SMI 606: Causality

Dr. Todd Hartman
Sheffield Methods Institute

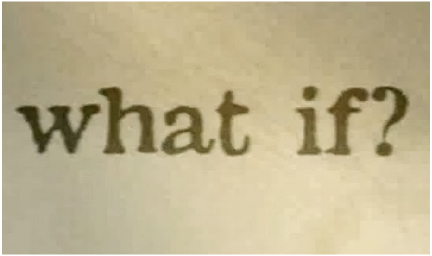
Roadmap: Week 2

- Lecture: Identifying causal effects
- Lab: Using R to test causality
  (SATE and SATT estimators)

Causal Effects and the Counterfactual

They key to demonstrating causality is to compare

- the **factual**, or what actually happened, with
- the **counterfactual**, or what would have happened if a key condition were different



Racial Discrimination in the Labour Market

Are employers less likely to interview minority candidates than whites?

An Example: The Resume Experiment

- Design: mail resumes to potential employers
- Treatment: resume with minority-sounding name
- Control: resume with white-sounding names

Resume Study: The Counterfactual

Causal Inference: Resume Study Main Results

TABLE 1—MEAN CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES

	Percent callback for White names	Percent callback for African-American names	Ratio	Percent difference (p-value)
Sample:				
All sent resumes	9.65	6.45	1.50	3.20
	[2,435]	[2,435]		(0.0000)
Chicago	8.06	5.40	1.49	2.66
	[1,352]	[1,352]		(0.0057)
Boston	11.63	7.76	1.50	4.05
	[1,083]	[1,083]		(0.0023)
Females	9.89	6.63	1.49	3.26
	[1,860]	[1,886]		(0.0003)
Females in administrative jobs	10.46	6.55	1.60	3.91
	[1,358]	[1,359]		(0.0003)
Females in sales jobs	8.37	6.83	1.22	1.54
	[502]	[527]		(0.3523)
Males	8.87	5.83	1.52	3.04
	[575]	[549]		(0.0513)

Resume Study: Callback Rates (Female Names)

White female			African-American female		
Name	Percent callback	Mother education	Name	Percent callback	Mother education
Emily	7.9	96.6	Aisha	2.2	77.2
Anne	8.3	93.1	Keisha	3.8	68.8
Jill	8.4	92.3	Tamika	5.5	61.5
Allison	9.5	95.7	Lakisha	5.5	55.6
Laurie	9.7	93.4	Tanisha	5.8	64.0
Sarah	9.8	97.9	Latoya	8.4	55.5
Meredith	10.2	81.8	Kenya	8.7	70.2
Carrie	13.1	80.7	Latonya	9.1	31.3
Kristen	13.1	93.4	Ebony	9.6	65.6

Resume Study: Callback Rates (Male Names)

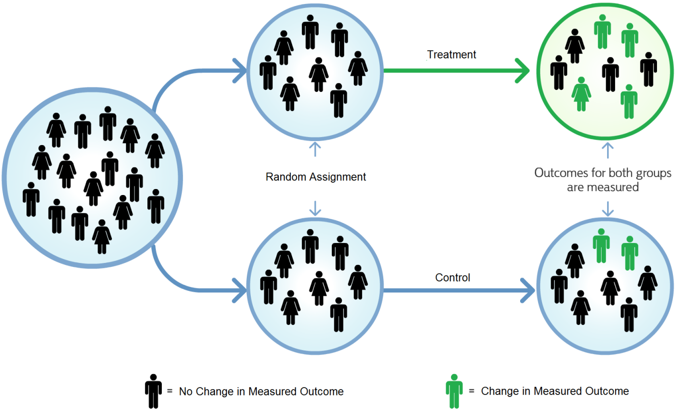
White male			African-American male		
Name	Percent callback	Mother education	Name	Percent callback	Mother education
Todd	5.9	87.7	Rasheed	3.0	77.3
Neil	6.6	85.7	Tremayne	4.3	—
Geoffrey	6.8	96.0	Kareem	4.7	67.4
Brett	6.8	93.9	Darnell	4.8	66.1
Brendan	7.7	96.7	Tyrone	5.3	64.0
Greg	7.8	88.3	Hakim	5.5	73.7
Matthew	9.0	93.1	Jamal	6.6	73.9
Jay	13.4	85.4	Leroy	9.4	53.3
Brad	15.9	90.5	Jermaine	9.6	57.5

Resume Study: Callback Rates (All Names)

White female			African-American female		
Name	Percent callback	Mother education	Name	Percent callback	Mother education
Emily	7.9	96.6	Aisha	2.2	77.2
Anne	8.3	93.1	Keisha	3.8	68.8
Jill	8.4	92.3	Tamika	5.5	61.5
Allison	9.5	95.7	Lakisha	5.5	55.6
Laurie	9.7	93.4	Tanisha	5.8	64.0
Sarah	9.8	97.9	Latoya	8.4	55.5
Meredith	10.2	81.8	Kenya	8.7	70.2
Carrie	13.1	80.7	Latonya	9.1	31.3
Kristen	13.1	93.4	Ebony	9.6	65.6
Average		91.7	Average		61.0
Overall		83.9	Overall		70.2
Correlation	−0.318	(p = 0.404)	Correlation	−0.383	(p = 0.309)

White male			African-American male		
Name	Percent callback	Mother education	Name	Percent callback	Mother education
Todd	5.9	87.7	Rasheed	3.0	77.3
Neil	6.6	85.7	Tremayne	4.3	—
Geoffrey	6.8	96.0	Kareem	4.7	67.4
Brett	6.8	93.9	Darnell	4.8	66.1
Brendan	7.7	96.7	Tyrone	5.3	64.0
Greg	7.8	88.3	Hakim	5.5	73.7
Matthew	9.0	93.1	Jamal	6.6	73.9
Jay	13.4	85.4	Leroy	9.4	53.3
Brad	15.9	90.5	Jermaine	9.6	57.5
Average		91.7	Average		66.7
Overall		83.5	Overall		68.9
Correlation	−0.0251	(p = 0.949)	Correlation	−0.595	(p = 0.120)

Randomized-Controlled Trials: The 'Gold' Standard



Sample Balancing

FAN2034560 [RF] © www.visualphotos.com

Experimental Data: Calculating Treatment Effects

Sample Average Treatment Effect

$$SATE = \frac{1}{n} \sum_{i=1}^{n} \underbrace{Y_i(1)}_{\text{Factual}} - \underbrace{Y_i(0)}_{\text{Counterfactual}}$$

The Importance of Random Assignment

- Randomization of treatment assignment means we can approximate SATE

Difference in Means Estimator

$$SATE = \frac{1}{n}\sum_{i=1}^{n} Y_i(1) - \frac{1}{n}\sum_{i=1}^{n} Y_i(0)$$

$$= \underbrace{\bar{Y}_i(1)}_{\text{Treatment}} - \underbrace{\bar{Y}_i(0)}_{\text{Control}}$$

Do Get Out The Vote Campaigns Work?

Social Pressure Experiment

- August 2006 Primary Statewide Election in Michigan
- Send postcards with different (randomly assigned) messages
  1. no message (control group)
  2. civic duty message
  3. "you are being studied" message (Hawthorne effect)
  4. neighborhood social pressure message

Name	Description
hhsize	household size of voter
messages	GOTV messages voter received (Civic, Control, Neighbours, Hawthorne)
sex	sex of voter (female or male)
yearofbirth	year of birth of voter
primary2004	whether a voter turned out in the 2004 Primary election (1=voted, 0=abstained)
primary2006	whether a voter turned out in the 2006 Primary election (1=voted, 0=abstained)

Field Experiments: 'Civic Duty' Treatment

30426-2 XXX

For more information: (517) 351-1975
email: etov@grebner.com
Practical Political Consulting
P. O. Box 6249
East Lansing, MI 48826

PRSRT STD
U.S. Postage
PAID
Lansing, MI
Permit # 444

ECRLOT **C002
THE JONES FAMILY
9999 WILLIAMS RD
FLINT MI 48507

Dear Registered Voter:

DO YOUR CIVIC DUTY AND VOTE!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

The whole point of democracy is that citizens are active participants in government; that we have a voice in government. Your voice starts with your vote. On August 8, remember your rights and responsibilities as a citizen. Remember to vote.

DO YOUR CIVIC DUTY — VOTE!

Field Experiments: 'Hawthorne' Treatment

30424-1

For more information: (517) 351-1975
email: etov@grebner.com
Practical Political Consulting
P. O. Box 6249
East Lansing, MI 48826

PRSRT STD
U.S. Postage
PAID
Lansing, MI
Permit # 444

ECRLOT **C001
THE SMITH FAMILY
9999 PARK LANE
FLINT MI 48507

Dear Registered Voter:

YOU ARE BEING STUDIED!

Why do so many people fail to vote? We've been talking about this problem for years, but it only seems to get worse.

This year, we're trying to figure out why people do or do not vote. We'll be studying voter turnout in the August 8 primary election.

Our analysis will be based on public records, so you will not be contacted again or disturbed in any way. Anything we learn about your voting or not voting will remain confidential and will not be disclosed to anyone else.

DO YOUR CIVIC DUTY — VOTE!

Named after the factory where researchers observed increase in productivity because subjects knew they were being observed.

Field Experiments: 'Neighbour' Treatment

Dear Registered Voter:

WHAT IF YOUR NEIGHBORS KNEW WHETHER YOU VOTED?

Why do so many people fail to vote? We've been talking about the problem for years, but it only seems to get worse. This year, we're taking a new approach. We're sending this mailing to you and your neighbors to publicize who does and does not vote.

The chart shows the names of some of your neighbors, showing which have voted in the past. After the August 8 election, we intend to mail an updated chart. You and your neighbors will all know who voted and who did not.

DO YOUR CIVIC DUTY — VOTE!

MAPLE DR	Aug 04	Nov 04	Aug 06
9995 JOSEPH JAMES SMITH	Voted	Voted	
9995 JENNIFER KAY SMITH		Voted	
9997 RICHARD B JACKSON		Voted	
9999 KATHY MARIE JACKSON		Voted	
9999 BRIAN JOSEPH JACKSON		Voted	
9991 JENNIFER KAY THOMPSON		Voted	
9991 BOB R THOMPSON		Voted	
9993 BILL S SMITH			
9989 WILLIAM LUKE CASPER		Voted	
9989 JENNIFER SUE CASPER		Voted	
9987 MARIA S JOHNSON	Voted	Voted	
9987 TOM JACK JOHNSON	Voted	Voted	
9987 RICHARD TOM JOHNSON		Voted	
9985 ROSEMARY S SUE		Voted	
9985 KATHRYN L SUE		Voted	
9985 HOWARD BEN SUE		Voted	
9983 NATHAN CHAD BERG		Voted	
9983 CARRIE ANN BERG		Voted	
9981 EARL JOEL SMITH			
9979 DEBORAH KAY WAYNE		Voted	
9979 JOEL R WAYNE		Voted	

Difference in Means Estimator (SATE)

```r
## Find means by treatment condition
tapply(social$primary2006, social$messages, mean)
```

```
## Civic Duty    Control  Hawthorne  Neighbors
##  0.3145377  0.2966383  0.3223746  0.3779482
```

```r
## Calculate SATE
mean(social$primary2006[social$messages == "Civic Duty"])
- mean(social$primary2006[social$messages == "Control"])
```

```
## [1] 0.01789934
```

```r
mean(social$primary2006[social$messages == "Hawthorne"]) -
 mean(social$primary2006[social$messages == "Control"])
```

```
## [1] 0.02573631
```

```r
mean(social$primary2006[social$messages == "Neighbors"]) -
 mean(social$primary2006[social$messages == "Control"])
```

```
## [1] 0.08130991
```

Validity

- Internal validity: Extent to which causal assumptions are satisfied

- External validity: Extent to which conclusions can be generalized

Observational Studies: Effects of Basque Terrorism

Basque Terrorism

- Basque region hit with terrorism in early 1970s
  - From 1973 to late 1990s, ETA killed almost 800 people
  - Goal: Independent Basque state
  - Financing: Kidnappings, extortion, robberies
  - Activity localized to Basque area

- What is the economic impact of terrorism?
  - Basque was the 3rd richest region in Spain at onset
  - Dropped to the 6th position by late 1990s
- Counterfactual: How would Basque economy have fared in the absence of the terrorism?

Applying 3 Identification Strategies

- Basque region experiences terrorism
- Other Spanish regions (control) did not

- 3 Identification strategies:
  1. Compare Basque and others after 1973 (Cross-section comparison)
  2. Compare Basque before and after 1973 (Before-and-after)
  3. Compare others before and after 1973 and subtract the difference from Basque's difference (Difference-in-differences)

- Each approach makes different assumptions

Load Basque Data

```r
basque <- read.csv("data/basque.csv")
summary(basque)

##                  region           year           gdpcap
##  Andalucia        : 43   Min.     :1955   Min.     : 1.24
##  Aragon           : 43   1st Qu.:1965   1st Qu.: 3.69
##  Baleares         : 43   Median :1976   Median : 5.28
##  Basque Country: 43   Mean     :1976   Mean     : 5.40
##  Canarias         : 43   3rd Qu.:1987   3rd Qu.: 6.87
##  Cantabria        : 43   Max.     :1997   Max.     :12.35
##  (Other)          :473
```

- region: 17 regions including Basque Country
- year: 1955 – 1997
- gdpcap: real GDP per capita (in 1986 USD, thousands)

Subset Basque Data into 4 Groups

```
## Basque before terrorism
treatedBefore <- subset(basque, (year < 1973) &
                              (region == "Basque Country"))
## Basque after terrorism
treatedAfter <- subset(basque, (year >= 1973) &
                              (region == "Basque Country"))
## others before terrorism
controlBefore <- subset(basque, (year < 1973) &
                              (region != "Basque Country"))
## others after terrorism
controlAfter <- subset(basque, (year >= 1973) &
                              (region != "Basque Country"))
```

1. Cross-section comparison

```r
mean(treatedAfter$gdpcap) - mean(controlAfter$gdpcap)

## [1] 1.13
```

3 Identification Strategies in Action

1. Cross-section comparison

```
mean(treatedAfter$gdpcap) - mean(controlAfter$gdpcap)

## [1] 1.13
```

2. Before-and-after design

```
mean(treatedAfter$gdpcap) - mean(treatedBefore$gdpcap)

## [1] 2.68
```

1. Cross-section comparison

```r
mean(treatedAfter$gdpcap) - mean(controlAfter$gdpcap)

## [1] 1.13
```

2. Before-and-after design

```r
mean(treatedAfter$gdpcap) - mean(treatedBefore$gdpcap)

## [1] 2.68
```

3. Difference-in-Differences design

```r
treatDiff <- mean(treatedAfter$gdpcap) -
    mean(treatedBefore$gdpcap)
controlDiff <- mean(controlAfter$gdpcap) -
    mean(controlBefore$gdpcap)
treatDiff - controlDiff

## [1] -0.483
```

Estimating "Treatment" Effects with Observational Data

- Sample Average Treatment Effect *for the Treated* (SATT)

Difference-in-Differences Estimator

$$SATT = \underbrace{(\bar{Y}^{after}_{treated} - \bar{Y}^{before}_{treated})}_{\text{difference for treatment group}} - \underbrace{(\bar{Y}^{after}_{control} - \bar{Y}^{before}_{control})}_{\text{difference for control group}}$$

Wrapping Up Causality: Do You Get This?