating Volume Bursts

Event classification (each category can be $+$, $-$, or neutral comments about the state)

1. ~~Collective Action Potential~~ 

    - protest or organized crowd formation outside the Internet
    - individuals who have organized or incited collective action on the ground in the past;
    - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. ~~Criticism of censors~~ 

3. ~~Pornography~~ 

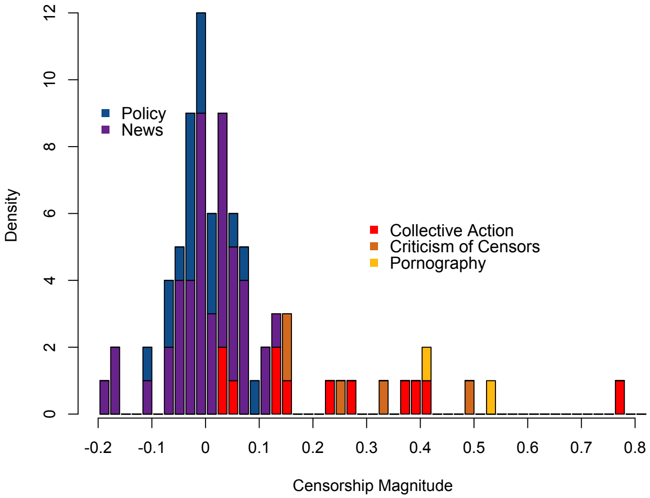
4. (Other) News 
5. Government Policies

Observational Test 2: The Event Generating Volume Bursts

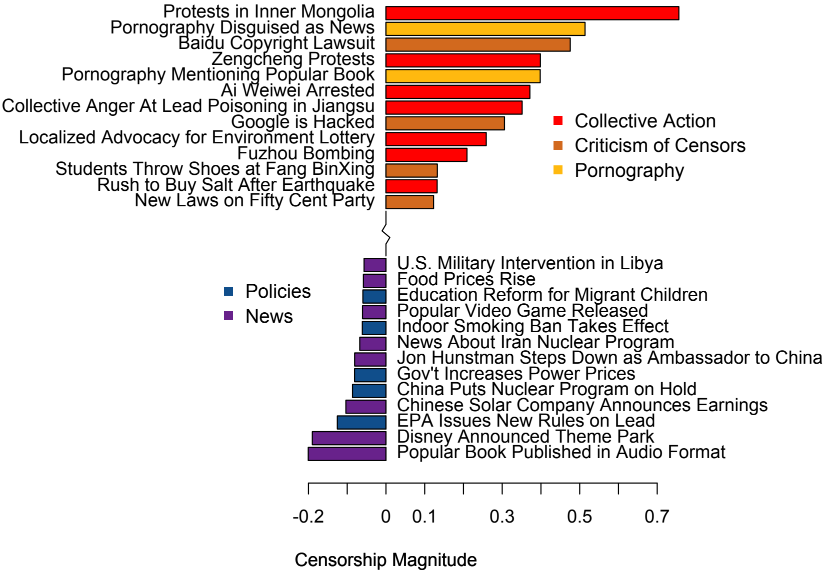
Event classification (each category can be +, −, or neutral comments about the state)

1. ~~Collective Action Potential~~ CENSORED
   - protest or organized crowd formation outside the Internet
   - individuals who have organized or incited collective action on the ground in the past;
   - topics related to nationalism or nationalist sentiment that have incited protest or collective action in the past.
2. ~~Criticism of censors~~ CENSORED
3. ~~Pornography~~ CENSORED
4. (Other) News
5. Government Policies

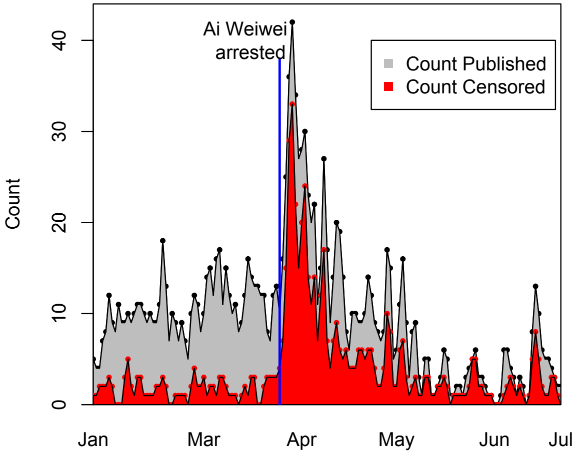
What Types of Events Are Censored?



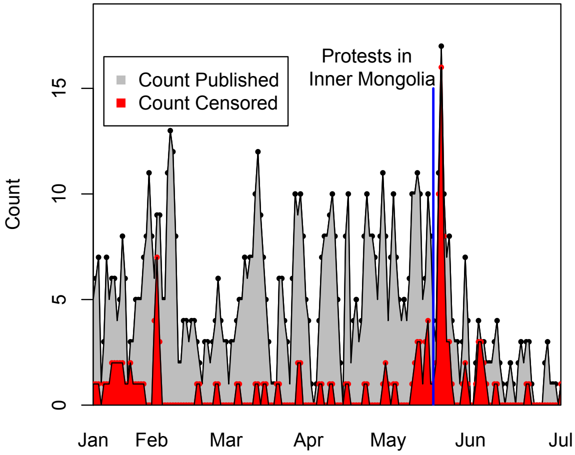
What Types of Events Are Censored?



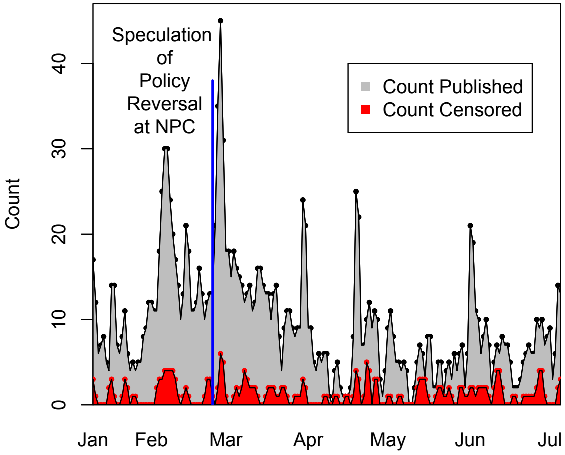
Censoring Collective Action: Ai Weiwei's Arrest



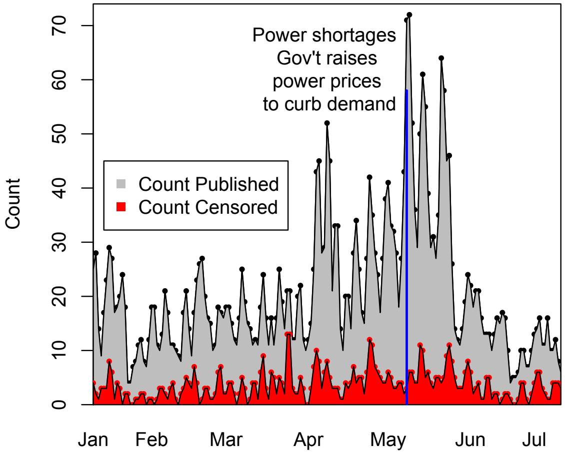
Censoring Collective Action: Protests in Inner Mongolia



Low Censorship on One Child Policy



Low Censorship on News: Power Prices



References

This Lecture:

- Gelman and Carlin (2014). "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors"
- Kastellec and Leoni (2007). "Using Graphs Instead of Tables in Political Science." *Perspectives on Politics*
- King, Pan and Roberts (2013) "How Censorship in China Allows Government Criticism but Silences Collective Expression"
- King, G. Tomz, M., and Wittenberg, J. (2000). "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science*
- Nunzo, R. (2014) "Scientific method: Statistical errors" *Nature*.

Where We've Been and Where We're Going...

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences
  - Wednesday:
    - causal inference

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences
  - Wednesday:
    - causal inference
- Next Week

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences
  - Wednesday:
    - causal inference
- Next Week
  - regression diagnostics

Where We've Been and Where We're Going...

- Last Week
  - matrix form of linear regression
  - inference and F-tests
- This Week
  - Monday:
    - making an argument in social sciences
  - Wednesday:
    - causal inference
- Next Week
  - regression diagnostics
- Long Run
  - regression $\rightarrow$ diagnostics $\rightarrow$ causal inference

Questions?

Causation

What's a cause?

Causation

What's a cause?

- Time precedence

Causation

What's a cause?

- Time precedence
- Constant Conjunction

Causation

What's a cause?

- Time precedence
- Constant Conjunction
  - Correlation is not causation

Causation

What's a cause?

- Time precedence
- Constant Conjunction
  - Correlation is not causation
  - We will see that correlation based definitions can lead to consideration of nonsensical causes

Causation

What's a cause?

- Time precedence
- Constant Conjunction
  - Correlation is not causation
  - We will see that correlation based definitions can lead to consideration of nonsensical causes
- Method of difference

Causation

What's a cause?

- Time precedence
- Constant Conjunction
  - Correlation is not causation
  - We will see that correlation based definitions can lead to consideration of nonsensical causes
- Method of difference
  - Identify identical units, except for treatment. Attribute cause to difference

Causation

What's a cause?

- Time precedence
- Constant Conjunction
  - Correlation is not causation
  - We will see that correlation based definitions can lead to consideration of nonsensical causes
- Method of difference
  - Identify identical units, except for treatment. Attribute cause to difference
- Granger Cause

Causation

What's a cause?

- Time precedence
- Constant Conjunction
  - Correlation is not causation
  - We will see that correlation based definitions can lead to consideration of nonsensical causes
- Method of difference
  - Identify identical units, except for treatment. Attribute cause to difference
- Granger Cause
  - Forecast based definition of cause

Causation

What's a cause?

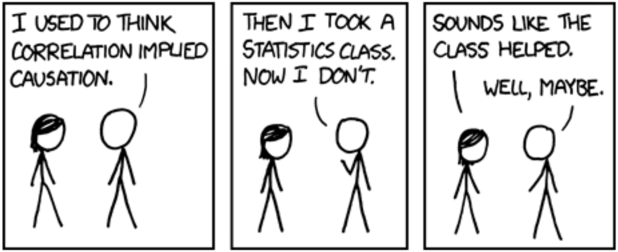
- Time precedence
- Constant Conjunction
  - Correlation is not causation
  - We will see that correlation based definitions can lead to consideration of nonsensical causes
- Method of difference
  - Identify identical units, except for treatment. Attribute cause to difference
- Granger Cause
  - Forecast based definition of cause
  - Problem of common causes

Causation

What's a cause?

- Time precedence
- Constant Conjunction
  - Correlation is not causation
  - We will see that correlation based definitions can lead to consideration of nonsensical causes
- Method of difference
  - Identify identical units, except for treatment. Attribute cause to difference
- Granger Cause
  - Forecast based definition of cause
  - Problem of common causes

Causation


I USED TO THINK CORRELATION IMPLIED CAUSATION.
THEN I TOOK A STATISTICS CLASS. NOW I DON'T.
SOUNDS LIKE THE CLASS HELPED.
WELL, MAYBE.

Neyman-Rubin Potential Outcomes Model

Figure: Neyman

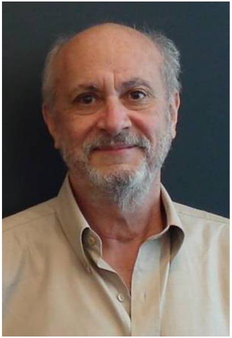


Figure: Rubin

Neyman-Rubin Model

Two possible conditions:

Neyman-Rubin Model

Two possible conditions:

- Treatment condition $T = 1$

Neyman-Rubin Model

Two possible conditions:

- Treatment condition $T = 1$
- Control condition $T = 0$

Neyman-Rubin Model

Two possible conditions:

- Treatment condition $T = 1$
- Control condition $T = 0$

Suppose that we have an individual i.

Neyman-Rubin Model

Two possible conditions:

- Treatment condition $T = 1$
- Control condition $T = 0$

Suppose that we have an individual i.

Key assumption: we can imagine a world where individual i is assigned to treatment and control conditions

Neyman-Rubin Model

Two possible conditions:

- Treatment condition $T = 1$
- Control condition $T = 0$

Suppose that we have an individual i.

Key assumption: we can imagine a world where individual i is assigned to treatment and control conditions

Potential Outcomes: responses under each condition, $Y_i(T)$

Neyman-Rubin Model

Two possible conditions:

- Treatment condition $T = 1$
- Control condition $T = 0$

Suppose that we have an individual i.

Key assumption: we can imagine a world where individual i is assigned to treatment and control conditions

Potential Outcomes: responses under each condition, $Y_i(T)$

- Response under treatment $Y_i(1)$

Neyman-Rubin Model

Two possible conditions:

- Treatment condition $T = 1$
- Control condition $T = 0$

Suppose that we have an individual i.

Key assumption: we can imagine a world where individual i is assigned to treatment and control conditions

Potential Outcomes: responses under each condition, $Y_i(T)$

- Response under treatment $Y_i(1)$
- Response under control $Y_i(0)$

Neyman-Rubin Model

Two possible conditions:

- Treatment condition $T = 1$
- Control condition $T = 0$

Suppose that we have an individual i.

Key assumption: we can imagine a world where individual i is assigned to treatment and control conditions

Potential Outcomes: responses under each condition, $Y_i(T)$

- Response under treatment $Y_i(1)$
- Response under control $Y_i(0)$

Definition: no differences between treatment and control worlds

A Concrete Example

A Concrete Example

Job Training Programs:

A Concrete Example

Job Training Programs:

- Treatment: Receive job training ($T = 1$)

A Concrete Example

Job Training Programs:

- Treatment: Receive job training ($T = 1$)
- Control: No Job training ($T = 0$)

A Concrete Example

Job Training Programs:

- Treatment: Receive job training ($T = 1$)
- Control: No Job training ($T = 0$)

A Concrete Example

Job Training Programs:

- Treatment: Receive job training ($T = 1$)
- Control: No Job training ($T = 0$)

Response: Income (Dollars per year)

A Concrete Example

Job Training Programs:

- Treatment: Receive job training ($T = 1$)
- Control: No Job training ($T = 0$)

Response: Income (Dollars per year)

	Treatment ($Y_i(1)$)	Control ($Y_i(0)$)
Person 1	45,000	32,000
Person 2	54,000	45,000
Person 3	34,000	34,000

A Definition of Causation

A Definition of Causation

The individual causal effect of treatment T for individual i is,

A Definition of Causation

The individual causal effect of treatment T for individual i is,

$$\text{Individual Causal Effect}_i \;=\; Y_i(1) - Y_i(0)$$

A Definition of Causation

The individual causal effect of treatment T for individual i is,

$$\text{Individual Causal Effect}_i \quad = \quad Y_i(1) - Y_i(0)$$

Causal effect:
Difference in individual i's potential outcomes

A Definition of Causation

The individual causal effect of treatment T for individual i is,

$$\text{Individual Causal Effect}_i = Y_i(1) - Y_i(0)$$

Causal effect:

Difference in individual i's potential outcomes

Compare responses in hypothetical worlds

Fundamental Problem of Causal Inference

Fundamental Problem of Causal Inference

Causal Effect:

Fundamental Problem of Causal Inference

Causal Effect:

Response Under Treatment - Response Under Control

Fundamental Problem of Causal Inference

Causal Effect:

Response Under Treatment - Response Under Control

Job Training Program:

Fundamental Problem of Causal Inference

Causal Effect:

Response Under Treatment - Response Under Control

Job Training Program:

	Treatment ($Y_i(1)$)	Control ($Y_i(0)$)
Person 1	45,000	32,000
Person 2	54,000	45,000
Person 3	34,000	34,000

Fundamental Problem of Causal Inference

Causal Effect:

Response Under Treatment - Response Under Control

Job Training Program:

	Treatment ($Y_i(1)$)	Control ($Y_i(0)$)
Person 1	?	32,000
Person 2	54,000	?
Person 3	34,000	?

Fundamental Problem of Causal Inference

Causal Effect:

Response Under Treatment - Response Under Control

Job Training Program:

	Treatment ($Y_i(1)$)	Control ($Y_i(0)$)
Person 1	?	32,000
Person 2	54,000	?
Person 3	34,000	?

Fundamental Problem of Causal Inference (Holland (1986)):

Fundamental Problem of Causal Inference

Causal Effect:

Response Under Treatment - Response Under Control

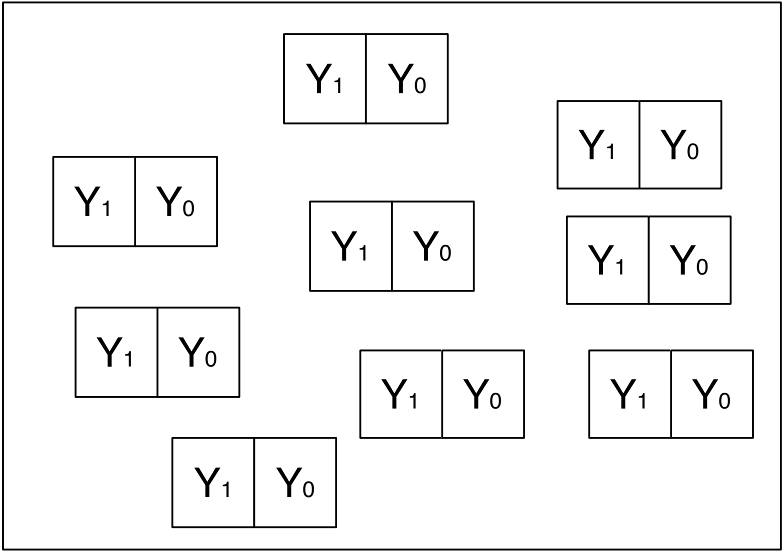
Job Training Program:

	Treatment ($Y_i(1)$)	Control ($Y_i(0)$)
Person 1	?	32,000
Person 2	54,000	?
Person 3	34,000	?

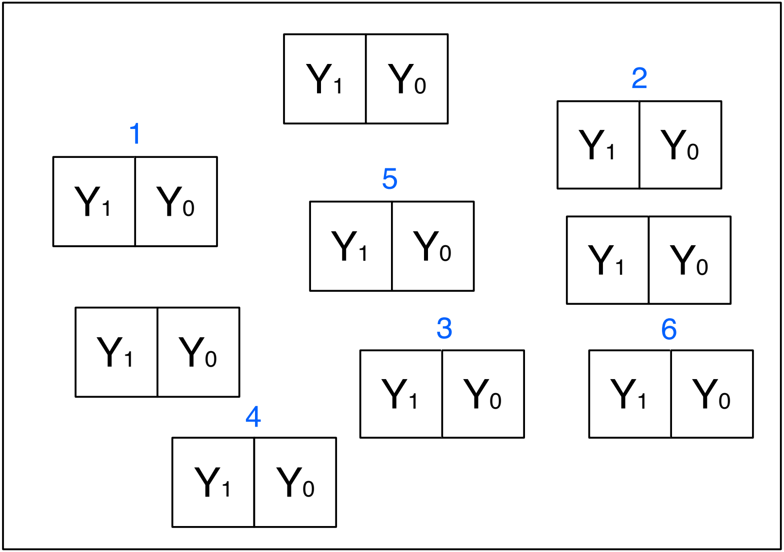
Fundamental Problem of Causal Inference (Holland (1986)):

It is impossible to observe both $Y_i(1)$ and $Y_i(0)$

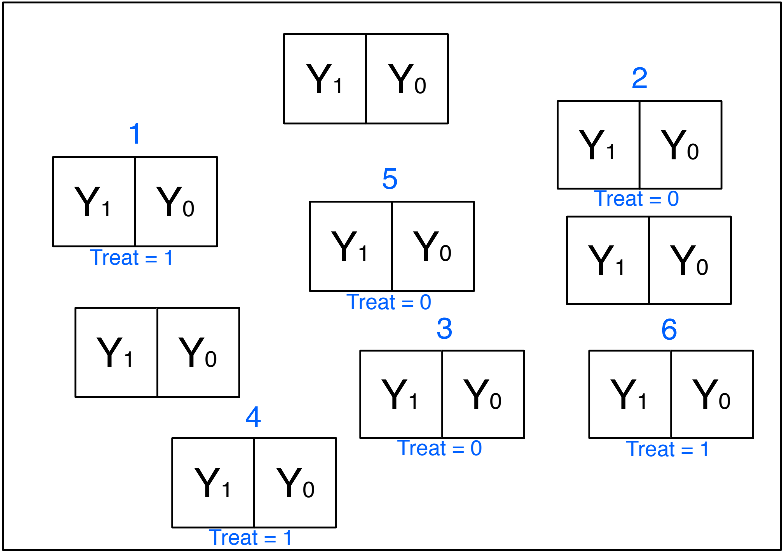
Neyman Urn Model



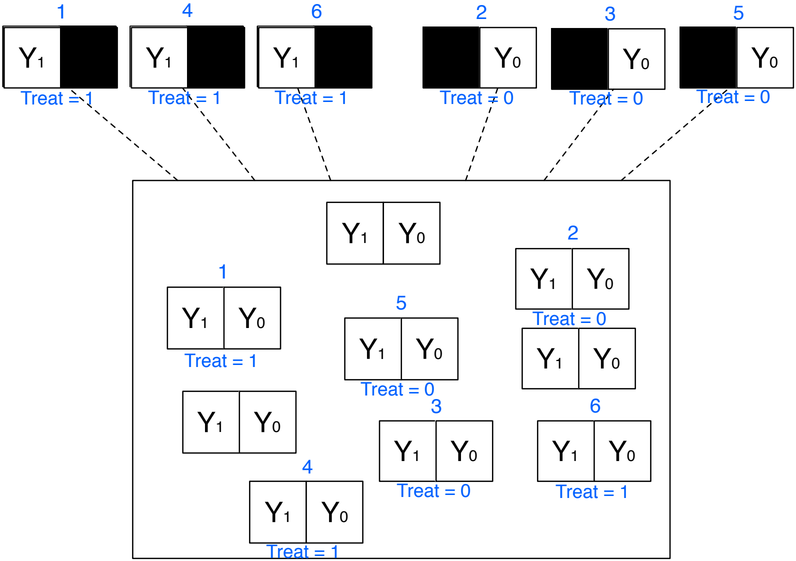
Neyman Urn Model



Neyman Urn Model



Neyman Urn Model



Some Useful Terms

Some Useful Terms

Definition (Treatment)

D_i: Indicator of treatment intake for unit i

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

Some Useful Terms

Definition (Treatment)

D_i: Indicator of treatment intake for unit i

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

Definition (Outcome)

Y_i: Observed outcome variable of interest for unit i. The treatment occurs temporally before the outcomes.

Some Useful Terms

Definition (Treatment)

D_i: Indicator of treatment intake for unit i

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

Definition (Outcome)

Y_i: Observed outcome variable of interest for unit i. The treatment occurs temporally before the outcomes.

Definition (Potential Outcome)

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

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

Some Useful Terms

Some Useful Terms

Definition (Causal Effect)

Causal effect of the treatment on the outcome for unit i is the difference between its two potential outcomes:

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

Some Useful Terms

Definition (Causal Effect)

Causal effect of the treatment on the outcome for unit i is the difference between its two potential outcomes:

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

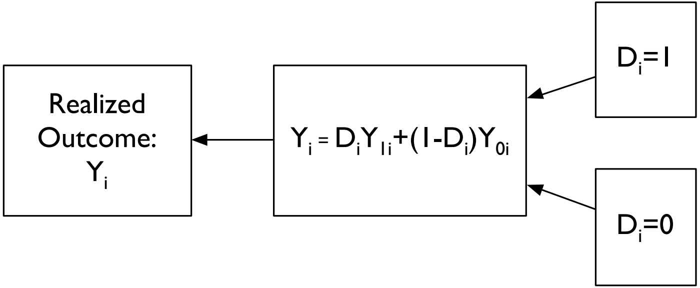
Assumption

Observed outcomes are realized as

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i} \textit{ so } Y_i = \begin{cases} Y_{1i} & \textit{if } D_i = 1 \\ Y_{0i} & \textit{if } D_i = 0 \end{cases}$$

Causal Inference as a Missing Data Problem

Causal Inference as a Missing Data Problem



Fundamental Problem of Causal Inference

Cannot observe both potential outcomes, so we how can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

Causal Inference as a Missing Data Problem

Causal Inference as a Missing Data Problem

Causal inference is difficult because it involves missing data. How can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

Causal Inference as a Missing Data Problem

Causal inference is difficult because it involves missing data. How can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

- Homogeneity is one solution:

Causal Inference as a Missing Data Problem

Causal inference is difficult because it involves missing data. How can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

- Homogeneity is one solution:
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across individuals, then cross-sectional comparisons will recover τ_i

Causal Inference as a Missing Data Problem

Causal inference is difficult because it involves missing data. How can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

- Homogeneity is one solution:
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across individuals, then cross-sectional comparisons will recover τ_i
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across time, then before and after comparisons will recover τ_i

Causal Inference as a Missing Data Problem

Causal inference is difficult because it involves missing data. How can we calculate $\tau_i = Y_{1i} - Y_{0i}$?

- Homogeneity is one solution:
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across individuals, then cross-sectional comparisons will recover τ_i
  - If $\{Y_{1i}, Y_{0i}\}$ is constant across time, then before and after comparisons will recover τ_i

In social phenomenon, unfortunately, homogeneity is very rare.

The Selection Problem

- Why is this difficult? selection bias

The Selection Problem

- Why is this difficult? selection bias
- The core idea is that the people who get treatment might look different from those who get control and thus they are not good counterfactuals for each other.

The Selection Problem

- Why is this difficult? selection bias
- The core idea is that the people who get treatment might look different from those who get control and thus they are not good counterfactuals for each other.
- Let's look at what we get from a naive difference in means with a binary treatment:

The Selection Problem

- Why is this difficult? selection bias
- The core idea is that the people who get treatment might look different from those who get control and thus they are not good counterfactuals for each other.
- Let's look at what we get from a naive difference in means with a binary treatment:

$$
\begin{aligned}
&E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]
\end{aligned}
$$

The Selection Problem

- Why is this difficult? selection bias
- The core idea is that the people who get treatment might look different from those who get control and thus they are not good counterfactuals for each other.
- Let's look at what we get from a naive difference in means with a binary treatment:

$$
\begin{aligned}
&E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]
\end{aligned}
$$

The Selection Problem

- Why is this difficult? selection bias
- The core idea is that the people who get treatment might look different from those who get control and thus they are not good counterfactuals for each other.
- Let's look at what we get from a naive difference in means with a binary treatment:

$$
\begin{aligned}
& E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
& \quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
& \quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
& \quad = \underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{\text{Average Treatment Effect on Treated}} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

The Selection Problem

- Why is this difficult? selection bias
- The core idea is that the people who get treatment might look different from those who get control and thus they are not good counterfactuals for each other.
- Let's look at what we get from a naive difference in means with a binary treatment:

$$
\begin{aligned}
&E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\quad = \underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{\text{Average Treatment Effect on Treated}} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

The Selection Problem

- Why is this difficult? selection bias
- The core idea is that the people who get treatment might look different from those who get control and thus they are not good counterfactuals for each other.
- Let's look at what we get from a naive difference in means with a binary treatment:

$$
\begin{aligned}
&E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\quad = \underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{\text{Average Treatment Effect on Treated}} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

- Naive estimator = Average Treatment Effect on Treated + Selection Bias

The Selection Problem

- Why is this difficult? selection bias
- The core idea is that the people who get treatment might look different from those who get control and thus they are not good counterfactuals for each other.
- Let's look at what we get from a naive difference in means with a binary treatment:

$$
\begin{aligned}
&E[Y_i|D_i = 1] - E[Y_i|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\quad = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 1] + E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0] \\
&\quad = \underbrace{E[Y_i(1) - Y_i(0)|D_i = 1]}_{\text{Average Treatment Effect on Treated}} + \underbrace{E[Y_i(0)|D_i = 1] - E[Y_i(0)|D_i = 0]}_{\text{selection bias}}
\end{aligned}
$$

- Naive estimator = Average Treatment Effect on Treated + Selection Bias
- Selection bias: how different the treated and control groups are in terms of their potential outcome under control.

Selection Makes Us Care About Assignment Mechanisms

Selection Makes Us Care About Assignment Mechanisms

Assignment Mechanism

"The process that determines which units receive which treatments, hence which potential outcomes are realized and thus can be observed, and, conversely, which potential outcomes are missing."
(Imbens and Rubin, 2015, p. 31)

Selection Makes Us Care About Assignment Mechanisms

Assignment Mechanism

"The process that determines which units receive which treatments, hence which potential outcomes are realized and thus can be observed, and, conversely, which potential outcomes are missing."
(Imbens and Rubin, 2015, p. 31)

Key Assumptions:

- Individualistic assignment: Limits the dependence of a particular unit's assignment probability on the values of the covariates and potential outcomes for other units

Selection Makes Us Care About Assignment Mechanisms

Assignment Mechanism

"The process that determines which units receive which treatments, hence which potential outcomes are realized and thus can be observed, and, conversely, which potential outcomes are missing."
(Imbens and Rubin, 2015, p. 31)

Key Assumptions:

- Individualistic assignment: Limits the dependence of a particular unit's assignment probability on the values of the covariates and potential outcomes for other units
- Probabilistic assignment: Requires the assignment mechanism to imply a non-zero probability for each treatment value, for every unit

Selection Makes Us Care About Assignment Mechanisms

Assignment Mechanism

"The process that determines which units receive which treatments, hence which potential outcomes are realized and thus can be observed, and, conversely, which potential outcomes are missing."
(Imbens and Rubin, 2015, p. 31)

Key Assumptions:

- Individualistic assignment: Limits the dependence of a particular unit's assignment probability on the values of the covariates and potential outcomes for other units
- Probabilistic assignment: Requires the assignment mechanism to imply a non-zero probability for each treatment value, for every unit
- Unconfounded assignment: Disallows dependence of the assignment mechanism on the potential outcomes

Assumptions: Be Careful When Defining Treatment

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...
   - Drug trial

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

- Drug trial
- Private Schooling

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

- Drug trial
- Private Schooling
- Practical Advice: are there hidden versions of your treatment? (suggests different interpretations)

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

- Drug trial
- Private Schooling
- Practical Advice: are there hidden versions of your treatment? (suggests different interpretations)

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

- Drug trial
- Private Schooling
- Practical Advice: are there hidden versions of your treatment? (suggests different interpretations)

2) Potential outcomes depend only on my treatment status ($Y(1)$, not $Y(1, 0, 0, 1, 0, \ldots, 0, 1)$ or $Y(\boldsymbol{T})$)

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

- Drug trial
- Private Schooling
- Practical Advice: are there hidden versions of your treatment? (suggests different interpretations)

2) Potential outcomes depend only on my treatment status ($Y(1)$, not $Y(1, 0, 0, 1, 0, \ldots, 0, 1)$ or $Y(\boldsymbol{T})$)

- Survey experiment

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

- Drug trial
- Private Schooling
- Practical Advice: are there hidden versions of your treatment? (suggests different interpretations)

2) Potential outcomes depend only on my treatment status ($Y(1)$, not $Y(1, 0, 0, 1, 0, \ldots, 0, 1)$ or $Y(\boldsymbol{T})$)

- Survey experiment
- AIDS drug trials

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

- Drug trial
- Private Schooling
- Practical Advice: are there hidden versions of your treatment? (suggests different interpretations)

2) Potential outcomes depend only on my treatment status ($Y(1)$, not $Y(1, 0, 0, 1, 0, \ldots, 0, 1)$ or $Y(\boldsymbol{T})$)

- Survey experiment
- AIDS drug trials
- Practical Advice: design study to avoid spillovers and contamination (unless question of interest, see Nickerson (2008) and Gerber, Green, and Larimer (2008))

Assumptions: Be Careful When Defining Treatment

1) There is only one version of the treatment, not T_1, T_2, ...

- Drug trial
- Private Schooling
- Practical Advice: are there hidden versions of your treatment? (suggests different interpretations)

2) Potential outcomes depend only on my treatment status ($Y(1)$, not $Y(1, 0, 0, 1, 0, \ldots, 0, 1)$ or $Y(\boldsymbol{T})$)

- Survey experiment
- AIDS drug trials
- Practical Advice: design study to avoid spillovers and contamination (unless question of interest, see Nickerson (2008) and Gerber, Green, and Larimer (2008))

Together: (1) and (2) constitute:
SUTVA: Stable Unit Treatment Value Assumption
Also sometimes referred to as the "No Interference" assumption.

The Trouble with Interference

The Trouble with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be a vector of treatment assignments for two units i (me) and j (you).

The Trouble with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be a vector of treatment assignments for two units i (me) and j (you).

How many elements in $\mathbf{D}$?

The Trouble with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be a vector of treatment assignments for two units i (me) and j (you).

How many elements in $\mathbf{D}$?

$$\mathbf{D} = \{(D_i = 0, D_j = 0), (D_i = 1, D_j = 0), (D_i = 0, D_j = 1), (D_i = 1, D_j = 1)\}$$

The Trouble with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be a vector of treatment assignments for two units i (me) and j (you).

How many elements in $\mathbf{D}$?

$$\mathbf{D} = \{(D_i = 0, D_j = 0), (D_i = 1, D_j = 0), (D_i = 0, D_j = 1), (D_i = 1, D_j = 1)\}$$

How many potential outcomes for unit i?

The Trouble with Interference

Let $\mathbf{D} = \{D_i, D_j\}$ be a vector of treatment assignments for two units i (me) and j (you).

How many elements in $\mathbf{D}$?

$$\mathbf{D} = \{(D_i = 0, D_j = 0), (D_i = 1, D_j = 0), (D_i = 0, D_j = 1), (D_i = 1, D_j = 1)\}$$

How many potential outcomes for unit i?

$$Y_{1i}(\mathbf{D}) = \begin{cases} Y_{1i}(1,1) \\ Y_{1i}(1,0) \end{cases} \quad Y_{0i}(\mathbf{D}) = \begin{cases} Y_{0i}(0,1) \\ Y_{0i}(0,0) \end{cases}$$

Potential Outcomes with Interference

How many causal effects for unit i?

Potential Outcomes with Interference

How many causal effects for unit i?

$$\tau_i(\mathbf{D}) = \begin{cases} Y_{1i}(1,1) - Y_{0i}(0,0) \\ Y_{1i}(1,1) - Y_{0i}(0,1) \\ Y_{1i}(1,0) - Y_{0i}(0,0) \\ Y_{1i}(1,0) - Y_{0i}(0,1) \\ Y_{1i}(1,1) - Y_{1i}(1,0) \\ Y_{0i}(0,1) - Y_{0i}(0,0) \end{cases}$$

Potential Outcomes with Interference

How many causal effects for unit i?

$$\tau_i(\mathbf{D}) = \begin{cases} Y_{1i}(1,1) - Y_{0i}(0,0) \\ Y_{1i}(1,1) - Y_{0i}(0,1) \\ Y_{1i}(1,0) - Y_{0i}(0,0) \\ Y_{1i}(1,0) - Y_{0i}(0,1) \\ Y_{1i}(1,1) - Y_{1i}(1,0) \\ Y_{0i}(0,1) - Y_{0i}(0,0) \end{cases}$$

How many potential outcome are observed for unit i?

Potential Outcomes with Interference

How many causal effects for unit i?

$$\tau_i(\mathbf{D}) = \begin{cases} Y_{1i}(1,1) - Y_{0i}(0,0) \\ Y_{1i}(1,1) - Y_{0i}(0,1) \\ Y_{1i}(1,0) - Y_{0i}(0,0) \\ Y_{1i}(1,0) - Y_{0i}(0,1) \\ Y_{1i}(1,1) - Y_{1i}(1,0) \\ Y_{0i}(0,1) - Y_{0i}(0,0) \end{cases}$$

How many potential outcome are observed for unit i?

Since we only observe one of the four potential outcomes, the missing data problem for causal inference is even more severe.

Potential Outcomes with Interference

Potential Outcomes with Interference

The No Interference assumption states that unit i's potential outcomes depends on D_i, not $\mathbf{D}$:

$Y_{1i}(1,1) = Y_{1i}(1,0)$ and $Y_{0i}(0,1) = Y_{0i}(0,0)$

Potential Outcomes with Interference

The No Interference assumption states that unit i's potential outcomes depends on D_i, not $\mathbf{D}$:

$Y_{1i}(1,1) = Y_{1i}(1,0)$ and $Y_{0i}(0,1) = Y_{0i}(0,0)$

This assumption furthermore allows us to define the effect for unit i as $\tau_i = Y_{1i} - Y_{0i}$.

Potential Outcomes with Interference

The No Interference assumption states that unit i's potential outcomes depends on D_i, not $\mathbf{D}$:

$Y_{1i}(1,1) = Y_{1i}(1,0)$ and $Y_{0i}(0,1) = Y_{0i}(0,0)$

This assumption furthermore allows us to define the effect for unit i as $\tau_i = Y_{1i} - Y_{0i}$.

No interference is an example of an **exclusion restriction**. We rely on outside information to rule out the possibility of certain causal effects (eg. you taking the treatment has no effect on my potential outcomes).

Potential Outcomes with Interference

Potential Outcomes with Interference

Some Examples of Interference:

Potential Outcomes with Interference

Some Examples of Interference:

- Contagion
- Displacement
- Communication
- Deterrence

Potential Outcomes with Interference

Some Examples of Interference:

- Contagion
- Displacement
- Communication
- Deterrence

Causal inference in the presence of interference between subjects is an area of active research. Specially tailored experimental designs have been developed to study these interactions, e.g. Miguel and Kremer (2004) and Sinclair, McConnell, and Green (2012).

What Gets to Be a Cause?

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

Problem: Immutable (or difficult to change) characteristics

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

Problem: Immutable (or difficult to change) characteristics

- Effect of gender on promotion

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

Problem: Immutable (or difficult to change) characteristics

- Effect of gender on promotion
- Effect of race on income

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

Problem: Immutable (or difficult to change) characteristics

- Effect of gender on promotion
- Effect of race on income

Consider causal effect of gender on promotion:

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

Problem: Immutable (or difficult to change) characteristics

- Effect of gender on promotion
- Effect of race on income

Consider causal effect of gender on promotion:

- Do we mean gender reassignment surgery?

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

Problem: Immutable (or difficult to change) characteristics

- Effect of gender on promotion
- Effect of race on income

Consider causal effect of gender on promotion:

- Do we mean gender reassignment surgery?
- Do we mean randomly assigning at birth? (a lot of other stuff different)

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

Problem: Immutable (or difficult to change) characteristics

- Effect of gender on promotion
- Effect of race on income

Consider causal effect of gender on promotion:

- Do we mean gender reassignment surgery?
- Do we mean randomly assigning at birth? (a lot of other stuff different)
- one idea: manipulate perceptions–women evaluated differently on paper

What Gets to Be a Cause?

We can imagine a world where individual i is assigned to treatment and control conditions

What is the Hypothetical Experiment?

Problem: Immutable (or difficult to change) characteristics

- Effect of gender on promotion
- Effect of race on income

Consider causal effect of gender on promotion:

- Do we mean gender reassignment surgery?
- Do we mean randomly assigning at birth? (a lot of other stuff different)
- one idea: manipulate perceptions–women evaluated differently on paper

No Causation Without Manipulation

Caveats and Implications

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society
  - Suggests: we need a different empirical strategy to evaluate claims

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society
  - Suggests: we need a different empirical strategy to evaluate claims
  - What facet of institutionalized racism (or its consequences) causes racial disparities?

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society
  - Suggests: we need a different empirical strategy to evaluate claims
  - What facet of institutionalized racism (or its consequences) causes racial disparities?
- Correlation problem (1) :

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society
  - Suggests: we need a different empirical strategy to evaluate claims
  - What facet of institutionalized racism (or its consequences) causes racial disparities?
- Correlation problem (1) :
  - Regression models can estimate coefficients for immutable characteristics

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society
  - Suggests: we need a different empirical strategy to evaluate claims
  - What facet of institutionalized racism (or its consequences) causes racial disparities?
- Correlation problem (1) :
  - Regression models can estimate coefficients for immutable characteristics
  - But are necessarily imprecise: what do scholars have in mind in models?

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society
  - Suggests: we need a different empirical strategy to evaluate claims
  - What facet of institutionalized racism (or its consequences) causes racial disparities?
- Correlation problem (1) :
  - Regression models can estimate coefficients for immutable characteristics
  - But are necessarily imprecise: what do scholars have in mind in models?
- Design Principle:

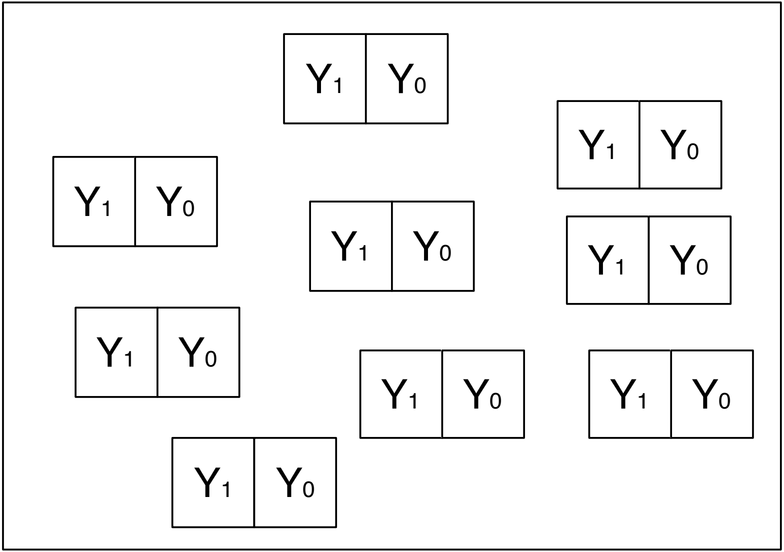
Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society
  - Suggests: we need a different empirical strategy to evaluate claims
  - What facet of institutionalized racism (or its consequences) causes racial disparities?
- Correlation problem (1) :
  - Regression models can estimate coefficients for immutable characteristics
  - But are necessarily imprecise: what do scholars have in mind in models?
- Design Principle:
  - Pretend you're God designing experiment

Caveats and Implications

- Does not dismiss claims of discrimination on immutable characteristics as legitimate
  - Pervasive effects of racism/sexism in society
  - Suggests: we need a different empirical strategy to evaluate claims
  - What facet of institutionalized racism (or its consequences) causes racial disparities?
- Correlation problem (1) :
  - Regression models can estimate coefficients for immutable characteristics
  - But are necessarily imprecise: what do scholars have in mind in models?
- Design Principle:
  - Pretend you're God designing experiment
  - If that experiment does not exist, be concerned

Back to the Neyman Urn Model



Average Treatment Effects

Average Treatment Effects

Move the goal posts:

Average Treatment Effects

Move the goal posts:
Focus on estimating Average Treatment Effect (ATE)

Average Treatment Effects

Move the goal posts:

Focus on estimating Average Treatment Effect (ATE)

Suppose we have N observations in population ($i = 1, \ldots, N$)

Average Treatment Effects

Move the goal posts:
Focus on estimating Average Treatment Effect (ATE)
Suppose we have N observations in population $(i = 1, \ldots, N)$

$$\text{ATE} = \frac{1}{N} \sum_{i=1}^{N} (Y_i(1) - Y_i(0))$$

Average Treatment Effects

Move the goal posts:
Focus on estimating Average Treatment Effect (ATE)
Suppose we have N observations in population ($i = 1, \ldots, N$)

$$\begin{aligned} \text{ATE} &= \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) \\ &= E[Y(1) - Y(0)] \text{ Average over population!!!} \end{aligned}$$

Average Treatment Effects

Move the goal posts:
Focus on estimating Average Treatment Effect (ATE)
Suppose we have N observations in population ($i = 1, \ldots, N$)

$$\begin{aligned} \text{ATE} &= \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) \\ &= E[Y(1) - Y(0)] \text{ Average over population!!!} \end{aligned}$$

- Population parameter

Average Treatment Effects

Move the goal posts:
Focus on estimating Average Treatment Effect (ATE)
Suppose we have N observations in population ($i = 1, \ldots, N$)

$$\begin{aligned} \text{ATE} &= \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0)) \\ &= E[Y(1) - Y(0)] \text{ Average over population!!!} \end{aligned}$$

- Population parameter
- It is fixed and unchanging

Estimating ATE under Random Assignment

Estimating ATE under Random Assignment

Estimator for ATE:

Estimating ATE under Random Assignment

Estimator for ATE:

$$\widehat{\text{ATE}} = \text{Average (Treated Units)} - \text{Average (Control Units)}$$

Estimating ATE under Random Assignment

Estimator for ATE:

$$\begin{aligned}\widehat{\text{ATE}} &= \text{Average (Treated Units)} - \text{Average (Control Units)} \\ &= \frac{\sum_{i=1}^N Y_i(1) T_i}{\sum_{i=1}^N T_i} - \frac{\sum_{i=1}^N Y_i(0)(1-T_i)}{\sum_{i=1}^N (1-T_i)}\end{aligned}$$

Estimating ATE under Random Assignment

Estimator for ATE:

$$
\begin{aligned}
\widehat{\text{ATE}} &= \text{Average (Treated Units)} - \text{Average (Control Units)} \\
&= \frac{\sum_{i=1}^{N} Y_i(1) T_i}{\sum_{i=1}^{N} T_i} - \frac{\sum_{i=1}^{N} Y_i(0)(1-T_i)}{\sum_{i=1}^{N}(1-T_i)} \\
&= \sum_{i=1}^{N} [\frac{Y_i(1) T_i}{n_t} - \frac{Y_i(0)(1-T_i)}{n_c}]
\end{aligned}
$$

Estimating ATE under Random Assignment

Estimator for ATE:

$$
\begin{aligned}
\widehat{\text{ATE}} &= \text{Average (Treated Units)} - \text{Average (Control Units)} \\
&= \frac{\sum_{i=1}^{N} Y_i(1) T_i}{\sum_{i=1}^{N} T_i} - \frac{\sum_{i=1}^{N} Y_i(0)(1 - T_i)}{\sum_{i=1}^{N}(1 - T_i)} \\
&= \sum_{i=1}^{N} [\frac{Y_i(1) T_i}{n_t} - \frac{Y_i(0)(1 - T_i)}{n_c}] \\
&= E[Y(1)|T = 1] - E[Y(0)|T = 0]
\end{aligned}
$$

Estimands

Because τ_i are unobservable, we shift what we are interested to:

Definition (Average Treatment Effect (ATE))

$$\tau_{ATE} = \text{Average of all treatment potential outcomes} - \text{Average of all control potential outcomes}$$

or

$$\tau_{ATE} = \frac{\sum_i^N Y_{1i}}{N} - \frac{\sum_i^N Y_{0i}}{N}$$

or

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

or

$$\tau_{ATE} = E[\tau_i]$$

Other Estimands

Definition (Average treatment effect on the treated (ATT))

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

Other Estimands

Definition (Average treatment effect on the treated (ATT))

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

Definition (Average treatment effect on the controls (ATC))

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

Other Estimands

Definition (Average treatment effect on the treated (ATT))

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

Definition (Average treatment effect on the controls (ATC))

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

Definition (Average treatment effects for subgroups)

$$\tau_{ATE(X)} = E[Y_{1i} - Y_{0i} | X_i = x]$$

or

$$\tau_{ATT(X)} = E[Y_{1i} - Y_{0i} | D_i = 1, X_i = x]$$

Average Treatment Effect

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATE?

Average Treatment Effect

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATE?

$E[Y_{1i} - Y_{0i}] = 1/4 \times (6 + -1 + 0 + 3) = 2$

Average Treatment Effect

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATE?

$E[Y_{1i} - Y_{0i}] = 1/4 \times (6 + -1 + 0 + 3) = 2$

Note: Average effect is positive, but τ_i are negative for some units!

Average Treatment Effect on the Treated

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATT and ATC?

Average Treatment Effect on the Treated

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATT and ATC?

$E[Y_{1i} - Y_{0i} | D_i = 1] = 1/2 \times (6 + -1) = 2.5$

Average Treatment Effect on the Treated

Imagine a study population with 4 units:

i	D_i	Y_{1i}	Y_{0i}	τ_i
1	1	10	4	6
2	1	1	2	-1
3	0	3	3	0
4	0	5	2	3

What is the ATT and ATC?

$$E[Y_{1i} - Y_{0i}|D_i = 1] = 1/2 \times (6 + -1) = 2.5$$

$$E[Y_{1i} - Y_{0i}|D_i = 0] = 1/2 \times (0 + 3) = 1.5$$

Naive Comparison: Difference in Means

Naive Comparison: Difference in Means

Comparisons between observed outcomes of treated and control units can often be misleading.

Naive Comparison: Difference in Means

Comparisons between observed outcomes of treated and control units can often be misleading.

- units which select treatment may not be like units which select control.

Naive Comparison: Difference in Means

Comparisons between observed outcomes of treated and control units can often be misleading.

- units which select treatment may not be like units which select control.
- i.e. selection into treatment is often associated with the potential outcomes

Naive Comparison: Difference in Means

Comparisons between <u>observed</u> outcomes of treated and control units can often be misleading.

- units which select treatment may not be like units which select control.
- i.e. selection into treatment is often associated with the potential outcomes
- this means we have violated the assumption of unconfoundness $(Y(1), Y(0)) \perp D$

Selection Bias

Selection Bias

Example: Church Attendance and Political Participation

Selection Bias

Example: Church Attendance and Political Participation

- Church goers likely to differ from non-Church goers on a range of background characteristics (e.g. civic duty)

Selection Bias

Example: Church Attendance and Political Participation

- Church goers likely to differ from non-Church goers on a range of background characteristics (e.g. civic duty)
- Given these differences, turnout for churchgoers could be higher than for non-churchgoers even if church had zero mobilizing effect

Selection Bias

Example: Church Attendance and Political Participation

- Church goers likely to differ from non-Church goers on a range of background characteristics (e.g. civic duty)
- Given these differences, turnout for churchgoers could be higher than for non-churchgoers even if church had zero mobilizing effect

Example: Gender Quotas and Redistribution Towards Women

Selection Bias

Example: Church Attendance and Political Participation

- Church goers likely to differ from non-Church goers on a range of background characteristics (e.g. civic duty)
- Given these differences, turnout for churchgoers could be higher than for non-churchgoers even if church had zero mobilizing effect

Example: Gender Quotas and Redistribution Towards Women

- Countries with gender quotas are likely countries where women are politically mobilized.

Selection Bias

Example: Church Attendance and Political Participation

- Church goers likely to differ from non-Church goers on a range of background characteristics (e.g. civic duty)
- Given these differences, turnout for churchgoers could be higher than for non-churchgoers even if church had zero mobilizing effect

Example: Gender Quotas and Redistribution Towards Women

- Countries with gender quotas are likely countries where women are politically mobilized.
- Given this difference, policies targeted towards women would be more common in quota countries even if these countries had not adopted quotas.

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill in, i.e. **estimate** missing potential outcomes.

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill in, i.e. **estimate** missing potential outcomes.
- In the causal inference literature, we typically make assumptions about the **assignment mechanism** to do so.

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill in, i.e. **estimate** missing potential outcomes.
- In the causal inference literature, we typically make assumptions about the **assignment mechanism** to do so.

Types of Assignment Mechanisms

- random assignment

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill in, i.e. **estimate** missing potential outcomes.
- In the causal inference literature, we typically make assumptions about the **assignment mechanism** to do so.

Types of Assignment Mechanisms

- random assignment
- selection on observables

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill in, i.e. **estimate** missing potential outcomes.
- In the causal inference literature, we typically make assumptions about the **assignment mechanism** to do so.

Types of Assignment Mechanisms

- random assignment
- selection on observables
- selection on unobservables

The Assignment Mechanism

- Since missing potential outcomes are unobservable we must make assumptions to fill in, i.e. **estimate** missing potential outcomes.
- In the causal inference literature, we typically make assumptions about the **assignment mechanism** to do so.

Types of Assignment Mechanisms

- random assignment
- selection on observables
- selection on unobservables

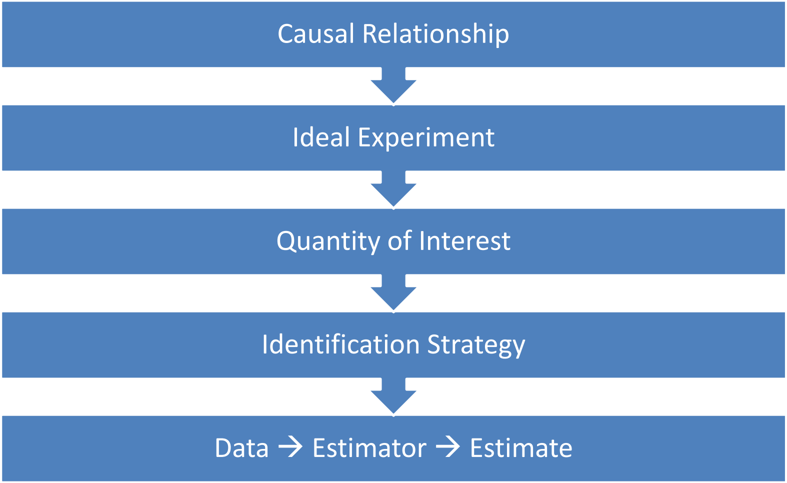
Most statistical models of causal inference attain identification of treatment effects by restricting the assignment mechanism in some way.

$\boxed{\text{No causation without manipulation?}}$

$\boxed{\text{No causation without manipulation?}}$

Always ask:
what is the experiment I would run if I had infinite resources and power?

Causal Inference Workflow



Summing Up: Neyman-Rubin causal model

Summing Up: Neyman-Rubin causal model

- Useful for studying the "effects of causes", less so for the "causes of effects".

Summing Up: Neyman-Rubin causal model

- Useful for studying the "effects of causes", less so for the "causes of effects".
- No assumption of homogeneity, allows for causal effects to vary unit by unit

Summing Up: Neyman-Rubin causal model

- Useful for studying the "effects of causes", less so for the "causes of effects".
- No assumption of homogeneity, allows for causal effects to vary unit by unit
  - No single "causal effect", thus the need to be precise about the target estimand.

Summing Up: Neyman-Rubin causal model

- Useful for studying the "effects of causes", less so for the "causes of effects".
- No assumption of homogeneity, allows for causal effects to vary unit by unit
  - No single "causal effect", thus the need to be precise about the target estimand.
- Distinguishes between observed outcomes and potential outcomes.

Summing Up: Neyman-Rubin causal model

- Useful for studying the "effects of causes", less so for the "causes of effects".
- No assumption of homogeneity, allows for causal effects to vary unit by unit
  - No single "causal effect", thus the need to be precise about the target estimand.
- Distinguishes between <u>observed</u> outcomes and <u>potential</u> outcomes.
- Causal inference is a missing data problem: we typically make assumptions about the assignment mechanism to go from descriptive inference to causal inference.

Summary: Observational Studies and Causal Inference

Experimental studies:

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Observational

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Observational

- Units (people, countries) control their treatment status

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
    - $E[Y(1)|T=1] \neq E[Y(1)|T=0]$, $E[Y(0)|T=0] \neq E[Y(0)|T=1]$

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T = 1] \neq E[Y(1)|T = 0]$, $E[Y(0)|T = 0] \neq E[Y(0)|T = 1]$
  - Observables: things we can see, measure, and use in our study

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T=1] \neq E[Y(1)|T=0]$, $E[Y(0)|T=0] \neq E[Y(0)|T=1]$
  - Observables: things we can see, measure, and use in our study
  - Unobservables: not observables (big problem)

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T=1] \neq E[Y(1)|T=0]$, $E[Y(0)|T=0] \neq E[Y(0)|T=1]$
  - Observables: things we can see, measure, and use in our study
  - Unobservables: not observables (big problem)
- Naive difference in means will be biased

Summary: Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T=1] \neq E[Y(1)|T=0]$, $E[Y(0)|T=0] \neq E[Y(0)|T=1]$
  - Observables: things we can see, measure, and use in our study
  - Unobservables: not observables (big problem)
- Naive difference in means will be biased
- Many, many, potential strategies for limiting bias

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)
  (1) can be relaxed by:

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)

  (1) can be relaxed by:
    - ⋆ Using a flexible nonlinear or nonparametric method

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)
  (1) can be relaxed by:
    - Using a flexible nonlinear or nonparametric method
    - "Preprocessing" data to make analysis robust to misspecification

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)

  (1) can be relaxed by:
    - ⋆ Using a flexible nonlinear or nonparametric method
    - ⋆ "Preprocessing" data to make analysis robust to misspecification

  (2) can be made plausible by:

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)
  (1) can be relaxed by:
    - Using a flexible nonlinear or nonparametric method
    - "Preprocessing" data to make analysis robust to misspecification
  (2) can be made plausible by:
    - Including carefully-selected control variables in the model

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)
  (1) can be relaxed by:
    - ⋆ Using a flexible nonlinear or nonparametric method
    - ⋆ "Preprocessing" data to make analysis robust to misspecification
  (2) can be made plausible by:
    - ⋆ Including carefully-selected control variables in the model
    - ⋆ Choosing a clever research design to rule out confounding

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)

  (1) can be relaxed by:
    - ⋆ Using a flexible nonlinear or nonparametric method
    - ⋆ "Preprocessing" data to make analysis robust to misspecification

  (2) can be made plausible by:
    - ⋆ Including carefully-selected control variables in the model
    - ⋆ Choosing a clever research design to rule out confounding

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)
  (1) can be relaxed by:
    - Using a flexible nonlinear or nonparametric method
    - "Preprocessing" data to make analysis robust to misspecification
  (2) can be made plausible by:
    - Including carefully-selected control variables in the model
    - Choosing a clever research design to rule out confounding
- We will discuss more in a few weeks.

Summary: Regression as a Causal Model

- Can regression be also used for causal inference?
- Answer: A very qualified yes
- For example, can we say that a one unit increase in inequality *causes* a 5.2 point increase in intensity?
- To interpret β as a causal effect of X on Y, we need very specific and often unrealistic assumptions:
  (1) $E[Y|X]$ is correctly specified as a linear function (linearity)
  (2) There are no other variables that affect both X and Y (exogeneity)
  (1) can be relaxed by:
    - Using a flexible nonlinear or nonparametric method
    - "Preprocessing" data to make analysis robust to misspecification
  (2) can be made plausible by:
    - Including carefully-selected control variables in the model
    - Choosing a clever research design to rule out confounding
- We will discuss more in a few weeks.
- For now while doing diagnostics, it is safest to treat β as a purely descriptive/predictive quantity

Graphical Models

Graphical Models

- A general framework for representing causal relationships based on directed acyclic graphs (DAG)

Graphical Models

- A general framework for representing causal relationships based on directed acyclic graphs (DAG)
- The work we discuss here comes out of developments by Judea Pearl and others

Graphical Models

- A general framework for representing causal relationships based on directed acyclic graphs (DAG)
- The work we discuss here comes out of developments by Judea Pearl and others
- Particularly useful for thinking through issues of identification.

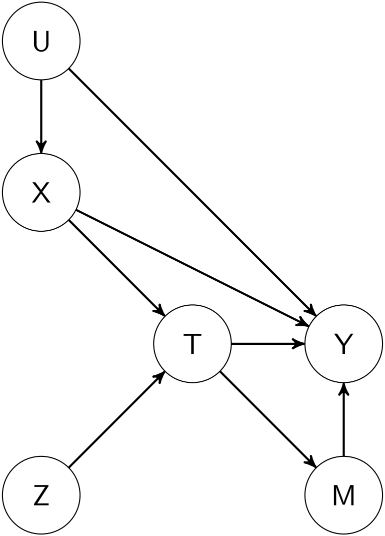
Graphical Models

- A general framework for representing causal relationships based on directed acyclic graphs (DAG)
- The work we discuss here comes out of developments by Judea Pearl and others
- Particularly useful for thinking through issues of identification.
- Provides a graphical representation of the models and a set of rules (do-calculus) for identifying the causal effect.

Graphical Models

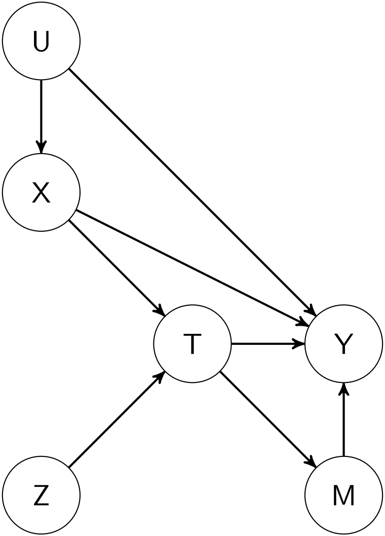
- A general framework for representing causal relationships based on directed acyclic graphs (DAG)
- The work we discuss here comes out of developments by Judea Pearl and others
- Particularly useful for thinking through issues of identification.
- Provides a graphical representation of the models and a set of rules (do-calculus) for identifying the causal effect.
- Nice software that takes the graph and returns an identification strategy: DAGitty at `http://dagitty.net`

Components of a DAG



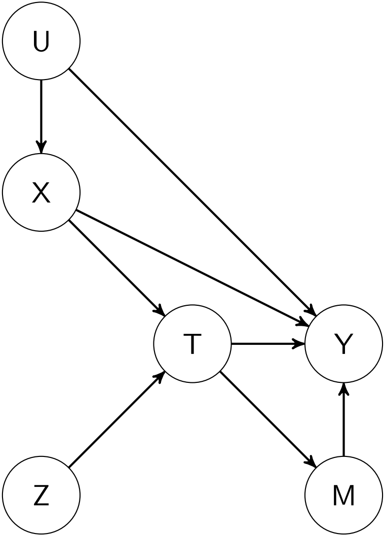
- nodes $\rightarrow$ variables
  (unobserved typically called U or V)

Components of a DAG



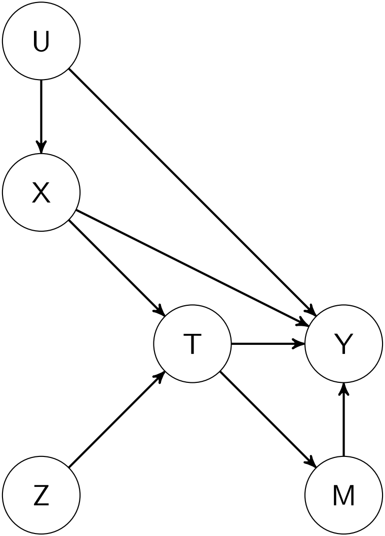
- nodes $\rightarrow$ variables
  (unobserved typically called U or V)
- (directed) arrows $\rightarrow$ causal effects

Components of a DAG



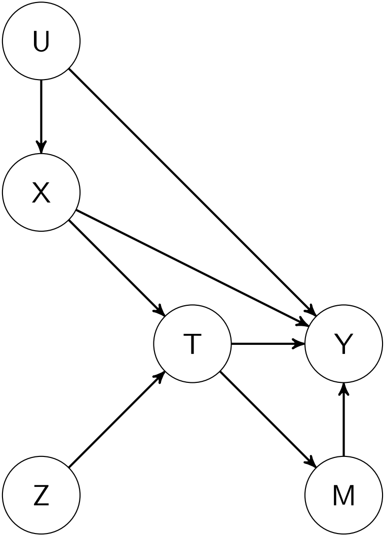
- nodes → variables
  (unobserved typically called U or V)
- (directed) arrows → causal effects
- absence of nodes → no common causes of any pair of variables

Components of a DAG



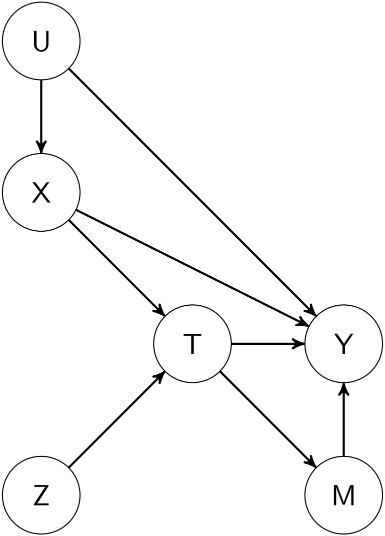
- nodes $\rightarrow$ variables
  (unobserved typically called U or V)
- (directed) arrows $\rightarrow$ causal effects
- absence of nodes $\rightarrow$ no common causes of any pair of variables
- absence of arrows $\rightarrow$ no causal effect

Components of a DAG



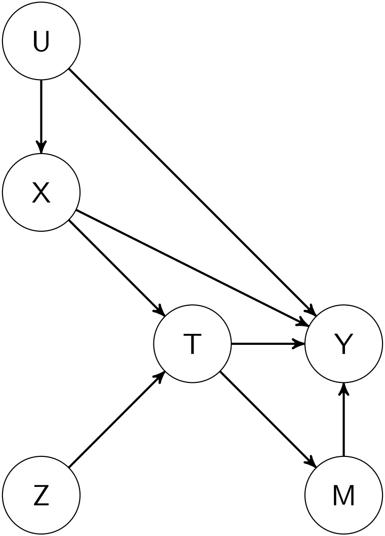
- nodes → variables
  (unobserved typically called U or V)
- (directed) arrows → causal effects
- absence of nodes → no common causes of any pair of variables
- absence of arrows → no causal effect
- positioning conveys no mathematical meaning but often is oriented left-to-right with causal ordering for readability.

Components of a DAG



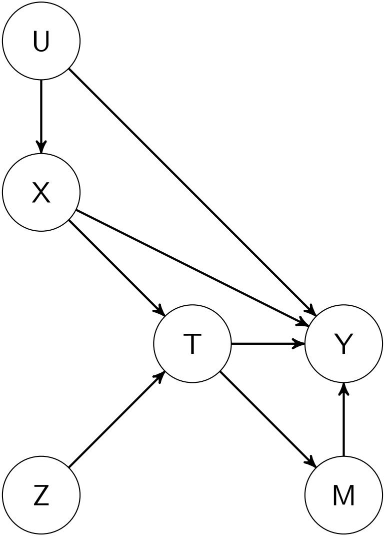
- nodes $\rightarrow$ variables
  (unobserved typically called U or V)
- (directed) arrows $\rightarrow$ causal effects
- absence of nodes $\rightarrow$ no common causes of any pair of variables
- absence of arrows $\rightarrow$ no causal effect
- positioning conveys no mathematical meaning but often is oriented left-to-right with causal ordering for readability.
- dashed lines are used in context dependent ways

Components of a DAG



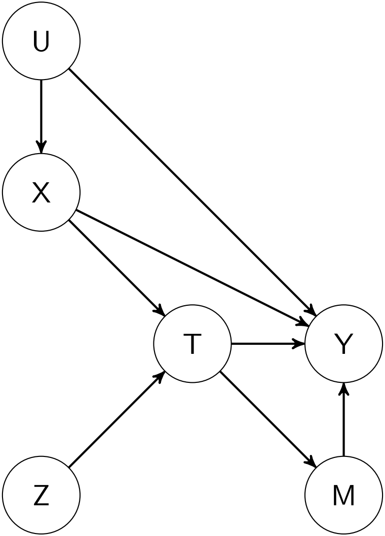
- nodes → variables (unobserved typically called U or V)
- (directed) arrows → causal effects
- absence of nodes → no common causes of any pair of variables
- absence of arrows → no causal effect
- positioning conveys no mathematical meaning but often is oriented left-to-right with causal ordering for readability.
- dashed lines are used in context dependent ways
- all relationships are non-parametric

Relationships in a DAG



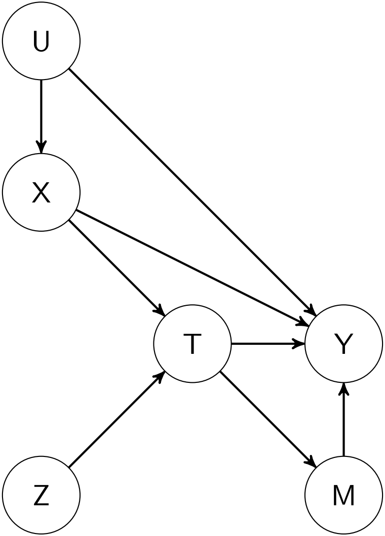
- Parents (Children): directly causing (caused by) a node

Relationships in a DAG



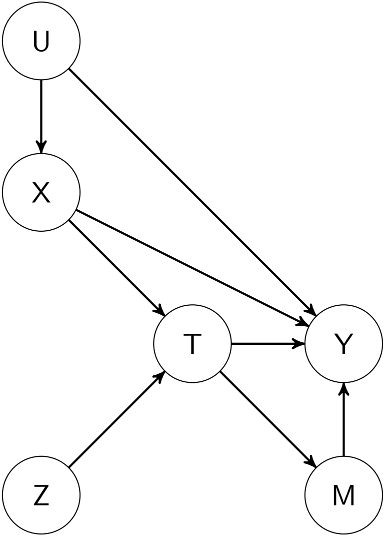
- Parents (Children): directly causing (caused by) a node
- Ancestors (Descendants): directly or indirectly causing (caused by) a node

Relationships in a DAG



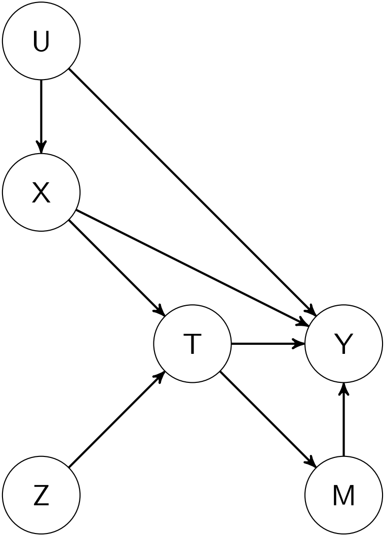
- Parents (Children): directly causing (caused by) a node
- Ancestors (Descendants): directly or indirectly causing (caused by) a node
- Path: a route that connects the variables (path is causal when all arrows point the same way)

Relationships in a DAG



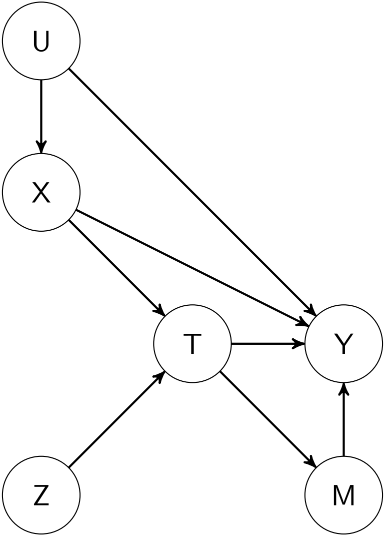
- Parents (Children): directly causing (caused by) a node
- Ancestors (Descendants): directly or indirectly causing (caused by) a node
- Path: a route that connects the variables (path is causal when all arrows point the same way)
- Acyclic implies that there are no cycles and a variable can't cause itself

Relationships in a DAG



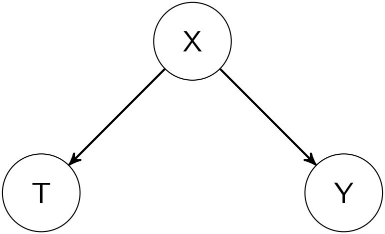
- Parents (Children): directly causing (caused by) a node
- Ancestors (Descendants): directly or indirectly causing (caused by) a node
- Path: a route that connects the variables (path is causal when all arrows point the same way)
- Acyclic implies that there are no cycles and a variable can't cause itself
- Causal Markov assumption: condition on its direct causes, a variable is independent of its non-descendents.

Relationships in a DAG



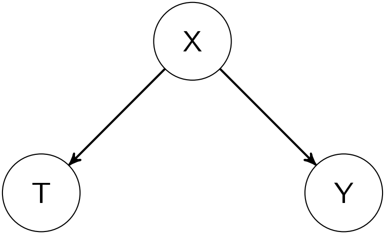
- Parents (Children): directly causing (caused by) a node
- Ancestors (Descendants): directly or indirectly causing (caused by) a node
- Path: a route that connects the variables (path is causal when all arrows point the same way)
- Acyclic implies that there are no cycles and a variable can't cause itself
- Causal Markov assumption: condition on its direct causes, a variable is independent of its non-descendents.
- We will talk in depth about two types of relationships: confounders and colliders

Confounders



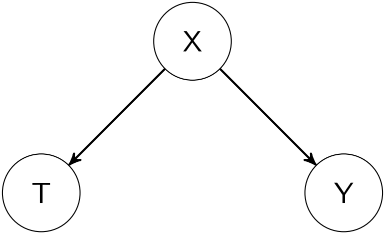
- X is a confounder (or common cause)

Confounders



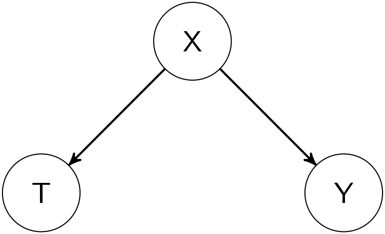
- X is a confounder (or common cause)
- Even without a causal effect or directed edge between T and Y they will have a marginal associational relationship

Confounders



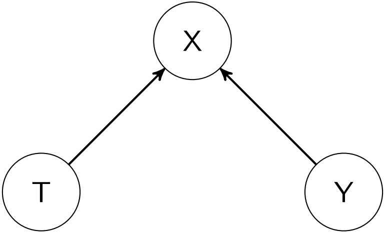
- X is a confounder (or common cause)
- Even without a causal effect or directed edge between T and Y they will have a marginal associational relationship
- Conditional on X, T and Y are unrelated in this graph.

Confounders



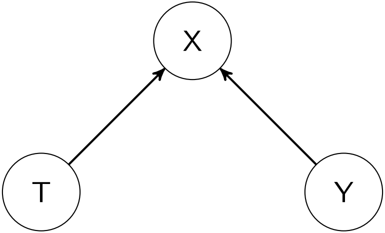
- X is a confounder (or common cause)
- Even without a causal effect or directed edge between T and Y they will have a marginal associational relationship
- Conditional on X, T and Y are unrelated in this graph.
- We can think of conditioning on a confounder as blocking the flow of association.

Colliders



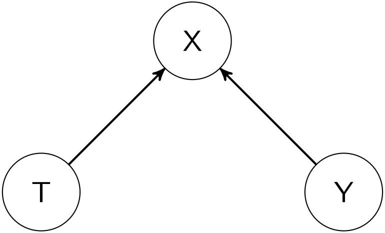
- X is now a collider because two arrows point into it

Colliders



- X is now a collider because two arrows point into it
- In this scenario T and Y are not marginally associated

Colliders



- X is now a collider because two arrows point into it
- In this scenario T and Y are not marginally associated
- If we control for X they become associated and create a connection between T and Y

Colliders are scary because you can induce dependence

ANNUAL REVIEWS **Further**

Click here for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable

Felix Elwert[1] and Christopher Winship[2]

[1]Department of Sociology, University of Wisconsin, Madison, Wisconsin 53706; email: elwert@wisc.edu

[2]Department of Sociology, Harvard University, Cambridge, Massachusetts 02138; email: cwinship@wjh.harvard.edu

Annu. Rev. Sociol. 2014. 40:31–53

First published online as a Review in Advance on June 2, 2014

The *Annual Review of Sociology* is online at soc.annualreviews.org

This article's doi: 10.1146/annurev-soc-071913-043455

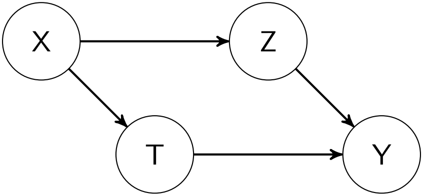
Keywords

causality, directed acyclic graphs, identification, confounding, selection

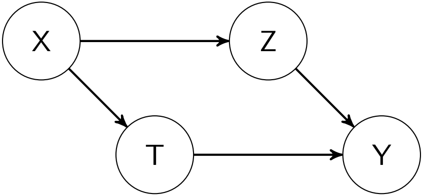
Abstract

Endogenous selection bias is a central problem for causal inference. Recognizing the problem, however, can be difficult in practice. This article introduces a purely graphical way of characterizing endogenous selection bias and of understanding its consequences (Hernán et al. 2004). We use causal graphs (direct acyclic graphs, or DAGs) to highlight that endogenous selection bias stems from conditioning (e.g., controlling, stratifying, or selecting) on a so-called collider variable, i.e., a variable that is itself caused by two other variables, one that is (or is associated with) the treatment and another that is (or is associated with) the outcome. Endogenous selection bias can result from direct conditioning on the outcome variable, a post-outcome variable, a post-treatment variable, and even a pre-treatment variable. We highlight the difference between endogenous selection bias, common-cause confounding, and overcontrol bias and discuss numerous examples from social stratification, cultural sociology, social network analysis, political sociology, social demography, and the sociology of education.

From Confounders to Back-Door Paths

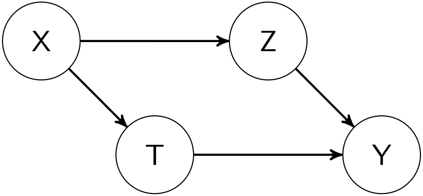


From Confounders to Back-Door Paths



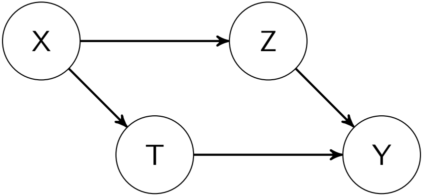
- Identify causal effect of T on Y by conditioning on X, Z or X and Z

From Confounders to Back-Door Paths



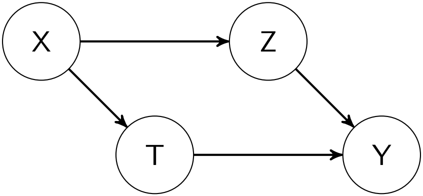
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path

From Confounders to Back-Door Paths



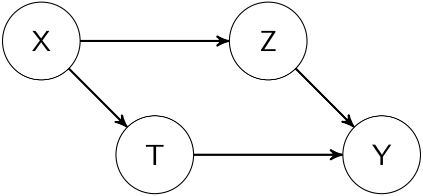
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path
- A back-door path is "a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable." (Morgan and Winship 2013)

From Confounders to Back-Door Paths



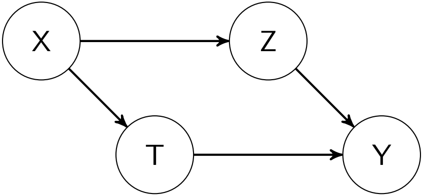
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path
- A back-door path is "a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable." (Morgan and Winship 2013)
- Two paths from T to Y here:
  1. $T \rightarrow Y$ (directed or causal path)
  2. $T \leftarrow X \rightarrow Z \rightarrow Y$ (back-door path)

From Confounders to Back-Door Paths



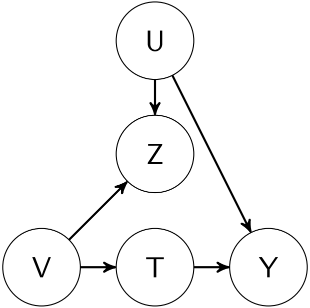
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path
- A back-door path is "a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable." (Morgan and Winship 2013)
- Two paths from T to Y here:
    1. $T \rightarrow Y$ (directed or causal path)
    2. $T \leftarrow X \rightarrow Z \rightarrow Y$ (back-door path)
- Observed marginal association between T and Y is a composite of these two paths and thus does not identify the causal effect of T on Y

From Confounders to Back-Door Paths



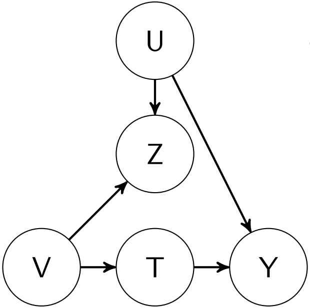
- Identify causal effect of T on Y by conditioning on X, Z or X and Z
- We can formalize this logic with the idea of a back-door path
- A back-door path is "a path between any causally ordered sequence of two variables that begins with a directed edge that points to the first variable." (Morgan and Winship 2013)
- Two paths from T to Y here:
  1. $T \rightarrow Y$ (directed or causal path)
  2. $T \leftarrow X \rightarrow Z \rightarrow Y$ (back-door path)
- Observed marginal association between T and Y is a composite of these two paths and thus does not identify the causal effect of T on Y
- We want to block the back-door path to leave only the causal effect

Colliders and Back-Door Paths



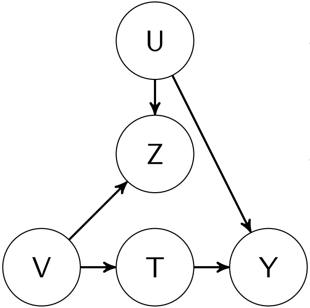
- Z is a collider and it lies along a back-door path from T to Y

Colliders and Back-Door Paths



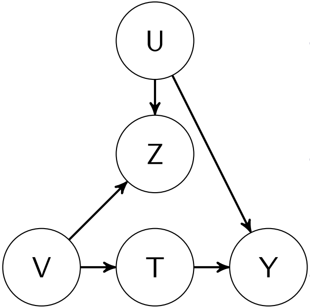
- Z is a collider and it lies along a back-door path from T to Y
- Conditioning on a collider on a back-door path does not help and in fact causes new associations

Colliders and Back-Door Paths



- Z is a collider and it lies along a back-door path from T to Y
- Conditioning on a collider on a back-door path does not help and in fact causes new associations
- Here we are fine unless we condition on Z which opens a path $T \leftarrow V \leftrightarrow U \rightarrow Y$ (this particular case is called M-bias)

Colliders and Back-Door Paths



- Z is a collider and it lies along a back-door path from T to Y
- Conditioning on a collider on a back-door path does not help and in fact causes new associations
- Here we are fine unless we condition on Z which opens a path $T \leftarrow V \leftrightarrow U \rightarrow Y$ (this particular case is called M-bias)
- So how do we know which back-door paths to block?

D-Separation

- Graphs provide us a way to think about conditional independence statements. Consider disjoint subsets of the vertices A, B and C

D-Separation

- Graphs provide us a way to think about conditional independence statements. Consider disjoint subsets of the vertices A, B and C
- A is D-separated from B by C if and only if C blocks every path from a vertex in A to a vertex in B

D-Separation

- Graphs provide us a way to think about conditional independence statements. Consider disjoint subsets of the vertices A, B and C
- A is D-separated from B by C if and only if C blocks every path from a vertex in A to a vertex in B
- A path p is said to be blocked by a set of vertices C if and only if at least one of the following conditions holds:

D-Separation

- Graphs provide us a way to think about conditional independence statements. Consider disjoint subsets of the vertices A, B and C
- A is D-separated from B by C if and only if C blocks every path from a vertex in A to a vertex in B
- A path p is said to be blocked by a set of vertices C if and only if at least one of the following conditions holds:
  1. p contains a chain structure $a \rightarrow c \rightarrow b$ or a fork structure $a \leftarrow c \rightarrow b$ where the node c is in the set C

D-Separation

- Graphs provide us a way to think about conditional independence statements. Consider disjoint subsets of the vertices A, B and C
- A is *D*-separated from B by C if and only if C blocks every path from a vertex in A to a vertex in B
- A path p is said to be blocked by a set of vertices C if and only if at least one of the following conditions holds:
  1. p contains a chain structure $a \rightarrow c \rightarrow b$ or a fork structure $a \leftarrow c \rightarrow b$ where the node c is in the set C
  2. p contains a collider structure $a \rightarrow y \leftarrow b$ where neither y nor its descendents are in C

D-Separation

- Graphs provide us a way to think about conditional independence statements. Consider disjoint subsets of the vertices A, B and C
- A is *D*-separated from B by C if and only if C blocks every path from a vertex in A to a vertex in B
- A path p is said to be blocked by a set of vertices C if and only if at least one of the following conditions holds:
    1. p contains a chain structure $a \rightarrow c \rightarrow b$ or a fork structure $a \leftarrow c \rightarrow b$ where the node c is in the set C
    2. p contains a collider structure $a \rightarrow y \leftarrow b$ where neither y nor its descendents are in C
- If A is not D-separated from B by C we say that A is *D*-connected to B by C

Backdoor Criterion

- Generally we want to know if we can nonparametrically identify the average effect of T on Y given a set of possible conditioning variables X

Backdoor Criterion

- Generally we want to know if we can nonparametrically identify the average effect of T on Y given a set of possible conditioning variables X
- Backdoor Criterion for X
  1. No node in X is a descendent of T
     (i.e. don't condition on post-treatment variables!)

Backdoor Criterion

- Generally we want to know if we can nonparametrically identify the average effect of T on Y given a set of possible conditioning variables X
- Backdoor Criterion for X
  1. No node in X is a descendent of T
     (i.e. don't condition on post-treatment variables!)
  2. X D-separates every path between T and Y that has an incoming arrow into T (backdoor path)

Backdoor Criterion

- Generally we want to know if we can nonparametrically identify the average effect of T on Y given a set of possible conditioning variables X
- Backdoor Criterion for X
  1. No node in X is a descendent of T
     (i.e. don't condition on post-treatment variables!)
  2. X D-separates every path between T and Y that has an incoming arrow into T (backdoor path)
- In essence, we are trying to block all non-causal paths, so we can estimate the causal path.

Backdoor Criterion

- Generally we want to know if we can nonparametrically identify the average effect of T on Y given a set of possible conditioning variables X
- Backdoor Criterion for X
  1. No node in X is a descendent of T
     (i.e. don't condition on post-treatment variables!)
  2. X D-separates every path between T and Y that has an incoming arrow into T (backdoor path)
- In essence, we are trying to block all non-causal paths, so we can estimate the causal path.
- Backdoor criterion is just one way to identify the effect: but its the most popular approach in the social sciences and what we are trying to do 99% of the time.

Backdoor Criterion

- Generally we want to know if we can nonparametrically identify the average effect of T on Y given a set of possible conditioning variables X
- Backdoor Criterion for X
  1. No node in X is a descendent of T (i.e. don't condition on post-treatment variables!)
  2. X D-separates every path between T and Y that has an incoming arrow into T (backdoor path)
- In essence, we are trying to block all non-causal paths, so we can estimate the causal path.
- Backdoor criterion is just one way to identify the effect: but its the most popular approach in the social sciences and what we are trying to do 99% of the time.
- See also Frontdoor Criterion in the social sciences in work by Glynn and Kashin

Thoughts on DAGs and Potential Outcomes

Thoughts on DAGs and Potential Outcomes

- Two very different languages for talking about and thinking about causal inferences

Thoughts on DAGs and Potential Outcomes

- Two very different languages for talking about and thinking about causal inferences
- Potential outcomes is very focused on thinking about the treatment assignment mechanism.

Thoughts on DAGs and Potential Outcomes

- Two very different languages for talking about and thinking about causal inferences
- Potential outcomes is very focused on thinking about the treatment assignment mechanism.
- Potential outcomes is also less of a "foreign language" for most statisticians, but in my experience lumps together a lot of identification assumptions in opaque ignorability conditions.

Thoughts on DAGs and Potential Outcomes

- Two very different languages for talking about and thinking about causal inferences
- Potential outcomes is very focused on thinking about the treatment assignment mechanism.
- Potential outcomes is also less of a "foreign language" for most statisticians, but in my experience lumps together a lot of identification assumptions in opaque ignorability conditions.
- Graphical Models with DAGs are very visually appealing but the operations on the graph can be challenging

Thoughts on DAGs and Potential Outcomes

- Two very different languages for talking about and thinking about causal inferences
- Potential outcomes is very focused on thinking about the treatment assignment mechanism.
- Potential outcomes is also less of a "foreign language" for most statisticians, but in my experience lumps together a lot of identification assumptions in opaque ignorability conditions.
- Graphical Models with DAGs are very visually appealing but the operations on the graph can be challenging
- DAGs very helpful for thinking through identification and the entire causal process

Thoughts on DAGs and Potential Outcomes

- Two very different languages for talking about and thinking about causal inferences
- Potential outcomes is very focused on thinking about the treatment assignment mechanism.
- Potential outcomes is also less of a "foreign language" for most statisticians, but in my experience lumps together a lot of identification assumptions in opaque ignorability conditions.
- Graphical Models with DAGs are very visually appealing but the operations on the graph can be challenging
- DAGs very helpful for thinking through identification and the entire causal process
- Note that both are about non-parametric identification and not estimation. This is good and bad.

Thoughts on DAGs and Potential Outcomes

- Two very different languages for talking about and thinking about causal inferences
- Potential outcomes is very focused on thinking about the treatment assignment mechanism.
- Potential outcomes is also less of a "foreign language" for most statisticians, but in my experience lumps together a lot of identification assumptions in opaque ignorability conditions.
- Graphical Models with DAGs are very visually appealing but the operations on the graph can be challenging
- DAGs very helpful for thinking through identification and the entire causal process
- Note that both are about non-parametric identification and not estimation. This is good and bad.
    - Good: provides a very general framework that applies in non-linear scenarios and interactions

Thoughts on DAGs and Potential Outcomes

- Two very different languages for talking about and thinking about causal inferences
- Potential outcomes is very focused on thinking about the treatment assignment mechanism.
- Potential outcomes is also less of a "foreign language" for most statisticians, but in my experience lumps together a lot of identification assumptions in opaque ignorability conditions.
- Graphical Models with DAGs are very visually appealing but the operations on the graph can be challenging
- DAGs very helpful for thinking through identification and the entire causal process
- Note that both are about non-parametric identification and not estimation. This is good and bad.
  - Good: provides a very general framework that applies in non-linear scenarios and interactions
  - Bad: identification results for identification only holds when variable is completely controlled for (which may be difficult!)

Next “Week” (Three Classes)

Next “Week” (Three Classes)

- Diagnostics

Next “Week” (Three Classes)

- Diagnostics
- Unusual and Influential Data $\rightarrow$ Robust Estimation (Day 1)

Next “Week” (Three Classes)

- Diagnostics
- Unusual and Influential Data $\rightarrow$ Robust Estimation (Day 1)
- Nonlinearity $\rightarrow$ Generalized Additive Models (Day 2)

Next "Week" (Three Classes)

- Diagnostics
- Unusual and Influential Data → Robust Estimation (Day 1)
- Nonlinearity → Generalized Additive Models (Day 2)
- Unusual Errors → Sandwich Standard Errors/Block Bootstrap (Day 3)

Next "Week" (Three Classes)

- Diagnostics
- Unusual and Influential Data → Robust Estimation (Day 1)
- Nonlinearity → Generalized Additive Models (Day 2)
- Unusual Errors → Sandwich Standard Errors/Block Bootstrap (Day 3)
- Reading:
  - Fox Chapters 11-13
  - Optional: Fox Chapter 19 Robust Regression
  - Optional: King and Roberts "How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It." *Political Analysis*, 2, 23: 159179.
  - Optional: Aronow and Miller Chapters 4.2-4.4 (Inference, Clustering, Nonlinearity)
  - Optional: Angrist and Pishke Chapter 8 (Nonstandard Standard Error Issues)

Fun with a Bundle of Sticks

Sen and Wasow (2016) "Race as a Bundle of Sticks: Designs that Estimate Effects of Seemingly Immutable Characteristics" *Annual Review of Political Science*.

No Causation Without Manipulation

No Causation Without Manipulation

- One of the difficulties that students have with causal inference is the need for manipulation or an ideal experiment.

No Causation Without Manipulation

- One of the difficulties that students have with causal inference is the need for manipulation or an ideal experiment.
- In many areas the key variables are immutable such as race or gender

No Causation Without Manipulation

- One of the difficulties that students have with causal inference is the need for manipulation or an ideal experiment.
- In many areas the key variables are immutable such as race or gender
- Sen and Wasow argue that we can improve our empirical work on this by seeing race/ethnicity as a composite variable or ‘a bundle of sticks’ which can be manipulated separately

The Trouble with Race As Treatment

The Trouble with Race As Treatment

There are three problems with race as a treatment in the causal inference sense

The Trouble with Race As Treatment

There are three problems with race as a treatment in the causal inference sense

1. Race cannot be manipulated

The Trouble with Race As Treatment

There are three problems with race as a treatment in the causal inference sense

1. Race cannot be manipulated
   - without the capacity to manipulate the question is arguably ill-posed and the estimand is unidentified

The Trouble with Race As Treatment

There are three problems with race as a treatment in the causal inference sense

1. Race cannot be manipulated
   - without the capacity to manipulate the question is arguably ill-posed and the estimand is unidentified
2. Everything else is post-treatment

The Trouble with Race As Treatment

There are three problems with race as a treatment in the causal inference sense

1. Race cannot be manipulated
   - without the capacity to manipulate the question is arguably ill-posed and the estimand is unidentified
2. Everything else is post-treatment
   - everything else comes after race which is perhaps unsatisfying

The Trouble with Race As Treatment

There are three problems with race as a treatment in the causal inference sense

1. Race cannot be manipulated
   - without the capacity to manipulate the question is arguably ill-posed and the estimand is unidentified
2. Everything else is post-treatment
   - everything else comes after race which is perhaps unsatisfying
   - this also presumes we are only interested in the total effect

The Trouble with Race As Treatment

There are three problems with race as a treatment in the causal inference sense

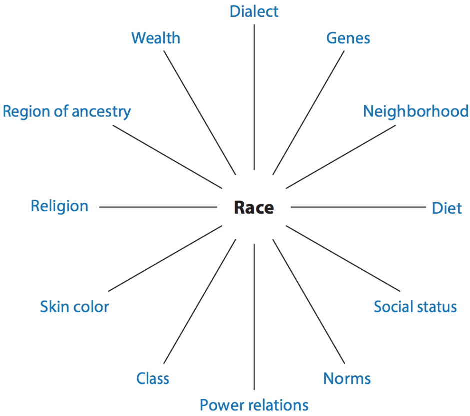
1. Race cannot be manipulated
   - without the capacity to manipulate the question is arguably ill-posed and the estimand is unidentified
2. Everything else is post-treatment
   - everything else comes after race which is perhaps unsatisfying
   - this also presumes we are only interested in the total effect
3. Race is unstable

The Trouble with Race As Treatment

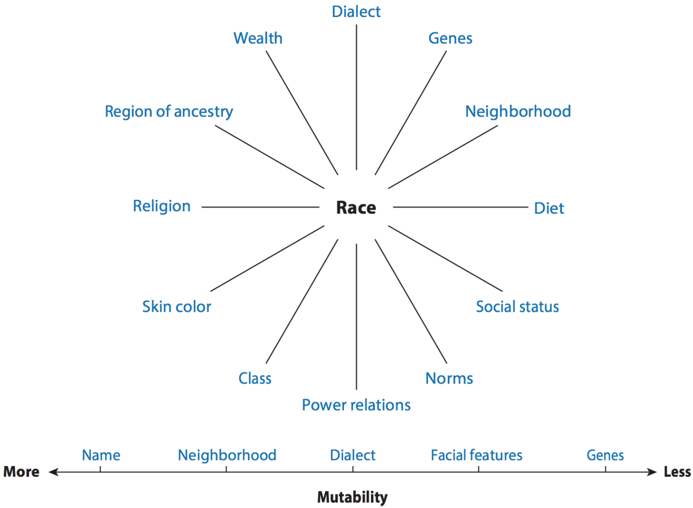
There are three problems with race as a treatment in the causal inference sense

1. Race cannot be manipulated
   - without the capacity to manipulate the question is arguably ill-posed and the estimand is unidentified
2. Everything else is post-treatment
   - everything else comes after race which is perhaps unsatisfying
   - this also presumes we are only interested in the total effect
3. Race is unstable
   - there is substantial variance across treatments which is a SUTVA violation

The Bundle of Sticks



The Bundle of Sticks



Design 1: Exposure Studies

- Approach

Design 1: Exposure Studies

- Approach
  a) "one or more elements of race is identified as a relevant cue"

Design 1: Exposure Studies

- Approach
    a) “one or more elements of race is identified as a relevant cue”
    b) “subjects are treated by exposure to the racial cue”

Design 1: Exposure Studies

- Approach
  a) "one or more elements of race is identified as a relevant cue"
  b) "subjects are treated by exposure to the racial cue"
  c) "unit of analysis is the individual or institution being exposed"

Design 1: Exposure Studies

- Approach
  a) "one or more elements of race is identified as a relevant cue"
  b) "subjects are treated by exposure to the racial cue"
  c) "unit of analysis is the individual or institution being exposed"
- Examples
  - Psychology (Steele 1997 on stereotype threat)

Design 1: Exposure Studies

- Approach
  a) "one or more elements of race is identified as a relevant cue"
  b) "subjects are treated by exposure to the racial cue"
  c) "unit of analysis is the individual or institution being exposed"
- Examples
  - Psychology (Steele 1997 on stereotype threat)
  - Audit/Correspondence Studies (Pager 2003, Bertrand and Mullainathan 2004)

Design 1: Exposure Studies

- Approach
  a) "one or more elements of race is identified as a relevant cue"
  b) "subjects are treated by exposure to the racial cue"
  c) "unit of analysis is the individual or institution being exposed"
- Examples
  - Psychology (Steele 1997 on stereotype threat)
  - Audit/Correspondence Studies (Pager 2003, Bertrand and Mullainathan 2004)
  - Survey Experiments with Racial Cues (Mendelberg 2001)

Design 1: Exposure Studies

- Approach
  a) "one or more elements of race is identified as a relevant cue"
  b) "subjects are treated by exposure to the racial cue"
  c) "unit of analysis is the individual or institution being exposed"
- Examples
  - Psychology (Steele 1997 on stereotype threat)
  - Audit/Correspondence Studies (Pager 2003, Bertrand and Mullainathan 2004)
  - Survey Experiments with Racial Cues (Mendelberg 2001)
  - Field Experiments with Racial Cues (Green 2004, Enos 2011)

Design 1: Exposure Studies

- Approach
  a) "one or more elements of race is identified as a relevant cue"
  b) "subjects are treated by exposure to the racial cue"
  c) "unit of analysis is the individual or institution being exposed"
- Examples
  - Psychology (Steele 1997 on stereotype threat)
  - Audit/Correspondence Studies (Pager 2003, Bertrand and Mullainathan 2004)
  - Survey Experiments with Racial Cues (Mendelberg 2001)
  - Field Experiments with Racial Cues (Green 2004, Enos 2011)
  - Observational Studies (Greiner and Rubin 2010, Wasow 2012)

Design 2: Within-Group Studies

- Approach: identify variation within the racial group along constitutive element.

Design 2: Within-Group Studies

- Approach: identify variation within the racial group along constitutive element.
- Example: Sharkey (2010) exploiting temporal variation in local homicides in Chicago to identify a significant neighborhood effect of proximity to violence on cognitive performance of African-American children

Concluding Thoughts

We can study race with causal inference, it just takes very careful design.

Table 2 Overview of exposure and within-group research designs

	Exposure	Within-Group
Unit	Individuals or institutions, potentially from any group	Members of a particular group
Typical treatment	Racial cue or signal (e.g., include distinctively ethnic names on a resume)	Constitutive element of the composite of race (e.g., address anxiety about social belonging in college)
Role of element of race	One "stick" is a proxy for the bundle (e.g., in a phone call with a landlord, dialect signals many traits associated with race)	One "stick" explains part of the bundle (e.g., Middle Passage might partly explain high rates of hypertension among African-Americans)
Examples	Correspondence and audit studies Implicit Association Tests	Experimental manipulation of a constitutive psychological dimension of race Within-race matching