Stat186/Gov2002: Causal Inference

Kosuke Imai

Professor of Government and of Statistics
Harvard University

Fall 2018

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
Office: CGIS Knafel K306
Email: Imai@Harvard.Edu
URL: https://Imai.Fas.Harvard.Edu
Office hours: Thursdays 3pm – 4:30pm (`https://tinyurl.com/Imai-OfficeHours` to sign up)
or simply stop by

Teaching Fellows

	Shom Mazumder	Azeem Zaman
Office:	CGIS Knafel Cafe	TBA
Office hours:	Wednesdays 4pm – 5:30pm	Mondays 4pm – 5pm and Tuesdays 4pm – 5pm
Email:	smazumder@g.harvard.edu	azaman@g.harvard.edu

2 Logistics

- Lectures: Tuesdays and Thursdays, 1:30pm – 2:45pm, Harvard Hall 202
- TF Sections: TBA

Use of laptops and cell phones during the lectures is discouraged. Research has shown that the use of these electronic devices often interfers 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 9am on the day of lecture. Those of you who wish to use laptops during the lectures will be asked to sit in a certain section of classroom so that the use of laptops does not affect other students.

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/fall2018/stat186gov2002` You will then be prompted to enter your `harvard.edu` email address to confirm your registration. All class announcements as well as electronic submissions of assignments will be made through Piazza.

4 Prerequisites

This course assumes the knowledge of basic probability and statistics at the level of either (1) Stat110 and Stat111 or (2) Gov2000 and Gov2001. In addition, we assume that students will be able to conduct basic data analysis using computer software. We support the use of statistical programming language R, but students may use other software programs such as Python and Stata. To understand the prerequisites, 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.

5 Course Requirements

The final grades are based on the following items:

- **Class participation** (10% of the course grade): We evaluate the level of your engagement in lectures, TF sections, and online discussion forum.
- **Problem sets** (40% of the course grade): A total of four problem sets (possibly split into a couple of shorter assignments) 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.* Unless otherwise stated, no collaboration is allowed. However, you are encouraged to consult with teaching fellows through sections, office hours, and Piazza. For some selected questions, a student may be allowed to collaborate with no more than two other students in the class. Even for those questions, 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 collaborators should be given at the beginning of your answer to each of these questions.
  - *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 Piazza.

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% for undergraduate 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 project** (50% for graduate students): The final project must be a collaborative project with another student in this class (unless you obtain a permission from the instructor to do an individual project). 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:

  - September 25: 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 13: Presentation slides with initial results (no more than 15 slides including the title page). We will ask you to do a short presentation in the following week and receive feedback from your peers as well as us.
  - December 20: Final report (no more than 20 double-space pages including the title, tables, figures, and references)

  The detailed instruction will be given later in the semester.

6 Textbooks

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

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

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

7 Course Outline

The topics we plan to cover in this course are given here:

1. Introduction
   (a) Overview of the course
   (b) Potential outcomes

2. Randomized Experiments from Fisher's Perspective
    (a) Permutation test
    (b) Randomization inference
3. Randomized Experiments from Neyman's Perspective
    (a) Inference for the average treatment effects
    (b) Stratified randomized experiments
4. Linear Regression and Randomized Experiments
    (a) Simple linear regression
    (b) Covariance adjustment
5. Instrumental Variables
    (a) Noncompliance in randomized experiments
    (b) More on instrumental variables
6. Regression Discontinuity Designs
    (a) Sharp regression discontinuity design
    (b) Fuzzy and other regression discontinuity designs
7. Confounding Bias
    (a) Observational studies
    (b) Confounder adjustment and directed acyclic graphs (DAGs)
8. Matching and Weighting
    (a) Matching methods
    (b) Weighting methods
9. Causal Mechanisms
    (a) Causal mediation analysis
    (b) Controlled direct effects
10. Panel and Time-Series Cross Section Data
    (a) Matching methods and fixed effects
    (b) Difference-in-differences
    (c) Synthetic control method
    (d) Dynamic treatment regimes and marginal structural models
11. Additional Topics
    (a) Causal inference with missing data
    (b) Causal inference with spillover effects
    (c) Causal inference and machine learning