Gov 50: 3. Causality

Matthew Blackwell

Harvard University

Fall 2018

1/ Today's agenda

Where are we?

- What you’ve been doing:

Where are we?

- What you've been doing:
  - DataCamp assignments

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)

Where are we?

- What you've been doing:
    - DataCamp assignments
    - Reading on Perusall (great job!)
    - Asking questions on Canvas

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)
  - Asking questions on Canvas
  - Getting set up with RStudio Cloud

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)
  - Asking questions on Canvas
  - Getting set up with RStudio Cloud
  - Looking at Google Calendar for class

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)
  - Asking questions on Canvas
  - Getting set up with RStudio Cloud
  - Looking at Google Calendar for class
- Sections start this week:

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)
  - Asking questions on Canvas
  - Getting set up with RStudio Cloud
  - Looking at Google Calendar for class
- Sections start this week:
  - Friday, 1:30-2:30pm

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)
  - Asking questions on Canvas
  - Getting set up with RStudio Cloud
  - Looking at Google Calendar for class
- Sections start this week:
  - Friday, 1:30-2:30pm
  - Friday, 3-4pm

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)
  - Asking questions on Canvas
  - Getting set up with RStudio Cloud
  - Looking at Google Calendar for class
- Sections start this week:
  - Friday, 1:30-2:30pm
  - Friday, 3-4pm
  - Monday, 6-7pm (taped for Extension School)

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)
  - Asking questions on Canvas
  - Getting set up with RStudio Cloud
  - Looking at Google Calendar for class
- Sections start this week:
  - Friday, 1:30-2:30pm
  - Friday, 3-4pm
  - Monday, 6-7pm (taped for Extension School)
  - Possibly more options to come

Where are we?

- What you've been doing:
  - DataCamp assignments
  - Reading on Perusall (great job!)
  - Asking questions on Canvas
  - Getting set up with RStudio Cloud
  - Looking at Google Calendar for class
- Sections start this week:
  - Friday, 1:30-2:30pm
  - Friday, 3-4pm
  - Monday, 6-7pm (taped for Extension School)
  - Possibly more options to come
- First HW goes out today, due Thursday 9/20

Where are we going?

1. Criminal record experiment

Where are we going?

1. Criminal record experiment
   - More practice working with data in R

Where are we going?

1. Criminal record experiment
   - More practice working with data in R
2. Causality

Where are we going?

1. Criminal record experiment
   - More practice working with data in R
2. Causality
   - Counterfactuals

Where are we going?

1. Criminal record experiment
   - More practice working with data in R
2. Causality
   - Counterfactuals
   - Fundamental problem of causal inference

2/ Data

Criminal record experiment

- Does having a criminal record affect job prospects?

Criminal record experiment

- Does having a criminal record affect job prospects?
- Experimental setting:

Criminal record experiment

- Does having a criminal record affect job prospects?
- Experimental setting:
  - Randomly assign 4 hired “confederates” (2 white, 2 black) to apply to different jobs in Milwaukee.

Criminal record experiment

- Does having a criminal record affect job prospects?
- Experimental setting:
  - Randomly assign 4 hired "confederates" (2 white, 2 black) to apply to different jobs in Milwaukee.
  - Men were matched on physical appearance, self-presentation, age, etc.

Criminal record experiment

- Does having a criminal record affect job prospects?
- Experimental setting:
  - Randomly assign 4 hired "confederates" (2 white, 2 black) to apply to different jobs in Milwaukee.
  - Men were matched on physical appearance, self-presentation, age, etc.
  - Confederates would alternate indicating they had a criminal record.

Criminal record experiment

- Does having a criminal record affect job prospects?
- Experimental setting:
  - Randomly assign 4 hired "confederates" (2 white, 2 black) to apply to different jobs in Milwaukee.
  - Men were matched on physical appearance, self-presentation, age, etc.
  - Confederates would alternate indicating they had a criminal record.
- Outcome of interest: receiving a callback from a potential employer.

Data

- Data file: `criminalrecord.csv`

Name	Description
`jobid`	Job ID number
`callback`	1 if tester received a callback, 0 if the tester did not receive a callback.
`black`	1 if the tester is black, 0 if the tester is white.
`crimrec`	1 if the tester has a criminal record, 0 if the tester does not.
`interact`	1 if tester interacted with employer during the job application, 0 if tester does not interact with employer.
`city`	1 is job is located in the city center, 0 if job is located in the suburbs.
`distance`	Job's average distance to downtown.

- Load the data and see what it looks like:

- Load the data and see what it looks like:

```
audit <- read.csv("data/criminalrecord.csv")

dim(audit)
```

- Load the data and see what it looks like:

```
audit <- read.csv("data/criminalrecord.csv")

dim(audit)
```

```
## [1] 696   7
```

- Load the data and see what it looks like:

```
audit <- read.csv("data/criminalrecord.csv")

dim(audit)
```

```
## [1] 696   7
```

```
head(audit)
```

- Load the data and see what it looks like:

```
audit <- read.csv("data/criminalrecord.csv")

dim(audit)
```

```
## [1] 696   7
```

```
head(audit)
```

```
##   jobid callback black crimrec interact city distance
## 1   108        1     0       1        1    0       15
## 2   113        0     0       0        1    0       20
## 3   101        1     0       0        0    0       15
## 4    64        1     0       0        0    1        7
## 5    33        0     0       1        0    1        5
## 6    73        0     0       1        0    1       10
```

What is our data?

- **Units**: objects being studied.

What is our data?

- **Units**: objects being studied.
  - Usually the rows of the data set.

What is our data?

- **Units**: objects being studied.
  - Usually the rows of the data set.
  - Survey respondents, countries, counties, members of congress, etc.

What is our data?

- **Units**: objects being studied.
  - Usually the rows of the data set.
  - Survey respondents, countries, counties, members of congress, etc.
  - Today's data: job applications.

What is our data?

- **Units**: objects being studied.
  - Usually the rows of the data set.
  - Survey respondents, countries, counties, members of congress, etc.
  - Today's data: job applications.
- **Variables**: measurements that can vary across units.

What is our data?

- **Units**: objects being studied.
  - Usually the rows of the data set.
  - Survey respondents, countries, counties, members of congress, etc.
  - Today's data: job applications.
- **Variables**: measurements that can vary across units.
  - Usually the columns of a data set.

What is our data?

- **Units**: objects being studied.
  - Usually the rows of the data set.
  - Survey respondents, countries, counties, members of congress, etc.
  - Today's data: job applications.
- **Variables**: measurements that can vary across units.
  - Usually the columns of a data set.
  - Political preferences, income, participation in an experiment.

Types of variables

- **Discrete** variable: each unit falls into one of several categories

Types of variables

- **Discrete** variable: each unit falls into one of several categories
  - Got a callback (1) or not (0)

Types of variables

- **Discrete** variable: each unit falls into one of several categories
  - Got a callback (1) or not (0)
  - Number of children: 0, 1, 2...

Types of variables

- **Discrete** variable: each unit falls into one of several categories
    - Got a callback (1) or not (0)
    - Number of children: 0, 1, 2...
    - Support for policy on a 1-5 scale

Types of variables

- **Discrete** variable: each unit falls into one of several categories
  - Got a callback (1) or not (0)
  - Number of children: 0, 1, 2...
  - Support for policy on a 1-5 scale
  - Differences between values are fixed.

Types of variables

- **Discrete** variable: each unit falls into one of several categories
  - Got a callback (1) or not (0)
  - Number of children: 0, 1, 2...
  - Support for policy on a 1-5 scale
  - Differences between values are fixed.
- **Continuous** variable:

Types of variables

- **Discrete** variable: each unit falls into one of several categories
  - Got a callback (1) or not (0)
  - Number of children: 0, 1, 2...
  - Support for policy on a 1-5 scale
  - Differences between values are fixed.
- **Continuous** variable:
  - Age, income

Types of variables

- **Discrete** variable: each unit falls into one of several categories
  - Got a callback (1) or not (0)
  - Number of children: 0, 1, 2...
  - Support for policy on a 1-5 scale
  - Differences between values are fixed.
- **Continuous** variable:
  - Age, income
  - Differences between values aren't fixed and can be arbitrarily small.

Logicals

- Often useful to ask R logical questions about values and vectors.

Logicals

- Often useful to ask R logical questions about values and vectors.

```
5 > 10
```

Logicals

- Often useful to ask R logical questions about values and vectors.

```r
5 > 10
```

```
## [1] FALSE
```

Logicals

- Often useful to ask R logical questions about values and vectors.

```r
5 > 10
```

```
## [1] FALSE
```

```r
5 >= 5
```

Logicals

- Often useful to ask R logical questions about values and vectors.

```r
5 > 10
```

```
## [1] FALSE
```

```r
5 >= 5
```

```
## [1] TRUE
```

Logicals

- Often useful to ask R logical questions about values and vectors.

```
5 > 10
```

```
## [1] FALSE
```

```
5 >= 5
```

```
## [1] TRUE
```

```
5 == 5
```

Logicals

- Often useful to ask R logical questions about values and vectors.

```
5 > 10
```

```
## [1] FALSE
```

```
5 >= 5
```

```
## [1] TRUE
```

```
5 == 5
```

```
## [1] TRUE
```

- We can apply logicals to vectors to get a vector of TRUE/FALSE entries:

- We can apply logicals to vectors to get a vector of TRUE/FALSE entries:

```
head(audit$crimrec)
```

- We can apply logicals to vectors to get a vector of TRUE/FALSE entries:

```
head(audit$crimrec)
```

```
## [1] 1 0 0 0 1 1
```

- We can apply logicals to vectors to get a vector of TRUE/FALSE entries:

```r
head(audit$crimrec)
```

```
## [1] 1 0 0 0 1 1
```

```r
head(audit$crimrec == 1)
```

- We can apply logicals to vectors to get a vector of TRUE/FALSE entries:

```
head(audit$crimrec)
```

```
## [1] 1 0 0 0 1 1
```

```
head(audit$crimrec == 1)
```

```
## [1]  TRUE FALSE FALSE FALSE  TRUE  TRUE
```

- We can then use this logical vector to subset the data to only certain rows:

- We can then use this logical vector to subset the data to only certain rows:

```
criminalrecs <- audit[audit$crimrec == 1, ]

head(criminalrecs[, c("jobid", "callback", "black", "crimrec")])
```

- We can then use this logical vector to subset the data to only certain rows:

```r
criminalrecs <- audit[audit$crimrec == 1, ]

head(criminalrecs[, c("jobid", "callback", "black", "crimrec")])
```

```
##    jobid callback black crimrec
## 1    108        1     0       1
## 5     33        0     0       1
## 6     73        0     0       1
## 7      4        0     0       1
## 8    125        1     0       1
## 10   110        0     0       1
```

Calculate means in subsets

- We want to know the average callback rate for the criminal and non-criminal applications.

Calculate means in subsets

- We want to know the average callback rate for the criminal and non-criminal applications.
- First, we use the subset we already created and then make a second subset for the other condition:

Calculate means in subsets

- We want to know the average callback rate for the criminal and non-criminal applications.
- First, we use the subset we already created and then make a second subset for the other condition:

```
mean(criminalrecs$callback)
```

Calculate means in subsets

- We want to know the average callback rate for the criminal and non-criminal applications.
- First, we use the subset we already created and then make a second subset for the other condition:

```
mean(criminalrecs$callback)
```

```
## [1] 0.101
```

Calculate means in subsets

- We want to know the average callback rate for the criminal and non-criminal applications.
- First, we use the subset we already created and then make a second subset for the other condition:

```
mean(criminalrecs$callback)
```

```
## [1] 0.101
```

```
noncriminalrecs <- audit[audit$crimrec == 0, ]
mean(noncriminalrecs$callback)
```

Calculate means in subsets

- We want to know the average callback rate for the criminal and non-criminal applications.
- First, we use the subset we already created and then make a second subset for the other condition:

```
mean(criminalrecs$callback)
```

```
## [1] 0.101
```

```
noncriminalrecs <- audit[audit$crimrec == 0, ]
mean(noncriminalrecs$callback)
```

```
## [1] 0.226
```

Calculate means in subsets

- We want to know the average callback rate for the criminal and non-criminal applications.
- First, we use the subset we already created and then make a second subset for the other condition:

```
mean(criminalrecs$callback)
```

```
## [1] 0.101
```

```
noncriminalrecs <- audit[audit$crimrec == 0, ]
mean(noncriminalrecs$callback)
```

```
## [1] 0.226
```

- Much higher callback rate in the non-criminal-record group!

3/ Causality

Causal questions

- Does the minimum wage increase the unemployment rate?

Causal questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased

Causal questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?

Causal questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?
- Does race affect one's job prospect?

Causal questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?
- Does race affect one's job prospect?
  - Jamal applied for a job but did not get it

Causal questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?
- Does race affect one's job prospect?
  - Jamal applied for a job but did not get it
  - Would Jamal have gotten a job if he were white?

Causal questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?
- Does race affect one's job prospect?
  - Jamal applied for a job but did not get it
  - Would Jamal have gotten a job if he were white?

- Comparison between factual and counterfactual

Causal questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?
- Does race affect one's job prospect?
  - Jamal applied for a job but did not get it
  - Would Jamal have gotten a job if he were white?

- Comparison between factual and counterfactual
- **Fundamental problem of causal inference**: We must infer counterfactual outcomes

Causal questions

- Does the minimum wage increase the unemployment rate?
  - Unemployment rate went up after the minimum wage increased
  - Would the unemployment rate have gone up, had the minimum wage increase not occurred?
- Does race affect one's job prospect?
  - Jamal applied for a job but did not get it
  - Would Jamal have gotten a job if he were white?

- Comparison between factual and counterfactual
- **Fundamental problem of causal inference**: We must infer counterfactual outcomes
- No causation without manipulation: **immutable characteristics**

Counterfactual

- Seeing the fundamental problem of causal inference in a movie: Sliding Doors (1998) https://www.youtube.com/watch?v=BvUbv4iwbDs

A tale of two applications

```
audit[4:5, c("jobid", "callback", "crimrec")]
```

```
##   jobid callback crimrec
## 4    64        1       0
## 5    33        0       1
```

- Did employer 33 not callback the applicant **because** they had a criminal record?

Notation

- **Unit** (indexed by i): job application for employer

Notation

- **Unit** (indexed by i): job application for employer
- **Treatment variable** (causal variable of interest) T_i: criminal record or not

Notation

- **Unit** (indexed by i): job application for employer
- **Treatment variable** (causal variable of interest) T_i: criminal record or not
- **Treatment group** (treated units): applications with criminal record

Notation

- **Unit** (indexed by i): job application for employer
- **Treatment variable** (causal variable of interest) T_i: criminal record or not
- **Treatment group** (treated units): applications with criminal record
- **Control group** (untreated units): applications without criminal record

Notation

- **Unit** (indexed by i): job application for employer
- **Treatment variable** (causal variable of interest) T_i: criminal record or not
- **Treatment group** (treated units): applications with criminal record
- **Control group** (untreated units): applications without criminal record
- **Outcome variable** (response variable) Y_i: callback

Notation

- **Unit** (indexed by i): job application for employer
- **Treatment variable** (causal variable of interest) T_i: criminal record or not
- **Treatment group** (treated units): applications with criminal record
- **Control group** (untreated units): applications without criminal record
- **Outcome variable** (response variable) Y_i: callback

	T_i (ex-felon)	Y_i (callback)
Ex-felon applicant	1	0
Non-ex-felon applicant	0	1

Causal effects & counterfactuals

- What does “T_i causes Y_i” mean? $\rightsquigarrow$ **counterfactuals**, “what if”

Causal effects & counterfactuals

- What does "T_i causes Y_i" mean? $\rightsquigarrow$ **counterfactuals**, "what if"
- Would the employer for application i have called back if the applicant didn't have a criminal record?

Causal effects & counterfactuals

- What does "T_i causes Y_i" mean? $\rightsquigarrow$ **counterfactuals**, "what if"
- Would the employer for application i have called back if the applicant didn't have a criminal record?
- Two **potential outcomes**:

Causal effects & counterfactuals

- What does "T_i causes Y_i" mean? $\rightsquigarrow$ **counterfactuals**, "what if"
- Would the employer for application i have called back if the applicant didn't have a criminal record?
- Two **potential outcomes**:
  - $Y_i(1)$: would employer i callback if applicant is an ex-felon?

Causal effects & counterfactuals

- What does "T_i causes Y_i" mean? $\rightsquigarrow$ **counterfactuals**, "what if"
- Would the employer for application i have called back if the applicant didn't have a criminal record?
- Two **potential outcomes**:
  - $Y_i(1)$: would employer i callback if applicant is an ex-felon?
  - $Y_i(0)$: would employer i callback if applicant was **not** an ex-felon?

Causal effects & counterfactuals

- What does "T_i causes Y_i" mean? $\rightsquigarrow$ **counterfactuals**, "what if"
- Would the employer for application i have called back if the applicant didn't have a criminal record?
- Two **potential outcomes**:
  - $Y_i(1)$: would employer i callback if applicant is an ex-felon?
  - $Y_i(0)$: would employer i callback if applicant was **not** an ex-felon?
- **Causal effect**: $Y_i(1) - Y_i(0)$

Causal effects & counterfactuals

- What does "T_i causes Y_i" mean? $\rightsquigarrow$ **counterfactuals**, "what if"
- Would the employer for application i have called back if the applicant didn't have a criminal record?
- Two **potential outcomes**:
  - $Y_i(1)$: would employer i callback if applicant is an ex-felon?
  - $Y_i(0)$: would employer i callback if applicant was **not** an ex-felon?
- **Causal effect**: $Y_i(1) - Y_i(0)$
- **Fundamental problem of causal inference**: only one of the two potential outcomes is observable.

Potential outcomes

	T_i (ex-felon)	Y_i (callback)	$Y_i(1)$	$Y_i(0)$
Ex-felon applicant	1	0	0	???
Non-ex-felon applicant	0	1	???	1

Potential outcomes

	T_i (ex-felon)	Y_i (callback)	$Y_i(1)$	$Y_i(0)$
Ex-felon applicant	1	0	0	???
Non-ex-felon applicant	0	1	???	1

- **Association is not causation**

Potential outcomes

	T_i (ex-felon)	Y_i (callback)	$Y_i(1)$	$Y_i(0)$
Ex-felon applicant	1	0	0	???
Non-ex-felon applicant	0	1	???	1

- **Association is not causation**
- Need to infer the missing counterfactuals!

How can we figure out counterfactuals?

- Find a similar unit! $\rightsquigarrow$ **matching** (Mill's method of difference)

How can we figure out counterfactuals?

- Find a similar unit! $\rightsquigarrow$ **matching** (Mill's method of difference)
- Did applicant fail to get a job offer because of his criminal record?

How can we figure out counterfactuals?

- Find a similar unit! $\rightsquigarrow$ **matching** (Mill's method of difference)
- Did applicant fail to get a job offer because of his criminal record?
  - $\rightsquigarrow$ find a non-ex-felon who is just like ex-felon applicant.

How can we figure out counterfactuals?

- Find a similar unit! $\rightsquigarrow$ **matching** (Mill's method of difference)
- Did applicant fail to get a job offer because of his criminal record?
  - $\rightsquigarrow$ find a non-ex-felon who is just like ex-felon applicant.
- NJ increased the minimum wage

How can we figure out counterfactuals?

- Find a similar unit! $\rightsquigarrow$ **matching** (Mill's method of difference)
- Did applicant fail to get a job offer because of his criminal record?
  - $\rightsquigarrow$ find a non-ex-felon who is just like ex-felon applicant.
- NJ increased the minimum wage
  - $\rightsquigarrow$ find a state similar to NJ that didn't increase minimum wage.

How can we figure out counterfactuals?

- Find a similar unit! $\rightsquigarrow$ **matching** (Mill's method of difference)
- Did applicant fail to get a job offer because of his criminal record?
  - $\rightsquigarrow$ find a non-ex-felon who is just like ex-felon applicant.
- NJ increased the minimum wage
  - $\rightsquigarrow$ find a state similar to NJ that didn't increase minimum wage.
- The problem: we cannot match on everything

How can we figure out counterfactuals?

- Find a similar unit! $\rightsquigarrow$ **matching** (Mill's method of difference)
- Did applicant fail to get a job offer because of his criminal record?
  - $\rightsquigarrow$ find a non-ex-felon who is just like ex-felon applicant.
- NJ increased the minimum wage
  - $\rightsquigarrow$ find a state similar to NJ that didn't increase minimum wage.

- The problem: we cannot match on everything
- Unobserved **confounders**: variables associated with treatment and outcome $\rightsquigarrow$ **selection bias**

4/ Assignment

Where we are going

- Next time:

Where we are going

- Next time:
  - One way to get around confounding: randomization.

Where we are going

- Next time:
  - One way to get around confounding: randomization.
- Your to-do list:

Where we are going

- Next time:
  - One way to get around confounding: randomization.
- Your to-do list:
  - DataCamp Assignment 1 by tonight.

Where we are going

- Next time:
  - One way to get around confounding: randomization.
- Your to-do list:
  - DataCamp Assignment 1 by tonight.
  - Log into Perusall and rstudio.cloud

Where we are going

- Next time:
  - One way to get around confounding: randomization.
- Your to-do list:
  - DataCamp Assignment 1 by tonight.
  - Log into Perusall and rstudio.cloud
  - Read Imai 2.1-2.4 if you haven't

Where we are going

- Next time:
  - One way to get around confounding: randomization.
- Your to-do list:
  - DataCamp Assignment 1 by tonight.
  - Log into Perusall and rstudio.cloud
  - Read Imai 2.1-2.4 if you haven't
- Problem Set 1:

Where we are going

- Next time:
  - One way to get around confounding: randomization.
- Your to-do list:
  - DataCamp Assignment 1 by tonight.
  - Log into Perusall and rstudio.cloud
  - Read Imai 2.1-2.4 if you haven't
- Problem Set 1:
  - Will go out today.

Where we are going

- Next time:
  - One way to get around confounding: randomization.
- Your to-do list:
  - DataCamp Assignment 1 by tonight.
  - Log into Perusall and rstudio.cloud
  - Read Imai 2.1-2.4 if you haven't
- Problem Set 1:
  - Will go out today.
  - Analysis of a randomized field experiment in Boston by a Gov faculty member (Ryan Enos).

Where we are going

- Next time:
  - One way to get around confounding: randomization.
- Your to-do list:
  - DataCamp Assignment 1 by tonight.
  - Log into Perusall and rstudio.cloud
  - Read Imai 2.1-2.4 if you haven't
- Problem Set 1:
  - Will go out today.
  - Analysis of a randomized field experiment in Boston by a Gov faculty member (Ryan Enos).
  - You'll be able to copy a project for HW 1 on rstudio.cloud that will have templates, data, etc.