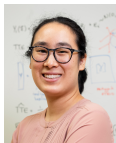# Causal Inference

STSCI / INFO / ILRST 3900

26 Aug 2025

Welcome to Causal Inference!

Christina Yu

Sam Wang

Filippo Fiocchi

Shira Mingelgrin

Why causal inference?

- ▶ Name, year, and major
- ▶ What was the highlight of your summer?
- ▶ What are you hoping to get out of the course?

Why causal inference?

Most of statistics is about

- ▶ Describing the current state of the world
- ▶ Making predictions when observing a system
- ▶ "What **is** the world like?"

Why causal inference?

Most of statistics is about
- ▶ Describing the current state of the world
- ▶ Making predictions when observing a system
- ▶ "What **is** the world like?"

Causal inference provides tools to
- ▶ Understand causal mechanisms
- ▶ Claims about what would have happened under an intervention
- ▶ "What **could** the world be like?"
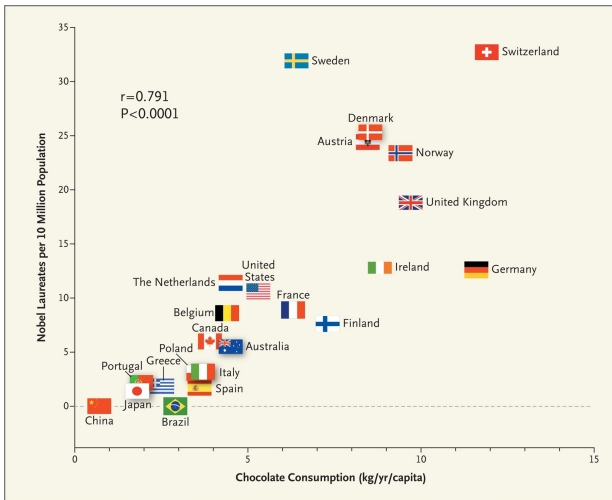
Causality
Reasoning about interventions

Figure: Chocolate vs Nobel Prizes from (Messerli, 2012, NEJM)

Examples of causal questions

▶ Card and Krueger (2000) ask if increasing the minimum wage results in a causes the employment rate to increase

Examples of causal questions

- ▶ Card and Krueger (2000) ask if increasing the minimum wage results in a causes the employment rate to increase
  - ▶ **Observe:** We may observe that the employment rate is higher/lower in states with a higher minimum wage
  - ▶ **Intervene:** Would increasing the minimum wage cause the employment rate to be higher/lower than it would have been if minimum wage remained the same?

Examples of causal questions

► Kearney and Levine (2015) ask if the reality TV show "16 and Pregnant" caused a decrease in US teen pregnancy rates

Examples of causal questions

- Kearney and Levine (2015) ask if the reality TV show "16 and Pregnant" caused a decrease in US teen pregnancy rates
  - **Observe:** We may observe that the teen pregnancy rate is lower after the TV show aired compared to before the show aired
  - **Intervene:** Did airing the show cause the pregnancy rate to decrease more than it would have if the show did not air?

Examples of causal questions

▶ Yam and Lopez (2019) ask if NFL teams would win more games if they tried to convert 4th downs more often

Examples of causal questions

- Yam and Lopez (2019) ask if NFL teams would win more games if they tried to convert 4th downs more often
  - **Observe:** We may observe that teams which attempt 4th down conversions more often end up winning more/less games
  - **Intervene:** Would deciding to go for fourth down more increase the number of wins over the number of wins which would have resulted if a team went for fourth down less often?

Why causal inference?

Correlation $\neq$ Causation

Why causal inference?

Correlation $\neq$ Causation

Correlation $+$ Assumptions $\stackrel{?}{=}$ Causation

Observing

Image source: Wikimedia

Observation:
People who eat a Mediterranean diet
have lower rates of cardiovascular disease

American Heart Association.

CLINICAL INVESTIGATION AND REPORTS

Mediterranean Diet, Traditional Risk Factors, and the Rate of Cardiovascular Complications After Myocardial Infarction

Final Report of the Lyon Diet Heart Study

Michel de Lorgeril, Patricia Salen, Jean-Louis Martin, Isabelle Monjaud, Jacques Delaye, and Nicole Mamelle

Intervening

CLINICAL INVESTIGATION AND REPORTS

Mediterranean Diet, Traditional Risk Factors, and the Rate of Cardiovascular Complications After Myocardial Infarction

Final Report of the Lyon Diet Heart Study

Michel de Lorgeril, Patricia Salen, Jean-Louis Martin, Isabelle Monjaud, Jacques Delaye, and Nicole Mamelle

Heart attack survivors randomized to

- ▶ advice to follow a Mediterranean diet   (treatment)
- ▶ advice to follow a prudent diet   (control)

Intervening

American Heart Association.

CLINICAL INVESTIGATION AND REPORTS

Mediterranean Diet, Traditional Risk Factors, and the Rate of Cardiovascular Complications After Myocardial Infarction
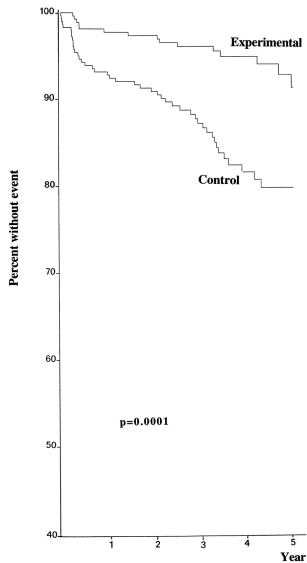### Final Report of the Lyon Diet Heart Study

Michel de Lorgeril, Patricia Salen, Jean-Louis Martin, Isabelle Monjaud, Jacques Delaye, and Nicole Mamelle

Heart attack survivors randomized to

- ▶ advice to follow a Mediterranean diet   (treatment)
- ▶ advice to follow a prudent diet   (control)

Outcome: Recurrent heart attack or death

Intervening

Course objectives

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 causal quantities by pairing those assumptions with statistical evidence

COURSE LOGISTICS

Who should take this course?

The course is designed for upper-division undergraduate students.

Prerequisites.
An introductory statistics course at the level of STSCI 2110, PAM 2100, PSYCH 2500, SOC 3010, ECON 3110, or similar courses.

Familiarity with R programming language
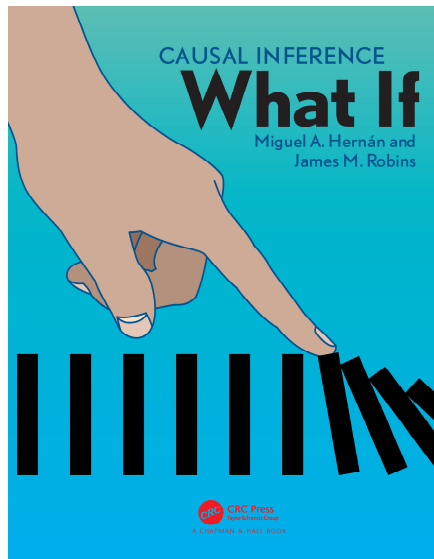(Can take STSCI 2120 concurrently)
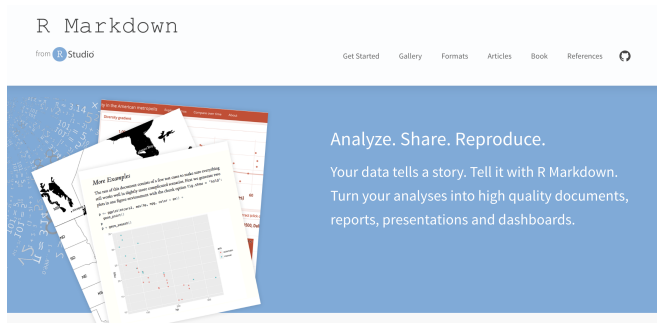
Course materials

All materials will be posted here:

causal3900.github.io

- ▶ Lectures will introduce new material
- ▶ Discussion sections will give practical details
- ▶ Post questions on Ed Discussion
- ▶ Office hours—listed at who we are page

Course readings

Typesetting

As soon as possible, you should

- ▶ Install R   (statistical software)
- ▶ Install RStudio   (user interface)
- ▶ Bookmark the RMarkdown cheat sheet   (documentation)

Note: 20% penalty for reported results that are not reproducible

Method of assessing student achievement

Problem sets	25%
In-class Quizzes	35%
Class project	35%
Peer grading	5%

Academic integrity

Each student in this course is expected to abide by the Cornell University Code of Academic Integrity. Any work submitted by a student in this course for academic credit must be the student's own work.

Generative AI tools should not be used for generating text/code on assignments unless explicitly allowed.

Collaboration

- Encouraged to work together
- Consulting help is great
- Should never involve one student having possession of a copy of all or part of work done by someone else, in the form of an email, an email attachment file, or a hard copy

Late work

- ▶ 5 flex days to be used on problem sets with no questions
- ▶ After your flex days are used up, each late day deducts 10% of the assignment's total points
  - ▶ max score after 1 day late is 90%
  - ▶ max score after 2 days late is 80%
- ▶ No credit for assignments which are more than 5 days late

Attendance

Public health matters—stay home if sick! Let us know.

Otherwise, we expect to see you in class and discussion.

Students with disabilities

You belong in this course. We are happy to work with you on appropriate accomodations—see the syllabus for details about working with Student Disability Services.

Mental health and wellbeing

Your health and wellbeing are important to us!

See syllabus for links to mental health resources. We hope our course helps you thrive a Cornell, and your thriving at Cornell is far more important than anything in this course.

Questions

We look forward to exploring causal inference together!

Card, D. and Krueger, A. B. (2000). Minimum wages and employment: a case study of the fast-food industry in new jersey and pennsylvania: reply. *American Economic Review*, 90(5):1397–1420.

Kearney, M. S. and Levine, P. B. (2015). Media influences on social outcomes: The impact of mtv's 16 and pregnant on teen childbearing. *American Economic Review*, 105(12):3597–3632.

Messerli, F. H. (2012). Chocolate consumption, cognitive function, and nobel laureates.

Yam, D. R. and Lopez, M. J. (2019). What was lost? a causal estimate of fourth down behavior in the national football league. *Journal of Sports Analytics*, 5(3):153–167.