Causal Inference

(Biostat M235)

Falco J. Bargagli-Stoffi
falco@ucla.edu

UCLA 2025

Summary of course content
The course aims to introduce the key concepts and state-of-the-art methods for causal inference from randomized experiments and observational studies. We will first introduce the basic concepts of the potential outcome framework (with a particular focus on the essential role of the treatment assignment mechanism) and the causal graphical model approach. Then, we will cover different situations corresponding to different assumptions concerning the assignment mechanism. We will discuss the design and analysis of experiments and how to make inference under different modes, including design (randomization)-based, frequentist, and Bayesian. We will cover the design and analysis of observational studies with regular assignment mechanisms where the unconfoundedness assumption is assumed to hold. We will introduce irregular assignment mechanisms, discuss strategies for dealing with experimental studies with broken randomizations, bridge with instrumental variable methods, and introduce the principal stratification and mediation frameworks. We will cover heterogeneous treatment effects and causal machine learning approaches to causal heterogeneity. We will explore quasi-experiments through the lens of the regression discontinuity design approach. Finally, we will delve into causal inference with panel data, examining the techniques and strategies used to establish cause-and-effect relationships in temporal datasets.

1. Foundations of causal inference and the potential outcome framework
   - Setting up the problem
   - Causal estimands
   - Causal assumption
   - Causal identification
   - The role of the assignment mechanism: definitions and taxonomy

2. Introduction to the causal graphical model approach
    - Directed Acyclic Graphs (DAG)
    - Bad and negative controls
    - Surrogates
    - Mediators
    - Dynamic Treatments
3. Introduction to randomized control studies
    - Design and analysis of randomized experiments
    - Pre-planning and power analysis
    - Modes of inference:
        - Fisher exact tests
        - Neyman frequentist perspective
        - Regression analysis
        - Bayesian model-based imputation
    - Introduction to broken-randomizations
4. Introduction to observational studies under unconfoundedness
    - The role of the propensity score
    - Designing observational studies
        - Matching
        - Weighting
        - Trimming
    - Analysis of observational studies:
        - Stratification
        - Weighting estimators
        - Matching estimators
        - methods based on the outcome models and regression
5. Observational studies under possible violations of unconfoundedness
    - Method combinations
        - Bias corrected estimators
        - Doubly robust estimators
    - Sensitivity analysis
        - The Rosenbaum-Rubin procedure
        - E-values
        - Bounds

6. Irregular assignment mechanisms and instrumental variables (IV)
    - Broken randomizations - revisited
    - Bayesian IV analysis: relaxing some of the assumptions
    - Point, partial, weak identification of causal effects
    - The role of covariates
    - IV and beyond:
        - Mediation analysis
        - Principal stratification
7. Heterogeneous treatment effects and causal machine learning
    - Heterogeneous treatment effects estimation
    - Subgroup identification
    - Policy targeting
    - Off-policy learning
    - Bandits and adaptive experiments
8. Quasi-experiments: Regression Discontinuity Designs (RDD)
    - The identification strategy: continuity vs local randomization
    - Sharp RDD
    - Fuzzy RDD and the local LATE interpretation of RDD
    - Graphical analysis, assessing identification assumptions
    - Bandwidth selection
    - Multiple thresholds, multiple forcing variables
9. Difference in difference, synthetic controls and beyond
    - DID and extensions (e.g., CIC, Synthetic DID)
    - Lagged dependent variables
    - Synthetic controls and permutation inference
    - Some recent developments (e.g., Matrix Completion, Time Series)

Teaching material

Selected chapters from the book list below. Articles in (bio)statistical and econometric journals.

- Imbens G. W., Rubin D. B. (2015) Causal Inference for Statistics, Social, and Biomedical Sciences, Cambridge University Press
- Ding P. (2023) A first Course in Causal Inference, Routledge
- Pearl J. (2009) Causality, Cambridge University Press

Final exam and grading

There will be 3 take-home assignments (simulation and real data exercises) and a final project. Final grading will be based on the 3 assignments (15% each), final essay (45%), and participation (10%).

Use of Large Language Models

The usage of large language models (LLMs) in this course is possible—even if not recommended—following the guidelines below:

- You must use LLM tools in a way that helps you learn, not hampers learning. Remember that these are tools to assist you, not a replacement for your own learning of the material, critical thinking, and writing.
- The only acceptable use of LLMs on assignments is for proofreading. This should only be for simple grammar checks, not extensive rewriting.
- It is acceptable to use LLMs to provide you with other explanations of concepts or organize your notes, and there is no need to disclose these. However, if the LLM gives you incorrect information and you use that incorrect information on an assignment, you are the one who will be held accountable.
- Be transparent: if you used an LLM for proofreading, you must acknowledge that. We expect that you will include a short paragraph at the end of the assignment or in the final report explaining which LLM tool you used, what you used it for, and why.
- If artificial intelligence tools are used in nefarious or unacknowledged ways, you may be subject to academic misconduct policies (see: https://www.deanofstudents.ucla.edu/academic-integrity).