Hertie School

Master of Public Policy
Spring Semester 2020

Course Syllabus, Version 24.01.2020

GRAD-C6-2001
Statistics II: Statistical Modeling and Causal Inference (with R)
Concentration: Policy Analysis

Simon Munzert, Lukas Stoetzer

1. General Information

Class Time	*3 hours per session!* **Group A (Munzert):** Mondays, 9-12h, R 3.61 **Group B (Munzert):** Mondays, 13-16h, R 3.61 **Group C (Munzert):** Tuesdays, 9-12h, R 3.61 **Group D (Stoetzer)**: Mondays, 9-12h, R 2.32 **Group E (Stoetzer)**: Mondays, 13-16h, R 2.32
Instructor	Prof. Simon Munzert Dr. Lukas Stoetzer
Instructor's office	3.13.1 (Munzert) Georgenstrasse 23 SCRIPTS 7th floor (Stoetzer)
Instructor's email	munzert@hertie-school.org lukas.stoetzer@hu-berlin.de
Instructor's phone number	+49 (0)30 259 219 450 (Munzert) +49 (0)30 2093 70234 (Stoetzer)
Assistants	Ayamba Kwoyila kwoyila@hertie-school.org +49 (0)30 259 219 121 3.59 Alex Karras karras@hertie-school.org +49 (0)30 259 219 156 2.45
Office Hours	Munzert: Tuesdays, 2-3pm Stoetzer: Mondays, 5-6pm (appointment by email required)
Teaching Assistants	Marina Wyss (Lab Sessions 1 + 2) Sebastian Ramirez Ruiz (Lab Sessions 3 + 4)
TA's email	m.wyss@mpp.hertie-school.org s.ramirez-ruiz@mpp.hertie-school.org
TA drop-in sessions	Wednesdays, 10-12h (Marina Wyss), 1.61 Thursdays, 10-12h (Marina Wyss), 1.61 Thursdays, 8-10h, (Sebastian Ramirez Ruiz), 3.61 Thursdays, 12-14h, (Sebastian Ramirez Ruiz), 1.61

Link to Module Handbook **MIA** and **MPP**
Link to **Study, Examination and Admission Rules**

Instructor Information:
Simon Munzert is Assistant Professor of Data Science and Public Policy at the Hertie School and part of the Hertie School Data Science Lab. His research interests include opinion formation in the digital age, public opinion, and the use of online data in social research. He is the principal investigator of an international cooperation project funded by the VolkswagenStiftung entitled "Paying Attention to Attention: Media Exposure and Opinion Formation in an Age of Information Overload". He received his Doctoral Degree in Political Science from the University of Konstanz.

Lukas F. Stoetzer is a Researcher at the SCRIPTS Cluster of Excellence at Humboldt University of Berlin, Germany. He graduated from the University of Mannheim at the Graduate School of Economic & Social Sciences in May 2015. His research focuses on comparative political behavior, public opinion as well as political methodology. Current research projects include pre-electoral coalition politics, the analysis of legislative behavior in parliament, competition between political parties in multi-party systems and the effect of political arguments on voter's decision making.

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, instrumental variables, difference-in-differences and fixed effects estimators, regression discontinuity designs, and techniques to explore moderated and mediated relationships. All classes 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

Teaching style:
Each session will start with a review of the assignments, followed by an interactive lecture on the session’s topic led by the instructor. From time to time, the implementation of statistical models with R will be shown and practiced.

Prerequisites:
Statistics I, basic knowledge of R

3. Grading and Assignments

Composition of Final Grade:

Assignment	Deadline/Date	Submit via	Weight for final grade
Series of weekly assignments	Sundays, 11.59pm (assignment) Tuesdays, 11.59pm (reviews)	GitHub	40%
In-class final exam	TBC – Final Exam Week	In class	30%
Replication project	27.04.2020, 11.59pm	GitHub	30%

Evaluation is conducted via a combination of one replication project (counts toward 30% of your final grade), one in-class final exam (30%), and a series of weekly assignments (40%). 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. We generally encourage you to study and learn to use the software together.

Late submission of assignments: For each hour the assignment is turned in late, the grade will be reduced by 10% (e.g. submission two hours 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 two out of 12 course sessions. 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, deceitful actions as well as free-riding in group work are not tolerated. See Examination Rules §16.

Compensation for Disadvantages: If a student furnishes evidence that he or she is 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). In this respect, the submission of adequate certificates may be required. See Examination Rules §14.

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 textbook:

.

2. Angrist, Joshua D. and Joern-Steffen Pischke (2014). *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

In addition, there will be selected chapters from other books and 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.

In a previous iteration of the course, the following textbooks were used. They are useful but not as accessible as *Mastering 'Metrics*:

- Morgan, Stephen L., and Christopher Winship (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research, Second Edition*. Cambridge University Press. (comprehensive, lengthy treatment of much of the content that is covered in this class; covers both the potential outcomes framework and causal graphs)
- Angrist, Joshua D. and Joern-Steffen Pischke (2009). *Mostly Harmless Econometrics*. Princeton University Press. (the not so harmless, more predecessor of *Mastering 'Metrics*)

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	**03./04.02.2020**	**Overview and introduction**
2	**10./11.02.2020**	**Causes and effects**
3	**17./18.02.2020**	**Revisiting regression estimators of causal effects**
4	**24./25.02.2020**	**Matching (and a quick primer to logit and probit)**
5	**02./03.03.2020**	**Instrumental variables**
6	**09./10.03.2020**	**Regression discontinuity designs**

Mid-term Exam Week: 16.03.2020—20.03.2020 – no class		
7	**23./24.03.2020**	**Panel data, difference-in-differences, fixed effects**
8	**30./31.03.2020**	**Causal explanation, moderators and mediators**
	06.04.2020	**(Optional) Bonus session**
	Date TBC	**In-class final exam**
	27.04.2020	**Replication project due at 11.59p.m.**

6. Course Sessions and Readings

Session 1: 03./04.02.2020 - Overview and introduction to counterfactual causality	
Basic Reading	Pearl and Mackenzie (2018), Introduction, Ch 1, Ch 4
Learning Objectives	Why causation matters in social science research and beyond. The ladder of causation. Associations, interventions, and counterfactuals.

Session 2: 10./11.02.2020 – Causes and effects	
Basic Reading	Angrist and Pischke (2014), Ch. 1 Elwert, F. (2013). Graphical causal models. In *Handbook of causal analysis for social research* (pp. 245-273). Springer, Dordrecht.
Learning Objectives	From associations to causes. Potential outcomes framework. Biases in estimating treatment effects. Assumptions identifying causal effects. Causal graphs. Randomized experiments.

Session 3: 17./18.02.2020 – Revisiting regression estimators of causal effects	
Basic Reading	Angrist and Pischke (2014), Ch. 2
Further Readings	Freedman, D. A. (1991). Statistical models and shoe leather. *Sociological Methodology* 21, 291–313. Keele, L., R.T. Stevenson, and F. Elwert (2020). The causal interpretation of estimated associations in regression models. *Political Science Research and Methods* 8: 1-13.
Application Paper	Burden, Barry C., and Amber Wichowsky (2014). "Economic discontent as a mobilizer: unemployment and voter turnout." *The Journal of Politics* 76(4): 887-898.
Learning Objectives	OLS mechanics and estimation. Regression from a causal perspective. Variable selection in regression models using causal graphs.

Session 4: 24./25.02.2020 – Matching (and a quick primer to logit and probit)

Basic Reading	Morgan, S. L. and Harding, D. J. (2006). Matching estimators of causal effects: Prospects and pitfalls in theory and practice. *Sociological Methods & Research*, 35(1):3–60. Fox, John. (2015). *Applied regression analysis and generalized linear models*. Sage Publications. Chapter 14-15.
Further Readings	Hainmueller, J. (2011). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis* 17(4), 400–417. Ho, D. E., K. Imai, G. King, and E. A. Stuart. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis* 15(3), 199–236. Iacus, S. M., G. King, and G. Porro. (2011). Causal inference without balance checking: coarsened exact matching. *Political Analysis* 19(4), 1–24. King, G. and R. Nielsen. (2018). Why Propensity Scores Should Not Be Used for Matching. *Political Analysis*. 1–34. Iacus, S., King, G., & Porro, G. (n.d.). A Theory of Statistical Inference for Matching Methods in Causal Research. *Political Analysis*, 1–23.
Application Paper	Boyd, C. L., Epstein, L., & Martin, A. D. (2010). Untangling the causal effects of sex on judging. *American Journal of Political Science*, *54*(2), 389-411.
Learning Objectives	Exact matching. Non-exact matching. Logit and probit. Propensity scores.

Session 5: 02./03.03.2020 – Instrumental variables	
Basic Reading	Angrist and Pischke (2014), Ch. 3
Further Readings	Bazzi, S. and M. Clemens. (2013). Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth. *American Economic Journal: Macroeconomics* 5(2), 152–186. 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. *Journal of the American Statistical Association* 90(430), 443–450. Sovey, A. J. and D. P. Green. (2011). Instrumental variables estimation in political science: A reader’s guide. *American Journal of Political Science* 55(1), 188–200. Yamamoto, T. (2012). Understanding the Past: Statistical Analysis of Causal Attribution. *American Journal of Political Science* 56: 237-256. Marbach, Moritz and Dominik Hangartner. (2020). Profiling Compliers and Noncompliers for Instrumental-Variable Analysis. *Political Analysis* XX. 1—10.
Application Paper	Kern, H.L. and J. Hainmueller. (2009). Opium for the Masses: How Foreign Media Can Stabilize Authoritarian Regimes. *Political Analysis* 17(4): 377-399.

Learning Objectives	Noncompliance. Partial compliance. IV for experimental and observational research. Natural experiments.

Session 6: 09./10.03.2020 - Regression discontinuity designs	
Basic Reading	Angrist and Pischke (2014), Ch. 4.
Further Readings	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. *Political Analysis* 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. *American Journal of Political Science* 59: 259-274.
Application Paper	Dahlgaard, Jens Olav. (2018). Trickle-Up Political Socialization: The Causal Effect on Turnout of Parenting a Newly Enfranchised Voter. *American Political Science Review* 112(3): 698–705.
Learning Objectives	Sharp RDD. Fuzzy RDD. Model selection and diagnostics.

Mid-term Exam Week: 16.03 – 20.03.2020 – no class

Session 7: 23./24.03.2020 – Panel data, difference-in-differences, fixed effects	
Basic Reading	Angrist and Pischke (2014), Ch. 5. Imai, K. and Kim, I. S. (2019), When Should We Use Unit Fixed Effects Regression Models for Causal Inference with Longitudinal Data?. *American Journal of Political Science*, 63: 467-490.
Further Readings	Abadie, A. , Diamond, A. and Hainmueller, J. (2015), Comparative Politics and the Synthetic Control Method. *American Journal of Political Science*, 59: 495-510. Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association* 105(490), 493–505.
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. *American Political Science Review* 112(4): 1050–1066.
Learning Objectives	Panel data basics. DID, Fixed effects.

Session 8: 30./31.03.2020 – Causal explanation, moderators and mediators	
Basic Reading	Baron, Reuben M., and David A. Kenny. (1986). "The moderator–mediator variable distinction in social psychological research: Conceptual,

	strategic, and statistical considerations." *Journal of personality and social psychology* 51(6): 1173. Brambor, Thomas, William Roberts Clark, and Matt Golder. (2006). "Understanding interaction models: Improving empirical analyses." *Political Analysis* 14(1): 63-82. (Discussion of Interaction Effect Models)
Further Readings	Acharya, A., M. Blackwell, and M. Sen. (2016). Explaining causal findings without bias: Detecting and assessing direct effects. *American Political Science Review* 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). *Journal of Personality and Social Psychology* 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. *American Political Science Review* 105(4): 765–89. Tingley, Dustin, et al. "Mediation: R package for causal mediation analysis." (2014).
Application Paper	Gidengil, Elisabeth, Hanna Wass, and Maria Valaste. (2016). "Political socialization and voting: The parent–child link in turnout." *Political Research Quarterly* 69(2): 373-383.
Learning Objectives	Moderators vs. mediators. Interaction effects. The value and challenge of causal explanation. Hands-on mediation analysis.

Bonus session: 06.04.2020 – Don't go yet! There's so much we wanted to tell you…	
Content	This session is entirely optional and offered as a joint event for all groups. We want to conclude this course on a high note by giving you an overview of other topics from the statistical toolbox that we did not cover in this class. Furthermore, you will get a sneak preview of other possible data science electives offered at Hertie.