

GRAD-C6-2001

Statistics II: Statistical Modeling and Causal Inference (with R)

Simon Munzert

1. General Information

Class Time	<i>3 hours per session!</i> Group A: Mondays, 9-12h Group B: Mondays, 13-16h Group C: Tuesdays, 16-19h
Venue	1.61 (Computer Lab)
Instructor	Dr. Simon Munzert
Instructor's office	3.13.1
Instructor's email	munzert@hertie-school.org
Instructor's phone number	+49 (0)30 259 219 450
Assistant	Bernadette Boddin boddin@hertie-school.org +49 (0)30 259 219 173 245
Office Hours	Tuesdays, 2-3pm (appointment by email required)
Teaching Assistant	Gabriel Tarriba
TA's email	g.tarriba@phd.hertie-school.org
TA drop-in sessions	Thursdays, 10-12h and 14-16h
Venue for TA drop-in sessions	1.61 (Computer Lab)

Instructor Information:

Simon Munzert is Lecturer in Political Data Science at the Hertie School of Governance. He received his doctoral degree in Political Science from the University of Konstanz. His research interests include measuring and forecasting public opinion, political representation, and the use of new media in society.

Gabriel Tarriba is a PhD candidate at the Hertie School of Governance. His dissertation focuses on the educational attainment of the children of immigrants in Western Europe.

2. Course Contents and Learning Objectives

Course contents:

This course continues the sequence in statistical modeling. Assuming prior knowledge in simple and multiple linear regression modelling, it introduces students to a new perspective on studying causes and effects in social science research. Based on a framework of causality, the course agenda covers various strategies to uncover causal relationships using statistical tools. We start with reflecting about causality, the ideal research design, and then learn to use a framework to study causal effects. Then, we revisit common regression estimators of causal effects and learn about their limits. Next, we will focus on matching and weighting, instrumental variables, panel and difference-in-differences

estimators, regression discontinuity designs, and techniques to explore moderated and mediated relationships. All classes take place in the computer lab and divide time between theory and application. Students are assigned a problem set at the end of each class covering that day's materials.

Main learning objectives:

The goals are to (1) acquaint you with some of the most common statistical methods, (2) enable you to implement these with statistical software, and (3) prepare you for our methods electives.

Software:

We will work with R to implement and practice the learned techniques. I assume you have some basic knowledge in using R from Statistics I. If not, resources to learn how to use R will be made available prior to the course.

Target group:

MPP 1st year students in the Policy Analysis track only

Prerequisites:

Statistics I, basic knowledge of R

3. Grading and Assignments

Evaluation is conducted via a combination of one replication project (counts toward 40% of your final grade), one in-class final exam (30%), and a series of weekly assignments that are graded for effort (30%). The replication project is based on a couple of research papers provided by the instructor; the student can choose one of the papers and replicate the analysis. The replication project itself is brief (maximum 7 pages) but bear in mind that coding takes considerable time. None of the assignments in this class are collaborative per se. However, I encourage you to study together and learn to use the software together.

Composition of Final Grade:

Assignment	Deadline/Date	Final grade
Series of weekly assignments	Sundays, 9.00pm	40%
In-class final exam	24.04.2019	30%
Replication project	10.05.2019	30%

Late submission of assignments:

For each day the assignment is turned in late, the grade will be reduced by 10% (e.g. submission two days after the deadline would result in 20% grade deduction).

Attendance: Students are expected to be present and prepared for every class session. Active participation during lectures and seminar discussions is essential. If unavoidable circumstances arise which prevent attendance or preparation, the instructor should be advised by email with as much advance notice as possible. Please note that students cannot miss more than one session. For further information please consult the Examination Rules §9.

Academic Integrity: The Hertie School of Governance is committed to the standards of good academic and ethical conduct. Any violation of these standards shall be subject to disciplinary action. Plagiarism, deceitful actions as well as free-riding in group work are not tolerated. See Examination Rules §15.

4. General Readings

In order to get attuned to the spirit of the class, please read the following book before the first session:

1. Pearl, Judea and Dana Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

The remainder of the course is built on the following two textbooks:

1. Morgan, Stephen L., and Christopher Winship (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research, Second Edition*. Cambridge University Press.
2. Angrist, Joshua D. and Joern-Steffen Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press.

Angrist/Pischke is a bit more difficult mathematically, while Morgan/Winship appeals more to intuition. Both complement each other well. **Important note:** The second edition of the Morgan/Winship (2014) book has been substantively revised. Be sure to get that one (with the blue cover) and not the first edition (with the red cover). The textbooks can be borrowed from the library.

In addition, there will be journal articles to read which provide illustrations and more background. Articles that are listed under “Further Reading” are optional, although they will be discussed in the lecture. Furthermore, there is one application paper listed for each session, which will serve as an example case in the lecture. Please skim this paper to gain familiarity with the topic in advance of each session.

This course has a clear focus on statistics as a tool for causal effects, that is, causal identification and inference. There are others books that focus more on statistical modeling and estimation. If you are interested in these topics, you might want to check out the following:

- Andrew Gelman and Jennifer Hill (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. (very accessible introduction to regression modelling, multilevel modelling, and applied Bayesian modelling)
- Gary King (1998). *Unifying Political Methodology. The Likelihood Theory of Statistical Inference*. University of Michigan Press. (conceptual framework for ML estimation and applications using various models)
- Jeffrey M. Wooldridge (2010). *Introductory Econometrics: A Modern Approach*. 4th ed. South-Western College Publishers. (a classic – many equations and examples from econ)
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman (2010). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. 2nd ed. Springer. (data science perspective on problems of classification using machine learning methods)
- James E. Monogan III (2015). *Political Analysis Using R*. Springer. (low-level introduction to common statistical models – more of an R manual with many examples)

5. Session Overview

Session	Session Date	Session Title
1	18./19.02.2019	Overview and introduction

2	25./26.02.2019	Causes and effects
3	04./05.03.2019	Revisiting regression estimators of causal effects
4	11./12.03.2019	Matching and weighting
Mid-term Exam Week: 18.-22. March 2019 – no class		
5	25./26.03.2019	Instrumental variables
6	01./02.04.2019	Panel and difference-in-differences estimators
7	08./09.04.2019	Regression discontinuity designs
8	15./16.04.2019	Causal explanation, moderators, and mediators
	23.04.2019	In-class final exam
	10.05.2019	Replication project due at 11.59p.m.

6. Course Sessions and Readings

Session 1: 18./19.02.2019 - Overview and introduction to counterfactual causality

Basic Reading	Pearl and Mackenzie (2018)
----------------------	----------------------------

Session 2: 25./26.02.2019 – Causes and effects

Basic Reading	Angrist and Pischke (2009), Ch. 1 + 2 Morgan and Winship (2014), Ch. 1 + 2 + 3
----------------------	---

Session 3: 04./05.03.2019 - Revisiting regression estimators of causal effects

Basic Reading	Angrist and Pischke (2009), Ch. 3.1, 3.2 Morgan and Winship (2014), Ch. 4, 6
Further Reading	Freedman, D. A. (1991). Statistical models and shoe leather. <i>Sociological Methodology</i> 21, 291–313.
Application Paper	Aronow, P. M. and C. Samii (2016). Does regression produce representative estimates of causal effects? <i>American Journal of Political Science</i> 60(1), 250–267.

Session 4: 11./12.03.2019 – Matching and weighting

Basic Reading	Angrist and Pischke (2009), Ch. 3.3, 3.4 Morgan and Winship (2014), Ch. 5, 7
Further Reading	Hainmueller, J. (2011). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. <i>Political Analysis</i> 17(4), 400–417.

	<p>Ho, D. E., K. Imai, G. King, and E. A. Stuart. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. <i>Political Analysis</i> 15(3), 199–236.</p> <p>Iacus, S. M., G. King, and G. Porro. (2011). Causal inference without balance checking: coarsened exact matching. <i>Political Analysis</i> 19(4), 1–24.</p> <p>King, G. and R. Nielsen. (2018). Why Propensity Scores Should Not Be Used for Matching. <i>Political Analysis</i>. 1–34.</p> <p>Iacus, S., King, G., & Porro, G. (n.d.). A Theory of Statistical Inference for Matching Methods in Causal Research. <i>Political Analysis</i>, 1–23.</p>
Application Paper	<p>Boyd, C. L., Epstein, L., & Martin, A. D. (2010). Untangling the causal effects of sex on judging. <i>American Journal of Political Science</i>, 54(2), 389–411.</p>

Session 5: 25./26.03.2019 - Instrumental variables

Basic Reading	<p>Angrist and Pischke (2009), Ch. 4</p> <p>Morgan and Winship (2014), Ch. 9</p>
Further Reading	<p>Bazzi, S. and M. Clemens. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. <i>American Economic Journal: Macroeconomics</i> 5(2), 152–186.</p> <p>Bound, J., D. A. Jaeger, and R. M. Baker. (1995). Problems with instrumental variables estimation when correlation between the instruments and the endogenous explanatory variable is weak. <i>Journal of the American Statistical Association</i> 90(430), 443–450.</p> <p>Sovey, A. J. and D. P. Green. (2011). Instrumental variables estimation in political science: A reader's guide. <i>American Journal of Political Science</i> 55(1), 188–200.</p> <p>Yamamoto, T. (2012). Understanding the Past: Statistical Analysis of Causal Attribution. <i>American Journal of Political Science</i> 56: 237–256.</p>
Application Paper	<p>Kern, H.L. and J. Hainmueller. (2009). Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes. <i>Political Analysis</i> 17(4): 377–399.</p>

Session 6: 01./02.04.2019 – Panel and difference-in-differences estimators

Basic Reading	<p>Angrist and Pischke (2009), Ch. 5</p> <p>Morgan and Winship (2014), Ch. 11</p>
Further Reading	<p>Abadie, A. , Diamond, A. and Hainmueller, J. (2015), Comparative Politics and the Synthetic Control Method. <i>American Journal of Political Science</i>, 59: 495–510.</p> <p>Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. <i>Journal of the American Statistical Association</i> 105(490), 493–505.</p>

Application Paper	Selb, P. and S. Munzert. (2018). Examining a Most Likely Case for Strong Campaign Effects: Hitler's Speeches and the Rise of the Nazi Party, 1927–1933. <i>American Political Science Review</i> 112(4): 1050–1066.
--------------------------	---

Session 7: 08./09.04.2019 - Regression discontinuity designs

Basic Reading	Angrist and Pischke (2009), Ch. 6
Further Reading	Green, D. P., T. Y. Leong, H. L. Kern, A. S. Gerber, and C. W. Larimer. (2009). Testing the Accuracy of Regression Discontinuity Analysis Using Experimental Benchmarks. <i>Political Analysis</i> 17(4): 400–417. Eggers, A. C., Fowler, A., J. Hainmueller, A.B. Hall, and J.M. Snyder. (2015). On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races. <i>American Journal of Political Science</i> 59: 259-274.
Application Paper	Dahlgard, Jens Olav. (2018). Trickle-Up Political Socialization: The Causal Effect on Turnout of Parenting a Newly Enfranchised Voter. <i>American Political Science Review</i> 112(3): 698–705.

Session 8: 15./16.04.2019 – Causal explanation, moderators, and mediators

Basic Reading	Morgan and Winship (2014), Ch. 10
Further Reading	Acharya, A., M. Blackwell, and M. Sen. (2016). Explaining causal findings without bias: Detecting and assessing direct effects. <i>American Political Science Review</i> 110(3): 512–29. Bullock, J. G., D. P. Green, and S. E. Ha. (2010). Yes, but what's the mechanism? (Don't expect an easy answer). <i>Journal of Personality and Social Psychology</i> 98(4), 550–558. Imai, K. L. Keele, D. Tingley, and T. Yamamoto. (2011). Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. <i>American Political Science Review</i> 105(4): 765–89.
Application Paper	Angrist, J.D., P.A. Pathak, and C.R. Walters. (2013). Explaining Charter School Effectiveness. <i>American Economic Journal: Applied Economics</i> 5(4): 1-27.