6.S059/15.C08/17.C08
CAUSAL INFERENCE

SPRING 2024

INSTRUCTORS: JOSEPH DOYLE, ROBERTO RIGOBON, TEPPEI YAMAMOTO

TAS: LICHENG LIU, BENJAMÍN MUÑOZ

MIT SCHWARZMAN COLLEGE OF COMPUTING

LOGISTICS

Lectures: Mondays and Wednesday, 3:00PM – 4:30PM, 66-168

Recitation Sessions: TBD

Contact Info and Office Hours:

	Doyle	Rigobon	Yamamoto
Email	jjdoyle@mit.edu	rigobon@mit.edu	teppei@mit.edu
Office Hours	Anytime, by appointment	Anytime, by appointment	M 1-2PM (or by appointment)
Office Location/Zoom			E53-401 (or on Zoom)

	Licheng	Benjamín
Email	liulch@mit.edu	benja_mr@mit.edu
Office Hours	W 1-2 pm (or by appointment)	T 2-3 pm (or by appointment)
Office Location/Zoom	E53-434 (or on Zoom)	E53–414 (or on Zoom)

Scheduling Notes:

- The first class will be on February 5 (Monday) and the last class will be on May 13 (Monday).
- There will be no class on February 19 (President's Day), March 25 and 27 (Spring Break), and April 15 (Patriots' Day).
- There will be a class on February 20, even though it is a Tuesday.

OVERVIEW AND LEARNING GOALS

Does ChatGPT really increase workers' productivity? Did poor working conditions in meatpacking plants contribute to the spread of COVID-19? Would raising minimum wage increase or decrease income inequality in society?

The course provides an accessible overview of modern quantitative methods for answering such questions – questions about whether an action *causes* an outcome to occur – using data. We will make heavy use of applied, real-world examples using Python or R and drawn from various domains, such as economics, political science, business, and public policy. The technical topics the course will cover include potential outcomes, causal graphs, randomized controlled trials, regression, matching, propensity scores, instrumental variable estimation, regression discontinuity, difference-in-differences and panel data methods, and how machine learning techniques can be adopted for answering causal questions.

After successfully completing the course, you will gain an intuitive understanding of the core concepts and techniques to help them produce and consume evidence of causal claims.

DEGREE REQUIREMENT SATISFACTIONS

For Course 6 majors, this subject satisfies the AUS2 requirement. It also counts as a decision-centric subject for Course 6-4 and satisfies the EE Systems Science track requirement.

PREREQUISITES

The official list of the subjects satisfying the course prerequisites includes 6.3800, 6.3900, 6.C01, 14.32, 17.803, 18.05, and 18.650. You are automatically allowed to take the course if you have taken and successfully completed at least one of them. We will also accept enrolment by "permission of instructors," i.e. if you make a case to us that you have taken a course substantively similar to one of these at an equivalent level. Please talk to us if you fit in the latter category.

COURSE REQUIREMENTS

Grades will be based on five problem sets (approximately biweekly, 60%), a midterm exam (held on March 20 in class, 30%), and participation (10%).

READINGS AND COURSE MATERIALS

There is no single required textbook for the course. Reading assignments, as well as other course materials, will be made available on the course Canvas site:

https://canvas.mit.edu/courses/25916

SOFTWARE

We will make heavy use of statistical programming packages, both for problem sets and in occasional in-class lab sessions. Students are allowed to choose between R and Python for their main language to complete assignments.

R is a free statistical programming language that has been widely used by statisticians and applied quantitative researchers. Since it has a long history as the preferred language among statisticians, newly developed methods are often made available in R first and implemented in other languages only later.

Python is a general purpose programming language that has recently become very popular among data scientists and machine learning practitioners. It has an increasingly broad user base, and so the set of statistical methods natively available in Python is rapidly growing.

SCHEDULE

MODULE 1: FUNDAMENTALS

FEB 5, MON

Introduction – Motivation and Examples

FEB 7, WED

Potential Outcomes

Readings:

- (Required) Angrist and Pischke (Mastering 'Metrics) Ch. 1
- (Optional, Advanced) Holland (1986)

FEB 12, MON

Causal Graphs

Readings:

- (Required) Cunningham Ch. 3
- (Optional, Example) Knox, Lowe and Mummolo (2020)
- (Optional, Advanced) Pearl (2009)

MODULE 2: RANDOMIZED EXPERIMENTS

FEB 14, WED

Basic Designs and Identification

Readings:

- (Required, theory) Gerber and Green, Ch. 2
- (Required, example) Gerber, Green, and Larimer (2008)
- (Optional, example) Broockman and Kalla (2016)

Problem Set 1 released (due Feb 28)

FEB 19, MON – NO CLASS (PRESIDENT'S DAY)

FEB 20, TUE (MONDAY SCHEDULE)

Statistical Inference for Randomized Experiments – Part I

Readings:

- (Required) Gerber and Green, Ch. 3.1-3.5

FEB 21, WED

Statistical Inference for Randomized Experiments – Part II

FEB 26, MON

Using Covariates in Randomized Experiments

Readings:

- (Required) Gerber and Green, Ch. 3.6
- (Optional, Advanced) Athey and Imbens (2017)

FEB 28, WED

R/Python Lab – ChatGPT and Productivity

Readings:

- (Required, review with code) Cunningham Ch. 4
- (Required, example) Noy and Zhang (2023)

Problem Set 1 DUE

Problem Set 2 Released (due Mar 18)

MODULE 3: OBSERVATIONAL STUDIES

MAR 4, MON

Controlling for Confounding – Subclassification and Matching

Readings:

- (Required) Cunningham Ch. 5

MAR 6, WED

Regression – Basic Concepts and Mechanics

Readings:

- (Required, primer) Angrist and Pischke (Mastering 'Metrics), Ch. 2
- (Required, theory) Wooldridge Ch.2.1-2.4, Ch.3.1-3.2 (rest of the chapters are also recommended but optional)

MAR 11, MON

How to Use Regression for Causal Inference

MAR 13, WED

Propensity Scores and Weighting

MAR 18, MON

R/Python Lab

Problem Set 2 DUE

MAR 20, WED – IN-CLASS MIDTERM

MAR 25 & 27 – NO CLASS (SPRING BREAK)

MODULE 4: MACHINE LEARNING FOR CAUSAL INFERENCE

APR 1, MON

Estimation and Approximation

Problem Set 3 Released (due April 17)

APR 3, WED

Neural Nets and Decision Trees – A Primer

APR 8, MON

Dimensionality Reduction via Principal Components

APR 10, WED

Dimensionality Reduction via LASSO

APR 15, MON – NO CLASS (PATRIOT'S DAY)

APR 17, WED

Data Collection – the "forgotten piece"

Problem Set 3 DUE
Problem Set 4 Released (due May 1)

MODULE 5: NATURAL EXPERIMENTS

APR 22, MON

Natural Randomization

Readings:

- Angrist and Pischke (Mostly Harmless Econometrics) Chapter 1 (copy on Canvas)

APR 24, WED

Panel Data Analysis with Fixed Effects Regressions

Readings:

- Wooldridge, p484-491

APR 29, MON

Panel Data Analysis with Difference-in-Differences

Readings:

- “Generative AI at Work” (p1-13; 27-31)

MAY 1, WED

Regression Discontinuity

Readings:

- “Regression Discontinuity Designs: An Introduction”

Problem Set 4 DUE
Problem Set 5 Released (due May 13)

MODULE 6: INSTRUMENTAL VARIABLES

MAY 6, MON

Instrumental Variables Part 1

Readings:

- Angrist & Pischke (Mastering ‘Metrics): CH.3, pp 98-110

MAY 8, WED

Instrumental Variables Part 2

Readings:

- Angrist & Pischke (Mastering ‘Metrics): CH.3, pp 142-144 (Appendix)
- Angrist and Krueger: Instrumental Variables and the Search for Identification

CONCLUSION

MAY 13, MON

What Lessons Did We Learn?

Problem Set 5 DUE