Stat286/Gov2002: Causal Inference with Applications

Kosuke Imai

Professor of Government and of Statistics
Harvard University

Fall 2025

Substantive questions in empirical scientific and policy research are often causal. Does voter outreach increase turnout? Are job training programs effective? Can a universal health insurance program improve people's health? This class will introduce students to both theory and applications of causal inference. As theoretical frameworks, we will discuss potential outcomes, causal graphs, randomization and model-based inference, sensitivity analysis, and partial identification. We will also cover various methodological tools including randomized experiments, regression discontinuity designs, matching, regression, instrumental variables, difference-in-differences, and dynamic causal models. The course will draw upon examples from political science, economics, education, public health, and other disciplines.

1 Contact Information

Instructor
Name: Kosuke Imai
Email: imai@harvard.edu
URL: https://imai.fas.harvard.edu
Office: CGIS K306
Office hours: Fridays 2:00pm – 3:30pm (sign up at https://tinyurl.com/KosukeImaiOfficeHours) or by an appointment

Teaching Fellows

Name:	Kentaro Nakamura (Gov2002)	Benedikt Koch (Stat286)	Longlin Wang
Office hours:	Thursday 9:00am – 10:30am	Friday 3:00pm – 4:30pm	
Location:	CGIS Knafel K109	SC 304	
Email:	knakamura@g.harvard.edu	benedikt_koch@g.harvard.edu	longlin_wang@g.harvard.edu

2 Logistics

- Lectures: CGIS South S010, Mondays and Wednesdays 3:00pm – 4:15pm
- Sections:
  - Gov2002: Fridays 1:30pm – 2:45pm, CGIS South S001 Kin-Chung Lam Room
  - Stat286: Thursdays 4:30pm – 5:45pm, SC 228

3 Prerequisites

This course assumes the solid knowledge of

- probability and statistical theory (based on calculus)
- linear models (based on matrix algebra)
- data analysis using R (students may use other software packages at their own risk)

at the level of either (1) Stat110, Stat111, and Stat139 for Stat286, or (2) Gov2001

4 What's the Difference between Gov2002 and Stat286?

The main differences between the two courses will be the following:

- Some of the exercise questions will be different between the two courses with those for Stat286 being more methodologically oriented and those for Gov2002 being more focused upon empirical applications.
- The TF sections will be separately run though students are welcome to attend both.
- The grading will be done separately for each course.
- Gov2002 tends to be taken by graduate students from Government, Public Policy, Sociology, and related disciplines while Stat286 tends to attract graduate and undergraduate students from Statistics, Biostatistics, and related disciplines.
- Some departments and programs may accept only Gov2002 or Stat286 (but not both) as a course that fullfils their requirement. For example, Gov2002 is part of the social science divisional distribution, whereas Stat286 is eligible for the science and engineering and applied science divisional distribution. We are unable to change these requirements. Be sure to enroll in the right course for the requirements you wish to fulfil.

5 The Instructional Tools

We use a variety of instructional tools to run this course.

- **Course website** (https://canvas.harvard.edu/courses/158099): This is the entry point to all the course materials. The links to the review questions and problem sets will be posted here too.
- **Course calendar** (https://tinyurl.com/stat286gov2002calendar): For your convenience, class meetings and TF sections as well as various deadlines are made available through this Google calendar.
- **Perusall**: This platform will host lecture videos and lecture slides. You can also ask questions about their contents by annotating relevant parts of the lecture slides and videos rather than directly emailing an instructional staff. We encourage you to watch their Getting Started video and learn about how to ask questions if you are new to Perusall. You can enter the Perusall course site from Canvas.

- **Ed** (https://edstem.org/us/courses/81097): This platform will host all assignments and their solutions. Questions about assignments should be posted here rather than directly emailing an instructional staff. You may find this user guide helpful to orient yourself to the platform.

- **Gradescope** (https://www.gradescope.com/courses/1074556): This is where you will submit all of your assignments. The grades for the assignments will be available here as well. There is a student guide you can check for any questions about the workflow.

6 The Structure of the Course

The course is divided into 10 modules. Each module is centered around a particular causal inference topic, and is roughly based on the following basic structure.

- **Pre-module assignments**: Watch video lectures available at Perusall *before* the class meetings of each module. *We will assume that you have watched these videos and asked questions you have about them at Perusall.* We also strongly recommend that you use the reading assignments for better understandings of the materials covered in each module.

- **Review Questions**: These questions are designed to better understand the materials presented in the pre-module video lectures. You may also find the assigned readings useful solving them. Review Questions are not graded, and you are not required to submit your answers. The solutions of these Review Questions will be made available before the next class meeting.

- **Class Meetings**: For each module, we have two or three class meetings. During those meetings, the instructor will give lectures by reviewing the basic materials, discussing some of Review Questions, and introducing additional materials. Lectures are designed to further solidify the concepts covered in each module.

- **Problem Sets**: There is one problem set for each module. You are required to complete the selected questions from Problem Set and submit your own answers to the selected questions. Different questions may be selected for Gov2002 and Stat286 students.

- **TF section**: The TF sections will review the course materials and provide some guidance to solving Problem Sets. There will be separate sections for Gov2002 and Stat286. The section attendance is not mandatory but encouraged. Students can also attend both sections if they wish.

7 Course Requirements

The final grade is based on the following components:

- **Class participation** (10% of the course grade): We evaluate the level of your engagement during the in-person class meetings and TF sections as well as Perusall and Ed online discussions.

- **Problem Sets** (20% of the course grade): You should choose a total of 6 modules out of 10 modules (3 modules before the midterm and 3 after the midterm), for which you are to submit your answers to the selected questions from Problem Sets. Each module will be

weighted equally. If you decide to submit the solutions to additional exercises, we will take your best scores among those submitted. The following rules apply to all problem sets:

  - *Submission policy.* All answers including the math and computer code are encouraged to be incorporated into the Rmarkdown file provided by the instructional staff. Handwritten math answers are acceptable once merged into one pdf document. For the exercises, each student should submit the pdf file electronically to Gradescope. Once you upload the PDF file, you will see a list of the questions in the assignment and thumbnails of your file. For each assigned question, click the PDF page(s) that contains your answer. No late submission will be accepted unless you obtain a prior approval from the instructor.

- **Mid-term and final exams** (70% of the course grade): The midterm and final exams will be held in class, each lasting two hours. They cover the first and second half of the materials, respectively, and are equally weighted. A practice exam will be given before each exam.

8 Collaboration Policy

Throughout this course, you are encouraged to collaborate with another student in the class so that you can learn from one another. However, for a problem set, you may choose no more than one student as a collaborator. In addition, you must write up your own solutions and make a separate submission. **Under no circumstances may you copy someone else's answer including computer code, mathematical derivation, and substantive interpretation.** Please do not forget to write down your partner's name on your solution to indicate collaboration.

You are also strongly encouraged to reach out to the instructional staff through Perusall (about lectures), Ed (questions about other course materials), and office hours about any questions you might have about the course materials. Students should also feel free to ask questions and answer the questions posed by others at Perusall and Ed, which will count towards class participation.

9 AI Policy

In this class, we apply the same collaboration policy to both human and generative artificial intelligence (GenAI). You may use GenAI tools, such as ChatGPT, to support your work on problem sets. However, just as you would not ask a human study partner to complete the assignment for you, you must not entirely rely on GenAI to solve problems on your behalf. Appropriate uses of GenAI include help with programming, identifying relevant articles or books, and brainstorming possible approaches. It is not acceptable to use GenAI to directly solve problems or to copy its answers. Such practices hinder your understanding of the course material and are likely to lead to poor performance on closed-book exams.

10 Textbook

The required textbook for this course is,

- Imbens, Guido W. and Rubin, Donald B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press.

The purchasing link for Harvard Coop is here. In addition, you may find the following books useful,

- Angrist, Joshua D. and Pischke, Jörn-Steffen. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.

- Ding, Peng. (2024). *A First Course in Causal Inference*. Chapman & Hall/CRC.
- Hernán M. A., Robins James M. (2020). Causal Inference. Boca Raton: Chapman & Hall/CRC. Available at https://miguelhernan.org/whatifbook

11 Course Plan

This course intends to provide a brief introduction to the following 10 topics in causal inference. The problem set for each module is due by 10 am on the day listed.

Module 1 Introduction

Class 1 (9/3)	Overview of the course
Video lecture	Potential outcomes
Reading	Imbens & Rubin, Chapter 1 (Optional Chapter 2)
Class 2 (9/8)	Potential outcomes
Problem Set 1	September 15

Module 2 Permutation Test

Video lecture	Permutation test
Reading	Imbens & Rubin, Chapter 5 (Optional Chapter 4)
Class 1 (9/10)	Inverting permutation tests
Class 2 (9/15)	Conditional randomization tests
Problem Set 2	September 22

Module 3 Inference for Average Treatment Effects

Video lecture	Average treatment effects
Reading	Imbens & Rubin, Chapters 6, 9 (Skip 9.6–9.7), and 10 (Skip 10.6–10.7)
Class 1 (9/17)	Evaluation of individualized treatment rules
Class 2 (9/22)	Stratified experiments
Problem Set 3	September 29

Module 4 Linear Regression and Randomized Experiments

Video lecture	Simple linear regression
Reading	Imbens & Rubin, Chapters 7, 9 (9.6–9.7), and 10 (10.6–10.7)
Class 1 (9/24)	Cluster randomized trials
Class 2 (9/29)	Covariate adjustment in randomized experiment
Problem Set 4	October 6

Module 5 Instrumental Variables

Video lecture	Noncompliance in randomized experiments
Reading	Imbens & Rubin, Chapters 23 and 24
Class 1 (10/1)	Instrumental variables

Class 2 (10/8) Two stage least squares

Problem Set 5 October 13

Midterm Week

Review Session TBA

Midterm October 15

Module 6 Regression Discontinuity Designs

Video lecture Sharp regression discontinuity designs

Reading Guido W. Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Jounal of Econometrics*, 142(2):615–635, February 2008. doi: 10.1016/j.jeconom.2007.05.001.

Class 1 (10/20) Diagnostics of regression discontinuity design

Class 2 (10/22) Fuzzy and other regression discontinuity designs

Problem Set 6 October 29

Module 7 Observational Studies

Video lecture Regression with observational data

Reading Imbens & Rubin, Chapters 21 and 22

Class 1 (10/27) Sensitivity analysis

Class 2 (10/29) Partial identification

Problem Set 7 November 5

Video lecture Directed acyclic graphs and backdoor criterion

Reading Felix Elwert. *Handbook of Causal Analysis for Social Research (ed. Stephen L. Morgan)*, chapter 13. Graphical Causal Models, pages 245–273. Springer, Dordrecht, 2013. ISBN 9789400760936.

Class 3 (11/3) do-calculus and frontdoor criterion

Module 8 Matching and Weighting

Video lecture Matching and weighting methods

Reading Imbens & Rubin, Chapters 13, 15, and 18 (Optional Chapters 12, 14, and 19)

Eli Ben-Michael, Avi Feller, David A. Hirshberg, and Jose R. Zubizarreta. The balancing act in causal inference. *arXiv preprint*, arXiv: 2110.14831, 2021.

Class 1 (11/5) Optimal matching and weighting methods

Class 2 (11/10) Semiparametric causal inference

Problem Set 8 November 17

Module 9 Causal Mechanisms and Heterogeneity

Video lecture Controlled Direct Effects, Natural Direct and Indirect Effects

Reading Kosuke Imai, Luke Keele, and Teppei Yamamoto. Identification, inference, and sensitivity analysis for causal mediation effects. *Statistical Science*, 25(1):51–71, February 2010. doi: 10.1214/10-STS321.

Class 1 (11/12) Causal mediation analysis

Video lecture Causal heterogeneity

Class 2 (11/17) Individual treatment rules

Problem Set 9 November 24

Module 10 Fixed Effects, Difference-in-Differences, and Synthetic Control Method

Video lecture Difference-in-differences and fixed effects regressions

Reading Angrist & Pischke, Chapter 5

Class 1 (11/19) Nonlinear difference-in-differences design

Class 2 (11/24) Matching and weighting methods for panel data

Problem Set 10 December 6

Video lecture Synthetic control method

Reading Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, June 2010. doi: 10.1198/jasa.2009.ap08746.

Class 3 (12/1) Difference-in-differences, regression, and synthetic control

Wrapup (12/3)

Final Exam Exam Period