Causal Inference: Course Recap

Cornell STSCI / INFO / ILRST 3900
Fall 2025
causal3900.github.io

Dec 4 2025

Logistics

- Quiz 6 today
- Final Project Due Dec 8 at 11:59
- PSET 5 and 6 grades will be posted this week

Learning goals for the course

As a result of participating in this course, students will be able to

- define counterfactuals as the outcomes of hypothetical interventions
- identify counterfactuals by causal assumptions presented in graphs
- estimate counterfactual outcomes by pairing those assumptions with statistical evidence


EXTRAVERGINE EXTRAVERGINE
olio
olio
olio

Potential outcomes

$$Y_i^a$$

the outcome Y
of person i
if given treatment $A = a$

Fundamental problem of causal inference

Holland 1986

Descriptive evidence

average lifespan − average lifespan

Causal claim

average lifespan − average lifespan

Causal inference is a **missing data** problem

	Outcome under Mediterranean diet	Outcome under standard diet
Person 1	lifespan	missing
Person 2	missing	lifespan
Person 3	lifespan	missing
Person 4	missing	lifespan
Person 5	lifespan	missing
Person 6	lifespan	missing
Person 7	missing	lifespan
Person 8	lifespan	missing

Outcome under Mediterranean diet	Outcome under standard diet
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan
lifespan	lifespan

Estimating unobserved potential outcomes

Estimating unobserved potential outcomes

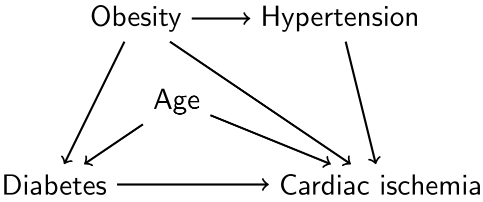
Exchangeability:

$$\underbrace{Y^{a=1}, Y^{a=0}}_{\text{potential outcomes}} \perp\!\!\!\perp \underbrace{A}_{\text{observed treatment}}$$

allows us to estimate counterfactual outcomes to observed outcomes when paired with consistency

$$\underbrace{\mathsf{E}(Y \mid A = 1)}_{\textit{Observed}} = \underbrace{\mathsf{E}(Y^1 \mid A = 1)}_{\textit{Counterfactual}} = \underbrace{\mathsf{E}(Y^1 \mid A = 0)}_{\textit{Counterfactual}}$$

Causal Graphs



Causal Graphs

$$\begin{tikzcd}
 & \text{Obesity} \arrow[r] \arrow[ddl] \arrow[ddr] & \text{Hypertension} \arrow[dd] \\
 & \text{Age} \arrow[dl] \arrow[dr] & \\
\text{Diabetes} \arrow[rr] & & \text{Cardiac ischemia}
\end{tikzcd}$$

- Open and blocked paths
- Checking for conditional exchangeability
- Selection bias can induce bias when conditioning on wrong variable

Causal identification by the backdoor criterion

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

Backdoor path starts with an edge pointing in to A and ends at Y

A set of variables satisfies the backdoor criterion if

1. Blocks all backdoor paths
2. Does not contain any descendant of A

Sufficient adjustment sets satisfy the backdoor criterion!

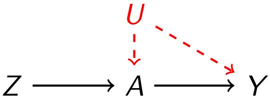
Estimation

If conditional exchangeability holds given $\vec{L}$,
then we need an estimator that statistically adjusts for $\vec{L}$

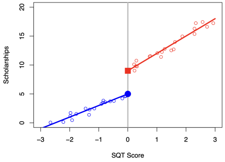
- Parametric G Formula
- Inverse probability weighting
- Matching

Identification without exchangeability

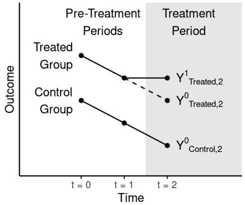
Instrumental Variables



Regression Discontinuity



Difference in Difference



Synthetic Control

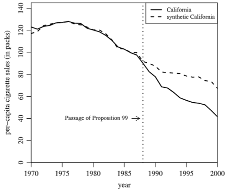
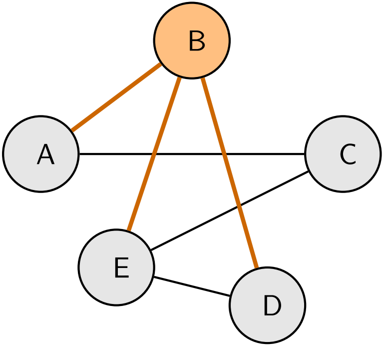


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Causal inference with interference



- Public health, social networks, marketing, etc.

Tied to concrete problems

Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination

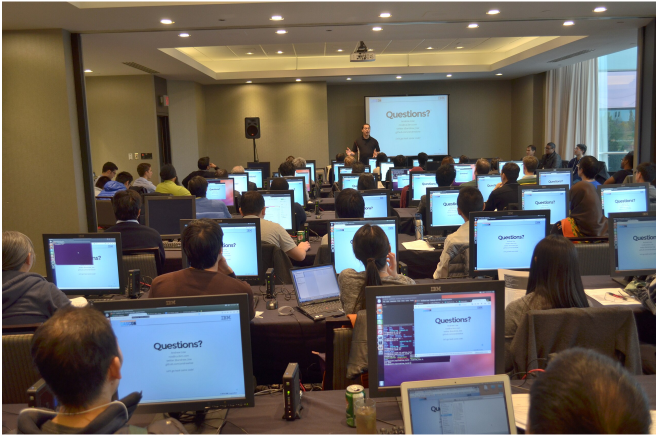
By MARIANNE BERTRAND AND SENDHIL MULLAINATHAN*

We study race in the labor market by sending fictitious resumes to help-wanted ads in Boston and Chicago newspapers. To manipulate perceived race, resumes are randomly assigned African-American- or White-sounding names. White names receive 50 percent more callbacks for interviews. Callbacks are also more responsive to resume quality for White names than for African-American ones. The racial gap is uniform across occupation, industry, and employer size. We also find little evidence that employers are inferring social class from the names. Differential treatment by race still appears to still be prominent in the U.S. labor market. (JEL J71, J64).

Every measure of economic success reveals significant racial inequality in the U.S. labor market. Compared to Whites, African-Americans are twice as likely to be unemployed and earn nearly 25 percent less when they are employed (Council of Economic Advisers, 1998). This inequality has sparked a debate as to whether employers treat members of different races differentially. When faced with observ- dates, employers might favor the African-American one.[1] Data limitations make it difficult to empirically test these views. Since researchers possess far less data than employers do, White and African-American workers that appear similar to researchers may look very different to employers. So any racial difference in labor market outcomes could just as easily be attributed to differences that are observable to

Experiments and Exchangeability on race and employment

Tied to concrete problems



Matching to estimate causal effect of job training program on income[1]

[1] By Raysonho @ Open Grid Scheduler / Grid Engine - Own work, CC0, https://commons.wikimedia.org/w/index.php?curid=36698205

Tied to concrete problems

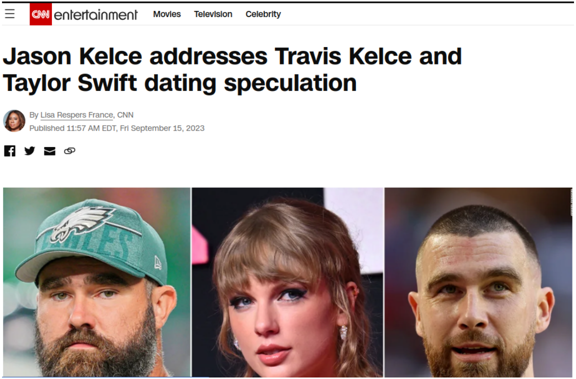
Instrumental variables to estimate causal effect of military service on income [1]

[1] https://commons.wikimedia.org/wiki/File:1969$_{d}raft_{l}ottery_{p}hoto.jpg$

Tied to concrete problems

Difference in difference to estimate the causal effect of raising minimum wage

Tied to concrete problems


entertainment Movies Television Celebrity

Jason Kelce addresses Travis Kelce and Taylor Swift dating speculation

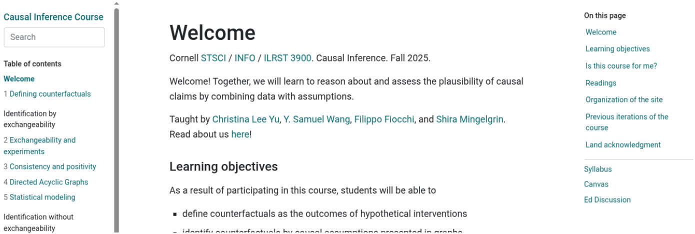
By Lisa Respers France, CNN
Published 11:57 AM EDT, Fri September 15, 2023

Synthetic control to estimate Taylor Swift's effect on Travis Kelce [1]

[1] https://commons.wikimedia.org/wiki/File:$1969_{d}raft_{l}ottery_{p}hoto.jpg$

Course structure

- Concepts introduced in lecture
- Hands-on practice in discussion (started coding!)
- Reinforced with problem sets
- Project to independently apply what you learned



Your thoughts

- What could we do to make this course better?
- What is your favorite thing you learned?
- What parts do you anticipate being most useful for your future work?

Evaluations

We want to hear from you!

We encourage **specific examples** for your TAs

- Recitation or discussion
  - Comments on the recitation or discussion section (include day and time of section or TA name)
  - If you would like to nominate a TA from this course for a teaching award, please identify the TA and explain briefly why.

Drop us a line!

In the future, if you are using any of the material from class, we'd love to know!

Learning goals for the course

As a result of participating in this course, students will be able to

- define counterfactuals as the outcomes of hypothetical interventions
- identify counterfactuals by causal assumptions presented in graphs
- estimate counterfactual outcomes by pairing those assumptions with statistical evidence