Stat186/Gov2002: Causal Inference

Kosuke Imai

Professor of Government and of Statistics
Harvard University

Fall 2019

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 statistical theory and practice 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 Knafel K306
Office hours: Tuesdays 1:15pm – 2:45pm or simply stop by anytime (open door policy)

Teaching Fellows

Name:	Niloy Biswas	Casey Petroff	Soichiro Yamauchi
Office hours:	Friday 12:00pm – 1:30pm	Monday 10:30am – 12:00pm	Friday 3:00pm – 4:30pm
Location:	SC 705	CGIS K031	CGIS K107
Email:	niloy_biswas@g.harvard.edu	petroff@g.harvard.edu	syamauchi@g.harvard.edu

2 Logistics

Lectures focus on the conceptual and theoretical issues while TF sections focus on their empirical applications.

- Lectures: Tuesdays, Thursdays 10:30am – 11:45am, Emerson 210
- TF Sections: Thursday 3:00pm – 4:00pm CGIS K050 (Casey), 6:00pm – 7:00pm SC 705 (Niloy) and Friday 2:00pm – 3:00pm CGIS K107 (Soichiro)

Use of laptops and cell phones during the lectures is not allowed unless you are given a permission from the instructor. Research has shown that the use of these electronic devices often

interferes your learning and has a negative impact on those around you. We will bring copies of lecture slides to class so that you can take notes directly on the slides. Lecture slides will be posted at Piazza by the end of the lecture day.

3 Questions, Announcements, and Submissions

- In addition to TF sections and office hours, please use the *Piazza Discussion Board* at `https://piazza.com/` when asking questions about lectures, problem sets, and other course materials. This allows all students to benefit from the discussion and to help each other understand the materials. Both students and instructors are encouraged to participate in discussions and answer any questions that are posted. To join the Stat186 / Gov2002 Piazza site, click `https://piazza.com/harvard/fall2019/stat186gov2002` You will then be prompted to enter your `harvard.edu` email address to confirm your registration. All class announcements will be made through Piazza.
- We will use Canvas for the submission of all assignments including the final exam and projects. The Canvas site for this course is at `https://canvas.harvard.edu/courses/58319`
- A Google calendar that contains the information about the course logistics is available at `http://bit.ly/stat186gov2002calendar`

4 Prerequisites

This course assumes the solid knowledge of

- probability theory
- statistical theory
- linear models
- data analysis using R

at the level of either (1) Stat110, Stat111, and Stat139 or (2) Gov2000 and Gov2001. Students may use other software programs such as Python and Stata, but we will only provide support for R. To understand the required level of mastery, students should consult the following books and make sure that they have an appropriate background:

- Blitzstein, Joseph K. and Hwang, Jessica (2014). *Introduction to Probability*, Chapman & Hall / CRC.
- Imai, Kosuke (2017). *Quantitative Social Science: An Introduction*, Princeton University Press.

Students are strongly encouraged to make sure that they have a sufficient preparation before deciding to enroll in the course.

5 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 in lectures, TF sections, and Piazza discussion forum.
- **Problem sets** (40% of the course grade): A total of four problem sets will be given throughout the semester. Each problem set will equally contribute to the final grade. The following instructions will apply to all problem sets:
  - *Collaboration policy.* Students are allowed to discuss the problem sets with other students in class. However, each student is expected to write up their own answers. UNDER NO CIRCUMSTANCES MAY STUDENTS COPY EACH OTHER'S ANSWER INCLUDING COMPUTER CODE AND MATHEMATICAL DERIVATION. In addition, the names of other students with whom you consulted should be given at the beginning of your answer to each of the questions.
  - *Online help and office hours.* You are strongly encouraged to consult with teaching fellows through sections and office hours. Students should also ask questions and answer the questions posed by others at Piazza too, which will count towards class participation.
  - *Submission policy.* All answers should be typed. Both the answers, which must be compiled as a pdf file, and source code should be submitted electronically to Canvas. We encourage the use of Rmarkdown so that the code, results, justification, and interpretation are presented together for each question. No late submission will be accepted unless you obtain a prior approval from the instructor.
- **Final take-home exam** (50% of the course grade for Stat186 students): The final take-home exam is similar to a problem set in its format. However, you should not consult with anyone including teaching fellows. Clarifying questions can be asked under a designated thread at Piazza. The final take-home exam will be posted at the beginning of the reading period and will be due sometime during the final exam period. The detailed instruction will be given later in the semester.
- **Final group project** (50% of the course grade for Gov2002 students): The final project must be a collaborative project with another student in this class. If you would like to do an individual project, which is generally discouraged, you must obtain a prior permission from the instructor as soon as possible. The final project can be methodological (e.g., development of a new causal inference method) or empirical (e.g., application of an existing causal inference method). The quality of project will be judged based on the originality of intellectual ideas presented in the final report. Students are required to meet the following milestones and are encouraged to discuss their projects with the instructor and teaching fellows throughout the semester:
  - October 1: One-page proposal, which should include a brief statement of the problem to be solved (or the question to be answered) and propose a feasible plan for conducting research
  - November 12: Presentation slides with initial results (no more than 15 slides including the title page)
  - November 19: Provide feedback to your classmates

– December 19: Final report (no more than 20 double-space pages including the title, tables, figures, and references). Before writing up the report, carefully read the memo, "How to write an empirical social science paper," available at `https://imai.fas.harvard.edu/teaching/files/HowToPaper.pdf`

The detailed instruction will be given later in the semester.

6 Recommended Textbooks

There are no required textbooks for this course. However, you may find the following three books useful as they cover the related materials. The first two are available at Coop for purchase.

- Angrist, Joshua D. and Pischke, Jörn-Steffen. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*, Princeton University Press.
- Imbens, Guido W. and Rubin, Donald B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press.
- Hernán M. A., Robins James M. (2019). Causal Inference. Boca Raton: Chapman & Hall/CRC, forthcoming. `https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/`

7 Course Plan

This course intends to provide a brief introduction to the following topics in causal inference.

Week 1 Introduction (September 3 – 6)

1. Overview of the course
2. Potential outcomes

Week 2 Randomized Experiments from Fisher's Perspective (September 10 – 13)

1. Fisher's exact test
2. Permutation inference

   Problem set 1: Posted on September 13, Due on September 24

Week 3 Randomized Experiments from Neyman's Perspective (September 17 – 20)

1. Inference for the average treatment effects
2. Stratified randomized experiments

Week 4 Linear Regression and Randomized Experiments (September 24 – 27)

1. Simple linear regression
2. Covariance adjustment

Week 5 Instrumental Variables (October 1 – 4)

1. Noncompliance in randomized experiments
2. Instrumental variables in observational studies

   Project proposal: Due on October 1

Problem set 2: Posted on October 4, Due on October 15

Week 6 Regression Discontinuity Designs (October 8 – 11)

1. Sharp and fuzzy regression discontinuity designs
2. Interrupted time series designs

Week 7 Observational Studies (October 15 – 18)

1. Partial identification
2. Directed acyclic graphs (DAGs)

Week 8 Matching and Weighting (October 22 – 25)

1. Matching methods
2. Weighting methods

Problem set 3: Posted on October 25, Due on November 5

Week 9 Causal Mechanisms (October 29 – November 1)

1. Causal mediation analysis
2. Controlled direct effects

Week 10 Fixed Effects and Difference-in-Differences (November 5 – 8)

1. Fixed effects regression
2. Difference-in-differences design

Week 11 Synthetic Control and Dynamic Treatment Regimes (November 12 – 15)

1. Synthetic control method
2. Marginal structural models

Initial project results: Due November 12

Week 12 Causal Heterogeneity (November 19 – 22)

1. Heterogeneous effects of a single treatment
2. Effects of multiple treatments

Initial project comments: Due November 19

Problem set 4: Posted on November 22, Due on December 3

Weeks 13 Spillover Effects (November 26 – 29)

1. Spillover effects
2. Thanksgivng holiday (No class)

Weeks 14 Wrap-up (December 3)