# Introduction to Quantitative Methods II: Causal Inference

**New York University**
**Spring 2023**

| | |
|---|---|
| Instructor: | Prof. Cyrus Samii (cds2083@nyu.edu) |
| Office hours: | `https://calendly.com/cdsamii` |
| Schedule: | Tuesdays and Thursdays, 10:00-11:50am |
| Location: | 19 W 4th St, Room 217 |
| Course site: | `https://cyrussamii.com/?page_id=3650` |
| Teaching Assistant: | Giacomo Lemoli (gl1759@nyu.edu) |
| Recitations: | TBD. |

## Overview

This course provides a current perspective on causal inference in quantitative social science research. We focus on non-parametric identification methods and then non-parametric and semi-parametric estimation and frequentist inference methods. We will emphasize research design and robust estimation and inference.

## Prerequisites and Restrictions

The course has two prerequisites. First, students should have working knowledge of basic probability theory and linear regression theory at the level of POL-GA 1250, Quant I. Second, students should have some background in writing scripts to implement statistical analyses in either R or Stata. There is also a restriction with respect to taking the course for credit. The course provides foundational methodological training to Politics PhD students in their first or second year as part of their required sequence of courses. With rare exception, only Politics PhD students will be allowed to take the course for a grade. (We only have one teaching assistant, so this limits our ability to service students from other departments, unfortunately.) People from other programs may audit or attend informally if space permits.

## Texts

Required textbooks for the course include the following:

- Imbens, Guido W., and Donald B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences.* Cambridge University Press, 2015.
- Cunningham, Scott. *Causal Inference: The Mixtape.* Yale University Press, 2021.

These books and the course lectures also reference journal articles and working papers, as indicated below. Previous versions of this course had used Angrist and Pischke and Morgan and Winship's textbooks, however I don't think those textbooks have kept up with the literature. The Imbens and Rubin textbook contains important statistical foundations and references for analyzing randomized experiments, while Cunningham is an up to date guide to observational methods.

# Software

You will have the choice to work in R or Stata. It is useful to obtain fluency in both. I encourage using RMarkdown or Stata Markdown for your assignments. This is a great investment that will pay off in the long run in terms of productivity as well as reproducibility.

- RMarkdown runs most easily through RStudio. Details here: `https://rmarkdown.rstudio.com/`
- Stata Markdown is a package that runs within Stata. Details here: `https://grodri.github.io/markstat/`

# Requirements and policies

## Homework

You will receive homework about every two weeks. You will have to submit your completed assignment within a week; exact deadlines will be made clear on the assignment. You can work with others, but to receive credit, your homework must comply with the following guidelines:

- You must turn in a PDF copy of your own homework by the stated deadline to both the professor and TA.
- The assignment that you turn in must clearly reflect your own thinking. Sets of verbatim copies of homework will have credit reduced by half.
- Homework assignments may be hand written or typed, but they must be clearly legible.
- Estimates obtained from data analysis programs (e.g., Stata or R) must be formatted properly into tables or graphs resembling journal presentation styles. You should use a table formatting function (e.g., outreg2 or esttab in Stata, or apsrtable or stargazer in R). Use a reasonable (2 or at most 3) number of digits after decimal points, report standard errors or confidence intervals along with coefficients, clarify what are the dependent variables in each table or figure, and explain in footnotes to your tables or figures what kinds of estimators or adjustments have been used. Print outs of raw screen output or commented logs will not receive any credit. However, you may include such output as an appendix so that the grader can troubleshoot.

- Mathematical derivations should include all key steps with explanations for important techniques.

Homework will be graded for points as indicated on each assignment and count toward 50% of your grade.

### Mid-term exam

An in-class mid-term exam will take place mid-way through the semester (exact date to be confirmed). The mid-term serves the purpose of evaluating individual progress, which in turn helps me to understand where to place emphasis for the remainder of the semester. If you are unable to make it to the exam, you must provide notice at least a week prior so that we can arrange an alternative time. The mid-term will count toward 15% of your grade.

### Final exam

A take-home final exam will be scheduled during the final examination period. The final also serves the purpose of evaluating individual progress, which in turn allows me to provide individualized recommendations on where students should apply effort to strengthen their methodological foundations. If you are unable to work during the exam period, you must provide notice at least a week prior so that we can arrange an alternative time. The final will count toward 25% of your grade.

### Attendance and participation

Attendance and participation in class discussions is required and counts toward 10% of your grade.

### Special needs

Students with special needs should come to office hours or schedule an appointment with the instructor to discuss possible accommodation.

## Topics

# 1 Introduction

- Syllabus review
- Causal inference in problem-focused social science research (observation versus intervention)
- Potential outcomes: "the science," defining causal effects (ATE, ATT, mediation effect, interference, probability of necessity), manipulability criterion.

- Directed acyclic graphs (DAGs) and non-parametric structural equation models (NPSEMs): how to encode a behavioral model into a DAG?

*Readings*: Cunningham (2021, Ch. 1, 3, 4), Imbens and Rubin (2015, Ch. 1), Samii (2016), Samii (2022).

# 2 Working with Potential Outcomes and DAGs

- Identifying the ATE under SUTVA and strong ignorability
- Using potential outcomes
- Using DAGs (graph moralization)

*Readings*: Elwert (2013), Greenland and Pearl (2014), Imbens and Rubin (2015, Ch. 3).

# 3 Statistical Inference for Descriptive Quantities

## Whole population

- Difference in means
- Linear regression

## Sample

- Difference in means
- Linear regression
- Analytical
- Bootstrap
- Cluster sample

*Readings*: Cunningham (2021, Ch. 2).

# 4 Statistical Inference for Causal Quantities

## Whole population

- Difference in means
- Linear regression
- Analytical

- Bootstrap
- Cluster assignment

### Sample

- Difference in means
- Linear regression
- Analytical
- Bootstrap
- Cluster assignment and sample

*Readings*: Abadie, Athey, Imbens, and Wooldridge (2020).

# Canonical Observational Designs

## 5 Conditioning to identify causal effects

- Backdoor criterion
- Collider biases: post-treatment bias, sample selection
- Bias amplification

## 6 Regression control

- One-way fixed effects (FWL, OVB)
- Continuous covariates (FWL, OVB)
- Double machine learning
- Effective sample under effect heterogeneity

## 7 Matching and Weighting

- Exact matching
- Approximate matching and propensity score theorem
- Bias correction
- IPW
- Balance weighting

## 8 Instrumental Variables Basics

- IV under structural model and under PO model
- 3 Key Conditions
- LATE Theorem
- Bound on the ATE
- Bias and weak instruments

## 9 Instrumental Variables Extensions

- Shift share
- Judge Fixed Effects/JIVE
- Granular IV

## 10 Front door criterion

## 11 Regression Discontinuity

- Parametric
- Nonparametric local regression
- Optimized RD and honest inference
- Threats

## 12 Regression Discontinuity Extensions

- Fuzzy RD
- Kink design

## 13 Difference in Differences for a Single Event

- Two period
- Pre-trends
- Event study

# 14 Difference in Differences for Multiple Events

- TWFE bias
- Stacking, Weighting, and Imputation

# 15 Synthetic Control for One Treated Unit

- Identification and estimation
- Statistical inference

# 16 Synthetic Control for Multiple Treated Units

- Interactive FE
- Augmented Synthetic Control

# Experimental Design

# 17 Basics of Experimental Design

- Randomization Inference
- MDE and power analysis

# 18 Experimental Design Effects

- Clustering
- Stratification
- Factorial designs and interactive effects

## 19 Adaptive and multiphase experiments

## 20 Moderator effects

## 21 Causal mediation

## 22 Interference and spillovers

## 23 Multiple endpoints

## 24 Missing data

## 25 Generalization and external validity

# References


Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020). Sampling-based versus design-based uncertainty in regression analysis. *Econometrica 88*(1), 265–296.

Cunningham, S. (2021). *Causal inference: The Mixtape*. Yale University Press.

Elwert, F. (2013). Graphical causal models. In *Handbook of causal analysis for social research*, pp. 245–273. Springer.

Greenland, S. and J. Pearl (2014). Causal diagrams. *Wiley StatsRef: Statistics Reference Online*, 1–10.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Samii, C. (2016). Causal empiricism in quantitative research. *The Journal of Politics 78*(3), 941–955.

Samii, C. (2022). Methodologies for “political science as problem solving”. *Technical Report, New York University*.