VRIJE UNIVERSITEIT AMSTERDAM

Statistical Methods for Causal Inference

SYLLABUS

VU Amsterdam Summer School

6-20 July 2024

Any general questions for the Summer School support team? Contact amsterdamsummerschool@vu.nl.

Do you have specific questions regarding the course? Contact
Sanchayan Banerjee: s.banerjee@vu.nl
Jack Fitzgerald: j.f.fitzgerald@vu.nl

VU
VRIJE
UNIVERSITEIT
AMSTERDAM

Course Details

Title	Statistical Methods for Causal Inference
Coordinator(s)	Dr. Sanchayan Banerjee, Jack Fitzgerald
Other lecturers	Marta Buso
Study credits	3 ECTS
Form(s) of tuition	Lectures and computer tutorials
Approximate contact hours	45
Approximate self-study hours	30

Teaching staff

Dr. Sanchayan Banerjee: Co-coordinator, primary lecturer on experimental design and RCTs
Jack Fitzgerald: Co-coordinator, Primary lecturer on quasi-experimental design
Marta Buso: Teaching assistant

Course description

There is great interest among students and practitioners today to understand the causal mechanisms underlying major events. Identifying cause-and-effect relationships is important for impact evaluation and effective policy design. Such identification can help us answer questions like: “What causes an economic downturn?”, “Does universal basic income reduce unemployment?” and “Does a carbon tax reduce greenhouse gas emissions?

However, identifying causal relationships using data is often error prone. Differentiating causality from simple correlation requires learning and applying sophisticated quantitative tools. The golden standard of identifying causal linkages relies on designing experiments, often through randomized control trials. But designing a randomized control trial is not always feasible or ethical. Moreover, some events might have already happened in the past, such as a financial crisis or a cyclone. How can one use observational data to analyze the causal effects of such events?

VU
VRIJE
UNIVERSITEIT
AMSTERDAM

This course provides a hands-on introduction to statistical methods for causal inference. Over two weeks, students are introduced to experimental and quasi-experimental methods which allow them to infer cause-and-effect relationships robustly. We teach these methods from both a theoretical and applied lens, supplementing lectures with hands-on computer tutorials in the R programming language to help students learn by doing.

Learning objectives

By the end of this course, students will be able to:

1. Understand the difference between correlation and causation.
2. Apply quantitative methods of statistical data analysis to infer causal relationships.
3. Identify confounding factors that threaten causal inference and hamper the internal and external validity of analytical findings.
4. Critically analyse data using statistical methods like experiments, matching analysis, difference-in-differences, regression discontinuity, and instrumental variables estimation.
5. Explore challenges and limitations in the use of quantitative methods of causal inference such as data availability, missing data, and measurement errors.
6. Apply diagnostic knowledge to inform impact evaluations and develop evidence-based policies.

Grading

Days 2-10 will begin with a short quiz covering the material from the previous lecture. The best 7/9 quiz scores will make up 20% of the final grade. You will submit a final research project that will be evaluated using a presentation (last day of the course, in-person) that will make up 20% of the final grade and a take-home assignment that will make up 60% of the final grade.

VRIJE
UNIVERSITEIT
AMSTERDAM

Recommended reading list

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

Cunningham, Scott. *Causal Inference: The Mixtape*. New Haven: Yale University Press, 2021.

Moffatt, Peter G. "Experimetrics: A Survey." *Foundations and Trends in Econometrics* 11, no. 1-2 (2021): 1-152. https://doi.org/10.1561/0800000035

Course Schedule

Day 1 (Monday July 8): Recap of Applied Statistical Methods

Morning Session: Sanchayan Banerjee
Lecture: Descriptive statistics, correlation and covariance, inferential statistics, ordinary least squares regression, controlling for variables, and inferential statistics

Afternoon Session: Marta Buso
Practical

Day 2 (Tuesday July 9): The Potential Outcomes Framework and RCTs

Morning Session: Sanchayan Banerjee
Quiz 1: Recap of Applied Statistical Methods
Lecture: Setting up an RCT, treatment effect estimation frameworks (ITT, LATE, ATT, ATE, CACE, TOT), review of parametric regression (i.e., linear and logistic models), correlation vs. causation, the identification problem, the SUTVA conditions, introduction to RCTs, and types of RCTs

Afternoon Session: Marta Buso (1400-1530)
Practical

Day 3 (Wednesday July 10): Power Analysis and Advanced Experimental Design

Morning Session: Sanchayan Banerjee
Quiz 2: The Potential Outcomes Framework

Lecture: Introduction to meta-science, statistical power, multiple hypothesis testing, heterogeneous treatment effects, joint hypothesis testing, and Monte Carlo simulation for power analysis

Day 4 (Thursday July 11): Practical Issues in Experiments

Morning Session: Sanchayan Banerjee
Quiz 3: Power Analysis and Advanced Experimental Design
Lecture: Instrumental variables in RCTs, the bad control problem, incentive compatibility, field experiments, online experiments, screeners, and attention checks

Afternoon Session: Marta Buso
Practical

Day 5 (Friday July 12): Matching and Weighting

Morning Session: Sanchayan Banerjee
Quiz 4: Practical Issues in Experiments
Lecture: Introduction to matching, propensity score matching and assumptions, distance-based matching methods, and algorithmic matching methods

Afternoon Session: Marta Buso
Practical

Day 6 (Monday July 15): Difference-in-Differences

Morning Session: Jack Fitzgerald
Quiz 5: Matching and Weighting
Lecture: General framework, the parallel trends assumption and testing procedures, placebo testing, staggered adoption methods, and extensions

Afternoon Session: Marta Buso
Practical

Day 7 (Tuesday July 16): Instrumental Variables Estimation

Morning Session: Jack Fitzgerald
Quiz 6: DID
Lecture: Two-stage least squares estimation, local average treatment effects, the exclusion restriction, weak instruments, efficiency, the Hausman test, control specifications, and shift-share design

Afternoon Session: Marta Buso
Practical

Day 8 (Wednesday July 17): Regression Discontinuity Design

Morning Session: Jack Fitzgerald
Quiz 7: IV
Lecture: Running variables, sharp regression discontinuity, nonparametric kernels, optimal bandwidth selection, and fuzzy regression discontinuity, multiple cutoffs and/or scores, and regression kink designs

Afternoon Session: Marta Buso
Practical

Day 9 (Thursday July 18): Synthetic Control

Morning Session: Jack Fitzgerald
Quiz 8: RDD
Lecture: Standard synthetic control, cross-validation techniques, (ridge) augmented synthetic control, staggered synthetic control, and generalized synthetic control

Afternoon Session: Marta Buso
Practical

Day 10 (Friday July 19): Research Presentations

Morning Session: Sanchayan Banerjee & Jack Fitzergald
Quiz 9: SCM
Lecture: Student research presentations