VRIJE UNIVERSITEIT AMSTERDAM

Statistical Methods for Causal Inference

SYLLABUS

VU Amsterdam Summer School

6-17 July 2026

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	N/a
Form(s) of tuition	Lectures and computer tutorials
Approximate contact hours	45
Approximate self-study hours	30

Teaching staff

Dr. Sanchayan Banerjee: Coordinator, Primary lecturer on experimental design and RCTs
Jack Fitzgerald: Co-coordinator, Primary lecturer on quasi-experimental design

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?

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

You will submit a research project that will be evaluated using a presentation (last day of the course, in-person) that will make up 40% of the final grade and the written report (joint work in a team of 2) that will make up 60% of the final grade.

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 6): Recap of Applied Statistical Methods

Morning Session: Sanchayan Banerjee (1000-1300)
Lecture: Basics of regression analysis – linear regression methods (OLS assumptions, simple and multivariate linear regression, interactions, t-tests, F-tests, backward and forward model selection, logistic regression methods, Wald test).

Afternoon Session: Sanchayan Banerjee (1400-1500)
Computer tutorial: Recap of applied regression analysis

Day 2 (Tuesday July 7): The Potential Outcomes Framework

Morning Session: Sanchayan Banerjee (1000-1300)
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: Sanchayan Banerjee (1400-1600)
Computer tutorial: Estimating treatment effects from a randomised controlled trial

Day 3 (Wednesday July 8): Power Analysis and Advanced Experimental Design

Morning Session: Sanchayan Banerjee (1000-1300)
Lecture: Introduction to meta-science, statistical power, multiple hypothesis testing, heterogeneous treatment effects, joint hypothesis testing, and Monte Carlo simulation for power analysis

Afternoon Session: N/A
VUASS Cultural Visits

Day 4 (Thursday July 9): Practical Issues in Experiments

Morning Session: Sanchayan Banerjee (1000-1300)
Lecture: Instrumental variables in RCTs, the bad control problem, incentive compatibility, field experiments, online experiments, screeners, and attention checks

Afternoon Session: Sanchayan Banerjee (1400-1600)
Computer tutorial: Power analysis, data management plans and pre-analysis plans

Day 5 (Friday July 10): Matching and weighting

Morning Session: Sanchayan Banerjee (1000-1300)
Lecture: Introduction to matching, propensity score matching and assumptions, distance-based matching methods, and algorithmic matching methods

Afternoon Session: Sanchayan Banerjee (1400-1600)
Computer tutorial: Using statistical matching methods

Day 6 (Monday July 13): Difference-in-Differences

Morning Session: Jack Fitzgerald (1000-1300)
Lecture: General framework, the parallel trends assumption and testing procedures, placebo testing, staggered adoption methods, and extensions

Afternoon Session: Jack Fitzgerald (1400-1600)
Computer tutorial: DID

Day 7 (Tuesday July 14): Instrumental Variables Estimation

Morning Session: Jack Fitzgerald (1000-1300)
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: Jack Fitzgerald (1400-1600)
Computer tutorial: IV

Day 8 (Wednesday July 15): Regression Discontinuity Design

Morning Session: Jack Fitzgerald (1000-1300)
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: N/A
VUASS Cultural Visits

Day 9 (Thursday July 16): Synthetic Control

Morning Session: Jack Fitzgerald (0900-1200)
Lecture: Standard synthetic control, cross-validation techniques, (ridge) augmented synthetic control, staggered synthetic control, and generalized synthetic control

Afternoon Session: Jack Fitzgerald (1400-1600)
Computer tutorial: RDD & Synthetic Control

Day 10 (Friday July 17): Presentations

Morning Session: Sanchayan Banerjee & Jack Fitzgerald (0900-1200
Presentations pitch (15 mins per team of 2 including Q&A)

Afternoon Session: Jack Fitzgerald (1300-1500)
Time to work on your own projects