Introduction to causal inference

Michael Kühhirt
michael.kuehhirt@uni-koeln.de

Winter 2018/19

Office Hours: Tue 10:00-11:30am
Office: Greinstr. 2, 0.05
Web: goo.gl/WEZeoX

Class Hours: Mon 10:00-11:30am, Wed 5:45-7:15pm
Class Room: Sociolab, Greinstr. 2

Class content

Do negative campaign ads help politicians win elections? Are there learning benefits to regular class attendance at university? Is maternal employment good for children's development and, if so, why? If you are interested in answering such or similar questions, you are in the business of causal inference; your aim is to use empirical data to learn about causal relations between specific variables.

This class provides an introduction to the state-of-the-art theory and practice for causal inference. It goes beyond the notion that statistical association doesn't necessarily imply causation and specifies precisely under which conditions we can endow correlations with a causal interpretation. In addition to conventional regression approaches the class introduces inverse probability weighting as a useful statistical technique for causal inference. The main topics of the class are the construction and application of graphical causal models, the precise definition of causal effects in terms of interventions and counterfactuals, and the conditions and techniques to learn causal effects from data.

The focus is never on the mathematical derivation of statistical methods, but on an intuition for the conditions under which these methods allow valid causal inference, and for the scenarios under which they break down. Providing step-by-step practical guidance for (1) specifying prior knowledge and assumptions regarding the research question, for (2) defining the effect of interest, for (3) assessing whether this effect can be learned from available data, and for (4) statistical estimation (in `Stata`) and (5) interpretation of the results, this class will strengthen your ability to evaluate existing research, aid you in the formulation of precise and novel research questions, and provide tools for answering them. To facilitate learning, lectures are complemented with assignments and student collaboration.

Goals and learning outcomes

The class has three broad goals, which I enumerate below. Each goal is associated with a number of specific learning outcomes, which you should be able to perform as the class concludes. Students who complete the class should be able to

1. formalize a research question about causal relations using causal graphs and counterfactuals. This includes
    - translating theoretical arguments about causal relations into a corresponding graphical model,
    - specifying the causal relation of interest and defining this relation in terms of counterfactual contrasts,
    - being able to conceptually distinguish this causal relation from statistical association.
2. use graphical models to devise strategies for identifying the causal relation of interest. For this, students
    - demonstrate that they are capable to derive empirical implications from a graphical causal model,
    - understand the theoretical assumptions necessary to test these implications,
    - critically evaluate whether these assumptions hold in applied social research.
3. obtain an estimate approximating the causal relation of interest and, if feasible, test underlying assumptions. To do so, students
    - adapt existing `Stata` code for their purposes,
    - correctly interpret the resulting estimates,
    - understand and perform tests of the validity of the analyses.

As a bonus, I hope the class leads you to appreciate the ubiquity of causal inference in research, work, and daily live along with the recent advances in the concepts and tools to systematically undertake it. The class, by far, doesn't cover everything there is. But it should provide you with enough understanding to learn more on your own.

Requirements and grading

For successful class completion you're awarded 9 credit points. Please be aware that class completion requires a substantial time investment outside class. This includes required reading, assignments, and preparation. I highly recommend to regularly get together in small groups to discuss class contents and also to prepare comments and questions to be discussed in class.

The final grade is based on an exam to be completed on January 30, 2019. The exam assesses your progress in achieving the learning goals and outcomes outlined above. It includes tasks that directly test your knowledge of the material and others that test its practical application. You can achieve a maximum of 60 points in the exam.

By timely submission of 13 homework assignments and by contributing to online Q&A (see first set of slides for details) you can earn up to 15 activity points. Any activity points gained, are added to the points achieved in the exam. Final grades are based on the combination of exam points and activity points. A minimum of 30 points total is required for a passing grade. *Please note: Activity points are only counted until you reach the maximum number of 60 points in combination with the exam.*

You are encouraged to work on assignments in pairs or small groups. Nonetheless, each of you must submit your own work in the end. Submissions that are identical in part or in full are awarded 0 points in each case, no exceptions. *Joint work on the exam is not permitted.*

The final grade is awarded according to the following rules:

Grade	Points
1.0	60 ≥ P ≤ 58
1.3	58 > P ≤ 55
1.7	55 > P ≤ 51
2.0	51 > P ≤ 48
2.3	48 > P ≤ 45
2.7	45 > P ≤ 42
3.0	42 > P ≤ 39
3.3	39 > P ≤ 36
3.7	36 > P ≤ 33
4.0	33 > P ≤ 30
n.p	30 > P ≤ 0

Class schedule

The class, for the most part, builds on the following three references, which—in the course schedule below—will be abbreviated as ELW, HR and PGJ, respectively:

- Elwert, F. (2013). Graphical causal models. In: *Handbook of Causal Analysis for Social Research.* Ed. by Morgan, S. L. New York: Springer, pp. 245–272. = ELW
- Hernán, M. A. & Robins, J. M. (2018). *Causal Inference* Boca Raton, FL: Chapman & Hall/CRC, forthcoming. = HR.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer.* West Sussex, UK: Wiley = PGJ.

Introduction (Oct 8)
Key point: *Causal inference is important!*
Further reading: Petersen & van der Laan (2014)
Assignment 1 (due Oct 10, 5:44pm)

Data and correlations (Oct 10 & 15)
Key point: *Correlations describe group differences in the data (whereas causal effects capture changes in the world)!*
Required reading: PGJ, Preface + HR, Preface to Ch. 1 + Shalizi (2018), Ch. 1.1
Further reading: PGJ, Ch. 1.3 (before 1.3.10) + Shalizi (2018), Ch. 1.2
Assignment 2 (due Oct 15, 9:59am)

Modelling correlations (Oct 17 & 22)
Key point: *Statistical models (like regression) interpolate and extrapolate correlations in sparse data (but don't estimate causal effects)!*
Required reading: HR, Ch. 11
Further reading: PGJ, Ch. 1.3.10 + 1.3.11 + Shalizi (2018), Ch. 1.3-1.5 + 11
Assignment 3 (due Oct 22, 9:59am)

The limits of correlations (Oct 24 & 29)
Key point: *Data, correlations, and statistical models alone are not sufficient for causal inference (we also need causal models)!*
Required reading: Shalizi (2018), Ch. 2.4 + PGJ, pp. 1-6
Further reading: Shalizi (2018), Ch. 2
Assignment 4 (due Oct 29, 9:59am)

Building causal models (Oct 31 & Nov 5)
Key point: *Causal models are theories/assumptions about the common causes of our phenomena of interest!*
Required reading: ELW, pp. 245-249
Further reading: PGJ Ch. 1.4 + 1.5
Assignment 5 (due Nov 5, 9:59am)

From causal models to correlations (Nov 7 & 12)
Key point: *Causal models imply a set of correlations we should see in data (but the same set of correlations is consistent with multiple causal models)!*
Required reading: ELW, pp. 249-254
Further reading: PGJ Ch. 2
Assignment 6 (due Nov 12, 9:59am)

Defining and identifying total effects (Nov 14 & 19)
Key point: *Causal inference addresses what-if questions by separating causal correlations from noncausal correlations on the basis of causal models (which is often very tricky and sometimes impossible)!*
Required reading: HR Ch. 1 + 3 + 6.4 + 4.1 + 4.3 + 6.6
Further reading: ELW, pp. 254-261 + PGJ Ch. 3 + 4
Assignment 7 (due Nov 19, 9:59am)

Confounding bias (Nov 21 & 26)
Key point: *Failing to control for specific variables can bias causal inference!*
Required reading: HR Ch. 7 (except 7.4)
Further reading: ELW, pp. 261-262
Assignment 8 (due Nov 26, 9:59am)

Collider bias and overcontrol bias (Nov 28 & Dec 3)
Key point: *Controlling for specific variables can bias causal inference, too!*
Required reading: HR Ch. 8
Further reading: Elwert & Winship (2013) + Hernán et al (2004)
Assignment 9 (due Dec 3, 9:59am)

Randomized experiments and bias (Dec 5 & 10)
Key point: *Randomization is a design-based solution for some (but not all) potential biases in causal inference!*
Required reading: HR Ch. 2.1 + PGJ Ch. 4.3.3
Further reading: Deaton & Cartwright (2018) + Mansournia et al (2017) + Sampson (2010)
Assignment 10 (due Dec 10, 9:59am)

Adjusting for covariates using regression (Dec 12 & 17)
Key point: *Under specific conditions, regression can be useful for causal inference!*
Required reading: HR, Ch. 10.1 + 10.2 + 10.5 + 15.1.
Assignment 11 (due Dec 17, 9:59am)

Adjusting for covariates through re-weighting (Dec 19 & Jan 7)
Key point: *Regression is one of many statistical tools that can be used for causal inference!*
Required reading: HR, Ch. 2.4 + 12 (except 12.3 + 12.6).
Assignment 12 (due Jan 7, 9:59am)

Adjusting for missing data through re-weighting (Jan 9 & 14)
Key point: *See above!*
Required reading: HR, Ch. 12.6
Assignment 13 (due Jan 14, 9:59am)

Further topics in causal inference (Jan 16, 21, 23 & 28)
Key point: *There are many many more things to know about causal inference!*
Further reading: see lecture slides for references

Exam due (Jan 30)
Key point: *Register on time for the exam on Klips!*

Class resources

Materials and homework submission on Ilias
https://www.ilias.uni-koeln.de/ilias/goto_uk_crs_2629351.html

Q&A on Piazza
This term we will be using Piazza for class discussion. The system is highly catered to getting you help fast and efficiently from classmates and myself. Rather than emailing questions, I encourage you to post your questions on Piazza. If you have any problems or feedback for the developers, email team@piazza.com.

Find our class page at:
https://piazza.com/uni-koeln.de/winter2018/basismodulsoziologieiii/home

Piazza is also available as a costfree app for Android and iOS.

DAGitty
Software to draw and analyze graphical causal models (DAGs) by Johannes Textor: http://www.dagitty.net/

Tutorial on using DAGitty by Scott Venners (video)

Other online resources

- Causal Diagrams by Miguel Hernán (free online class)
- Causal Inference Bootcamp (videos)
- Causal mediation analysis by Tyler VanderWeele (videos)
- Tutorial on non-parametric causal models by Thomas Richardson (video)
- Introduction to causal inference by Maya Petersen and Laura Balzer (course materials)
- Judea Pearl's blog

- Twitter #Causalinference (selection):
  Laura Balzer, Elias Bareinboim, Rhian Daniel, Fernando Martel García, Maria Glymour, Miguel Hernán, Stephen L. Morgan, Ellie Murray, Manjari Narayan, Judea Pearl, Sherri Rose, Bianca De Stavola, Peter Tennant, Johannes Textor

Overview articles (from different disciplines)

Causal inference:

- Gangl, M. (2010). Causal inference in sociological research. *Annual Review of Sociology* 36, pp. 21– 47.
- Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47 (1), pp. 5–86.
- Keele, L. (2015). The statistics of causal inference: A view from political methodology. *Political Analysis* 23 (3), pp. 313–335.
- Petersen, M. L. and Laan, M. J. van der (2014). Causal models and learning from data: Integrating causal modeling and statistical estimation. *Epidemiology* 25 (3), pp. 418–426.
- Shalizi, C. (2018). *Advanced data analysis from an elementary point of view.* New York: Cambridge University Press, Ch. 21-24.

Graphical causal models:

- Elwert, F. (2013). Graphical causal models. In: *Handbook of Causal Analysis for Social Research.* Ed. by Morgan, S. L. New York: Springer, pp. 245–272.
- Glymour, M. M. and Greenland, S. (2008). Causal diagrams. In: *Modern Epidemiology. Third Edition.* Ed. by Rothman, K. J., Greenland, S., and Lash, T. L. Philadelphia, PA: Lippincott Williams & Wilkins, pp. 183–209.
- Rohrer, J. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science* 1 (1), pp. 27-42.
- Steiner, P. M. et al. (2017). Graphical models for quasi-experimental designs. *Sociological Methods & Research* 46 (2), pp. 155–188.

Randomized controlled trials:

- Deaton, A. and Cartwright, N. (2016). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine.*
- Jackson, M. and Cox, D. R. (2013). The principles of experimental design and their application in sociology. *Annual Review of Sociology* 39, pp. 27–49.
- Mansournia, M. A., Higgins, J. P. T., Sterne, J. A. C., and Hernán, M. A. (2017). Biases in randomized trials: A conversation between trialists and epidemiologists. *Epidemiology* 28 (1), pp. 54–59

- Sampson, R. J. (2010). Gold standard myths: Observations on the experimental turn in quantitative criminology. *Journal of Quantitative Criminology* 26 (4), pp. 489–500.

Causal mediation analysis:

- Green, D. P., Ha, S. E., and Bullock, J. G. (2010). Enough already about "black box" experiments: Studying mediation is more difficult than most scholars suppose. *Annals of the American Academy of Political and Social Science* 628 (1), pp. 200-208.
- Keele, L. (2015). Causal mediation analysis: Warning! Assumptions ahead. *American Journal of Evaluation* 36 (4), pp. 500-513.
- Knight, C. and Winship, C. (2013). The causal implications of mechanistic thinking: Identification using directed acyclic graphs (DAGs). In: *Handbook of Causal Analysis for Social Research*. Ed. by Morgan, S. L. Dordrecht u.a.: Springer, pp. 275-299.
- VanderWeele, T. J. (2016). Mediation analysis: A practitioner's guide. *Annual Review of Public Health*. 37, pp. 17-32.

Causality bookshelf

Introductions to causal inference:

- Hernán, M. A. and Robins, J. M. 2018. *Causal Inference*. Boca Raton, FL: Chapman & Hall/CRC.
- Imbens, G. W. and Rubin, D. B. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction*. New York: Cambridge University Press.
- Morgan, S. L. and Winship, C. 2015. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Second Edition. New York: Cambridge University Press.
- Pearl, J., Glymour, M., and Jewell, N. P. 2016. *Causal Inference in Statistics: A Primer*. West Sussex, UK: Wiley.
- Pearl, J. and Mackenzie, D. 2018. *The Book of Why: The New Science of Cause and Effect*. New York: Basic Books.
- Rosenbaum, P. R. 2017. *Observation and Experiment: An Introduction to Causal Inference*. Cambridge, MA: Harvard University Press.

Graphical causal models:

- Glymour, C. 2001. *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*. Cambridge, MA: MIT Press.
- Pearl, J. 2009[2000]. *Causality: Models, Reasoning, and Inference*. Second Edition. New York: Cambridge University Press.
- Sloman, S. 2005. *Causal Models: How People Think About the World and its Alternatives*. Oxford, UK: Oxford University Press.
- Spirtes, P., Glymour, C., and Scheines, R. 2001[1993]. *Causation, Prediction, and Search*. Second Edition. Cambridge, MA: MIT Press.

Causal mediation (and interaction) analysis:

- Hong, G. 2015. *Causality in a Social World: Moderation, Mediation, and Spill-Over.* West Sussex, UK: Wiley- Blackwell.
- VanderWeele, T. J. 2015. *Explanation in Causal Inference: Methods for Mediation and Interaction.* New York: Oxford University Press.

Causality:

- Berzuini, C., Dawid, P., and Bernardinelli, L. (ed.) 2012. *Causality: Statistical Perspectives and Applications.* West Sussex, UK: Wiley.
- Cartwright, N. 2007. *Hunting Causes and Using Them: Approaches in Philosophy and Economics.* New York: Cambridge University Press.
- Illari, P. and Russo, F. 2014. *Causality: Philosophical Theory Meets Scientific Practice.* New York: Oxford University Press.

Research design and methods:

- Angrist, J. D. and Pischke, J.-S. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton University Press.
- Angrist, J. D. and Pischke, J.-S. 2015. *Mastering 'Metrics. The Path from Cause to Effect.* Princeton, NJ: Princeton University Press.
- King, G., Keone, R., and Verba, S. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research.* Princeton,NJ: Princeton University Press.
- Manski, C. F. 1995. *Identification Problems in the Social Sciences.* Cambridge, MA: Harvard University Press.
- Manski, C. F. 2007. *Identification for Prediction and Decision.* Cambridge, MA: Harvard University Press.
- Murnane, R. J. and Willett, J. B. 2010. *Methods Matter: Improving Causal Inference in Educational and Social Science Research.* Oxford University Press.
- Peters, J., Janzing, D., and Schölkopf, B. 2017. *Elements of Causal Inference: Foundations and Algorithms.* Cambridge, MA: MIT Press.
- Rosenbaum, P. R. 2010. *Design of Observational Studies.* New York: Springer.
- Shadish, W., Cook, T., and Campbell, D. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Belmont, CA: Wadsworth Cengage Learning.
- van der Laan, M. and Rose, S. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data.* New York: Springer.