Hertie School Spring Semester 2025

Course Syllabus

GRAD-C6-2001: Statistics II: Statistical Modeling and Causal Inference with R

1. General information

Course Format	Onsite
Instructors	Asya Magazinnik Maria Uttenthal
Instructors' e-mail	a.magazinnik@hertie-school.org m.uttenthal@hertie-school.org
Faculty Assistant Teaching Assistants	Amanda Slater; slater@hertie-school.org Sebastian Ruiz (ramirez-ruiz@hertie-school.org) (Head TA) Sofia Garcia-Durer (s.garcia-durrer@students.hertie-school.org) William Fernandez (w.fernandez-tinoco@phd.hertie-school.org) Carolina Diaz (a.suarez@hertie-school.org)
Instructors' Office Hours	**Asya Magazinnik:** Please email Amanda Slater to make a 30-minute appointment, Thursdays 9-12 **Maria Uttenthal:** Drop-in office hours Wednesday, 16:30-17:30, or email Maria to make an appointment

Link to Study, Examination and Admission Rules and MIA, MDS and MPP Module Handbooks

For information on **course room, times and session dates,** please consult the Course Plan on *MyStudies*.

Instructor Information:

Asya Magazinnik is the Professor of Social Data Science at the Hertie School. Her research interests include electoral geography, federalism, local politics, and law enforcement. She also works on political methodology, in particular at the intersection of causal inference and formal theory. Her work has appeared in the *American Journal of Political Science*, the *Journal of Politics*, and other outlets. Previously, she was an Assistant Professor of Political Science at MIT. She earned a PhD in Political Science from Princeton University in 2020 and holds an MPP from the Harris School at the University of Chicago.

Maria Uttenthal is a Lecturer and Academic Counsellor at the Hertie School. Her research focuses on building closer connections between quantitative methods and systematic conceptual work in the social sciences. She brings extensive teaching experience within Statistics and Research Design, including as a Lecturer of Statistics at Bard College Berlin. At the Hertie School, Maria also teaches and counsels students on research skills for their master's theses. She obtained a PhD from the Hertie School in 2023, and further holds an MSc in European Public Policy from University College London.

2. Course Contents and Learning Objectives

Course contents:

This course continues the sequence in statistical modeling, helping students build a more diverse and sophisticated data analysis toolkit. In the first half of the course, we introduce a new perspective on studying causes and effects in social science and present various strategies to uncover causal relationships using statistical tools, from conditioning on observables (regression and matching) to the use of natural experiments (instrumental variables and regression discontinuity). The second half of the course covers other essential tools for analyzing data from the social world, including strategies for measuring theoretical constructs, analyzing time series and categorical data, and creating beautiful and informative data visualizations. All lectures divide time between theory and application, while labs and assignments emphasize hands-on data analysis with R.

Main learning objectives:

The goals are to (1) acquaint you with some of the most common statistical methods used to demonstrate causality, (2) expand your statistical toolkit to analyse and interpret different data types, including categorical and panel data (3) enable you to implement these new methods with statistical software, and (4) prepare you for Hertie's methods electives.

Software:

We will work with R, RStudio and RMarkdown to implement and practice the new techniques. It is assumed that you have some basic knowledge of these tools from Statistics I. If not, we strongly encourage you to familiarize yourself with them prior to the course so as to be able to focus on the substance. We suggest the following resources for self-guided learning:

R and RStudio:

- https://education.rstudio.com/learn/beginner/
- https://posit.co/resources/videos/a-gentle-introduction-to-tidy-statistics-in-r/
- https://r4ds.had.co.nz/

RMarkdown:

- https://rmarkdown.rstudio.com/lesson-1.html

Teaching style:

Each session will consist of an in-person lecture, which, alongside the course readings, serves to familiarize you with the session's topic. To become acquainted with R and to learn how to implement various statistical techniques in practice, you will be enrolled in a small-group support lab session taught weekly by a teaching assistant.

Prerequisites:

Statistics I, basic command of R

Diversity Statement:

In this classroom and at the Hertie School more broadly, we respect and value differences in neurodivergence, race, gender, ethnicity, age, physical and language abilities, culture, religion, sexual orientation, and other identities. Such differences contribute to the strength of our community, and upholding an atmosphere of inclusion and mutual respect is of the utmost importance.

This class will also contain a diversity of backgrounds and experiences with mathematics, coding, and statistics. It is everyone's responsibility to promote and maintain an environment where all students feel comfortable asking questions and voicing their struggle. Mutually supporting one another will be at the heart of everyone's success in this class, and we hope that more advanced students will help their peers.

Students are encouraged to bring up any concerns relating to these policies, or any other hardships that they encounter, with us as early as possible, and we will make every effort to resolve them.

3. Grading and Assignments

Composition of Final Grade:

Assignment 1: 4 take-home problem sets	Deadline: about a week after publication	Submit via Moodle	45% (15% each for the best 3 submissions)
Assignment 2: Midterm exam	Deadline: TBA (tentatively week of March 24)	On campus	25%
Assignment 3: Final exam	Deadline: TBA (week of May 19)	On campus	30%

Assignment Details

Assignment 1

This assignment comprises 4 take-home problem sets evenly spaced throughout the semester. Problem sets emphasize hands-on learning and application of the tools of the course to real-world data. The assignment will have to be submitted as a knitted R Markdown file containing both R code and your answers to the questions.

The problem sets will be graded by awarding 15% each to the best 3 submissions. That means that one of the problem sets is optional: students are welcome not to submit it. Students are free to choose which of the four problem sets they would like to skip. If students wish to submit all four, they will receive grades on all four, and the best three grades will be counted toward their final course grade.

Assignment 2

There will be a midterm exam taking place roughly half-way through the semester (date TBA). The exam will cover the mathematical foundations and theoretical components of the course. The exam may include some pen and paper computations or proofs as well as open-ended discussion questions related to research design.

Assignment 3

There will be a final exam during exam week, similar to the midterm above. The final will cover all topics from throughout the course, with greater emphasis on the second half.

Late submission of assignments: For each day an assignment is late, the grade will be reduced by 10% (e.g. submission two days later would result in 20% grade deduction).

Attendance: Students are expected to be present and prepared for every lecture and lab session. However, we will allow at most 2 absences each from lecture and lab. Please do not inform us that you are taking these two absences; simply take them, no questions asked. If you need to miss a lecture or lab more than two times in the semester, please email Amanda Slater providing the reason for your absence. For further information please consult the Examination Rules §10.

Academic Integrity: The Hertie School is committed to the standards of good academic and ethical conduct. Any violation of these standards shall be subject to disciplinary action. Plagiarism, misuse of AI, free riding in group work, and other deceitful actions are not tolerated. See Examination Rules §16, the Hertie Plagiarism Policy, and the Hertie Guidelines for Artificial Intelligence Tools.

Compensation for Disadvantages: If a student furnishes evidence that they are not able to take an examination as required in whole or in part due to disability or permanent illness, the Examination Committee may upon written request approve learning accommodation(s). The submission of adequate certificates may be required. See Examination Rules §14.

Extenuating circumstances: An extension can be granted due to extenuating circumstances (i.e., for reasons like illness, personal loss or hardship, or caring duties). In such cases, please contact the course instructor and Examination Office *in advance* of the assignment deadline.

4. Course Sessions and Readings

Below you find the required and suggested readings for each week. Articles listed under **Application Reading** serve as examples during the lecture and are also mandatory, but you are not required to read every page closely. It is a good skill to develop to read articles quickly to grasp the key elements: research question, research design, results, and implications.

All course readings can be accessed on the course Moodle page.

Session 1: Counterfactual Causality	
Required Reading	**The Book of Why** (Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books): **Introduction**

Session 2: The Potential Outcomes Framework and Experiments	
Required Reading	**Mastering Metrics** (Angrist, J. D., & Pischke, J.-S. (2015). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton: Princeton University Press)**: Chapter 1**

Session 3: Causal Graphs	
Required Reading	**The Mixtape** (Cunningham, S. (2021). *Causal Inference: The Mixtape*. New Haven: Yale University Press)**: Chapter 3**
Application Reading	Knox, Dean, Will Lowe, and Jonathan Mummolo. 2020. "Administrative Records Mask Racially Biased Policing." *American Political Science Review* 114(3): 619-637.
Suggested Reading	**The Book of Why:** Chapter 4

Session 4: Conditioning on Observables: Regression and Matching	
Required Reading	**The Mixtape**: **Chapter 5**
Application Reading	Gilligan, Michael J. and Ernest J. Sergenti. 2008. "Do UN Interventions Cause Peace? Using Matching to Improve Causal Inference." *Quarterly Journal of Political Science* 3: 89-122.

Session 5: Natural Experiments I: Instrumental Variables	
Required Reading	**Mastering Metrics: Chapter 3**
Application Reading	Miguel, Edward, Shanker Satyanath and Ernest Sergenti. 2004. "Economic Shocks and Civil Conflict: An Instrumental Variables Approach." *Journal of Political Economy* 112(4): 725-753. Wasow, Omar. (2020). "Agenda Seeding: How 1960s Black Protests Moved Elites, Public Opinion, and Voting." *American Political Science Review* 114(3): 638-659.

Session 6: Natural Experiments II: Regression Discontinuity	
Required Reading	**Mastering Metrics: Chapter 4**
Application Reading	Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics* 142: 675–97. Caughey, Devin and Jasjeet S. Sekhon. 2011. "Elections and the Regression Discontinuity Design: Lessons from Close U.S. House Races, 1942-2008." *Political Analysis* 19: 385-408.

Session 7: Time I: Panel Data and Fixed Effects	
Required Reading	**Introductory Econometrics** (Wooldridge, J. M. (2020b). *Introductory econometrics: A modern approach* (Seventh edition). Cengage Learning). **Chapter 13.1** **The Effect** (Huntington-Klein, N. (2022). *The effect: An introduction to research design and causality*. CRC Press, Taylor & Francis Group). **Chapter 16**
Application Reading	O'Grady, T. (2019). "How Do Economic Circumstances Determine Preferences? Evidence from Long-Run Panel Data." *British Journal of Political Science* 49(4): 1381–1406.

Session 8: Time II: Difference-in-Differences	
Required Readings	**The Effect: Chapter 18**
Application Reading	Card, D., & Krueger, A. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, *84*(4), 772–793

Session 9: Nonlinear Regressions	
Required Reading	**Introduction to Econometrics** (Stock, J. H., & Watson, M. W. (2020). *Introduction to econometrics* (Fourth edition, global edition). Pearson): **Chapter 8**
Application Reading	Hakhverdian, A., & Mayne, Q. (2012). Institutional Trust, Education, and Corruption: A Micro-Macro Interactive Approach. *The Journal of Politics*, *74*(3), 739–750.

Session 10: Categorical Data	
Required Reading	**Introduction to Econometrics: Chapter 11**
Application Reading	Delhey, J., & Newton, K. (2003). Who trusts?: The origins of social trust in seven societies. *European Societies*, *5*(2), 93–137.

Session 11: Measurement	
Required Reading	**Latent Variable Modeling with R** (Finch, W. H. & French, B.F. (2015). *Latent Variable Modeling with R*. New York: Routledge)**: Chapters 2 and 3**
Application Reading	OECD. (2017). *OECD Guidelines on Measuring Trust*. OECD: Chapter 2

Session 12: Data Visualization	
Required Reading	**Data Visualization** (Healy, K. (2019). *Data Visualization: A Practical Introduction.* Princeton: Princeton University Press)**: Chapters 1 and 3**
Suggested Reading	The Visual Display of Quantitative Information