# Advanced Research Methods I (Causal Inference)

Ilona Lahdelma

Fall, 2024

E-mail: `ilahdelm@clio.uc3m.es`
Office Hours: TBA
Office: Getafe campus, 18.2.E.14
Web:**www.ilonalahdelma.com**
Class Hours: Mondays 10.00-13:00
Class Room: 18.1.A01

---

## Course Description

Do hospitals make people healthier? This is a classic example in causal inference, where looking at the observational data leads us to distorted views of what we would get intuitively right. Causal inference in statistics is the art and science of making causal claims about data where such pitfalls of misinterpreting the data exist. How can we *know* that *a* leads to *b*? What if *a* and *b* are merely correlational because they are both caused by something unknown? How can we model this? How can we trust our estimates? How do we go beyond correlations?

This course introduces students to causality, which is just as much a theory of statistics as computation. Students will be introduced to causal diagrams, informed data modelling and the potential outcomes framework, as well as the standard econometric methods to draw causal inferences from data: Randomized control trials, matching, instrumental variables, regression discontinuity and differences-in-differences.

At the end of the course, students will showcase what they learned in a take-home exam, which will consist of exercise covering the entire range of the course.

## Required Materials

All slides and code will be provided, so students only need to come with basic understanding of statistical analysis. Students need to have the latest version of R downloaded to their laptops and come with their laptops to class. The course will mainly follow the book *The Effect* by Nick Huntington-Klein (2021) and *Mostly Harmless Econometrics* by Joshua Angrist and Jörn-Steffen Pischke (2009). Those who want to go beyond what we learn in class to understand and study the fundamental concepts of causal inference across the sciences, are welcome to read *Causal Inference for Statistics, Social, and Biomedical Sciences* by Guido Imbens and Donald Rubin (2015). It is not a prerequisite to have read the books, but students are encouraged to consult the books throughout the course. On weeks 2,4,6,8, 10, and 12 we will read and replicate published journal articles using methods of causal inference. The used paper will be announced on the previous week. The code for each week will be uploaded prior to the lesson to Aula Global.

## Prerequisites

Introduction to statistics.

## Course Objectives

Successful students will have an understanding at the end of the course about

1. Why and how causal inference is used in the social sciences
2. What types of research questions and problems lend themselves to causal inference
3. Different ways of generating causal estimates
4. The main stages of the research design process
5. The main stages of data management and analysis
6. How reliable causal estimates can be
7. How to stay up to date with the constantly evolving methods literature

## Course Structure

### Class Structure

Typically the first half of each sessions (1,5 hours) focuses on the *how* and *why* of experimental methods while the second half (1,5 hours) will involve a lab session for replication of existing research for learning coding skills. Every other week we will discuss an existing paper applying or advancing the causal inference method learned the previous week and we will critically evaluate this paper and discuss its identification strategy.

### Assessments

Continuous evaluation of class participation (25 %) and final project (75 %)

## Course Policies

### During Class

Please refrain from using computers for anything but activities related to the class.

### Attendance Policy

Attendance is expected in all lecture and lab sections. Valid excuses for absence will be accepted before class. In extenuating circumstances, valid excuses with proof will be accepted after class.

# Schedule and weekly learning goals

The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts you should grasp after each week, and also as a study guide.

**Week 1** What is causality? Introduction to the potential outcomes framework

- Learning to draw directed acyclic graphs (DAGs)
- Learning to use the potential outcome notation
- Understanding what confoundedness and selection-bias are and how to avoid them

**Week 2** Reading discussion

- Reading TBA

**Week 3** The experimental benchmark: Randomized Control Trials

- Learn what an optimal sample and randomization procedure is like
- Learn to draw inferences from experiments
- Learn the code for randomization inference and power calculations

**Week 4** Reading discussion

- Reading TBA

**Week 5** Instrumental Variables

- Learn how instruments can make up for imperfect compliance in randomized trials
- Learn how instruments can be used in observational data to generate quasi-randomization
- Learn to code instrumental variable regressions in R

**Week 6** Reading discussion

- Reading TBA

**Week 7** Matching

- Understand how matching uses back-door adjustment to replicate the experimental condition
- Learn to carry out propensity-score matching, distance-based matching and genetic matching in R
- Understand the assumptions and pitfalls of matching

**Week 8** Reading discussion

- Reading TBA

**Week 9** Regression discontinuity designs

- Learn to exploit artificial cut-offs to generate quasi-experiments
- Learn the difference between sharp and fuzzy RDDs
- Learn to run RDD regressions in R

**Week 10** Reading discussion

- Reading TBA

**Week 11** Panel data and difference-in-differences

- Learn how pre-and post-dimensions of time can yield causal estimates
- Learn to carry out difference-in-differences analysis in R
- Understand the parallel trends assumptions and the use of fixed effects

**Week 12** Reading discussion

- Reading TBA