Introduction to causal inference

Michael Kühhirt
michael.kuehhirt@uni-koeln.de

Winter 2017

Office Hours: Wed 2:00–3:00pm
Office: Greinstr. 2, 0.05
Web: goo.gl/WEZeoX

Class Hours: Mon 2:00–5:30pm
Class Room: Sociolab, Greinstr. 2

Course content

Do negative campaign ads help politicians win elections? Are there learning benefits to regular course attendance at university? Is maternal employment bad 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 course 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 interpret statistical associations as estimates of causal effects. In addition to conventional regression approaches the course introduces inverse probability weighting as a useful alternative technique for causal inference. The main topics of the course are the construction and application of graphical causal models, the precise definition of causal effects in terms of interventions and counterfactual variables, and the conditions and techniques for estimating total effects, direct and indirect effects (i.e., mediation analysis), and the effects of time-varying variables.

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 about the variables of interest, for (2) defining the effect to be estimated, for (3) assessing whether this effect can be

Version: 2017-12-09

estimated with given data, and for (4) statistical estimation (in `Stata`) and (5) interpretation of the results, this course 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, theoretical lectures are complemented with exercises and practical `Stata` labs.

Goals and learning outcomes

The course 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 course concludes. Students who complete the course 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. estimate 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 course leads you to appreciate the opportunities offered by recent advances in the theory and tools for causal inference. The course, 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

In case of successful course completion you're awarded 9 credit points. Please be aware that course completion requires a substantial time investment outside class. This includes required readings, exercises and practical labs. I highly recommend to regularly get together in small groups to discuss course contents and also to prepare comments and questions to be discussed in class.

Methodological prerequisites for the course include basic familiarity with linear regression, statistical inference (e.g., confidence intervals, standard errors) and concepts like variables, levels of measurement, and distributions. Knowing some Stata won't hurt, but isn't necessary. For the course, working with Stata mostly consists of adapting existing code files with relevant commands to your purposes.

The final grade is based on a 60-minute exam written in the last week of the course. The exam assesses your progress in achieving the learning goals 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. A minimum of 30 points is required for a passing grade. I will distribute a mock exam later in the semester. By completing three Stata labs you can earn up to 30 bonus points for the final grade. For each lab, you can earn 10 points maximum. Labs are completed through submission on Ilias until a specic deadline. It is allowed (even encouraged!) to work on labs in pairs or small groups. Nonetheless, each of you must submit your own work in the end. Identical submissions are awarded 0 points in each case.

Please note: Always check your submissions carefully with regard to spelling, grammar, and formatting (particularly ensuring that `Stata` *output is readable). Bonus points are only counted until you reach the maximum number of 60 points in combination with the exam.*

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

Course schedule

The course, for the most part, builds on the following three textbooks, which—in the course schedule—will be abbreviated as PGJ, HR, and VDW, respectively, when excerpts are designated as readings for class:

Pearl, J., Glymour, M., and Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer.* West Sussex, UK: Wiley = PGJ.

Hernán, M. A. and Robins, J. M. (2018). *Causal Inference (v. 04-10-17).* Boca Raton, FL: Chapman & Hall/CRC. = HR.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction.* New York: Oxford University Press = VDW.

Required reading not taken from these books are linked in the course schedule.

---

01. Introduction and course organization
Slides

02. Statistical relations and their interpretation
Reading: PGJ, Ch. 1.1–1.3.10.
Slides

03. Graphical causal models
Reading: PGJ, Ch. 1.4–2.5.
Slides

04. Definition of (average total) causal effects
Reading: PGJ, Ch. 3.1, 4.1–4.2.2, 4.3.2.; HR, Ch. 1.1–1.3, 1.5, 4.1, 4.3.
Slides

05. Identification of total effects I
Reading: PGJ, Ch. 3.2 (until p.58); HR, Ch. 2.1, 7.1, 7.2, 8.1, 8.2, 8.3, 8.4.
Slides

06. Identification of total effects II
Reading: PGJ, Ch. 3.3.; HR, Ch. 2.2, 2.3, 3.1–3.3, 3.6, 4.2, 4.4–4.6.
Slides

07. Estimation of total effects I
Reading: HR, Ch. 10.1, 10.2, 10.5, 11, 15.1.
Slides

08. Estimation of total effects II
Reading: HR, Ch. 2.4, 12.
Slides | Lab 1

09. Direct and indirect effects I
Reading: VDW, Ch. 1.2.1, 1.3.1, 2.16, 2.2–2.3, 2.13–2.15.
Slides

10. Direct and indirect effects II
Reading: VDW, Ch. 2.1, 2.6, 2.12.
Slides

11. Direct and indirect effects III
Reading: VDW, Ch. 5.3.
Slides

12. Cumulative effects I
Reading: HR, Ch. 19.
Slides

13. Cumulative effects II
Reading: HR, Ch. 20. Cole and Hernán (2008)
Slides

14. Cumulative effects III
Reading: Fewell et al. (2004)
Slides

Resources

Course materials and lab submission on Ilias

https://www.ilias.uni-koeln.de/ilias/goto_uk_crs_2272154.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

Other resources

Causal Diagrams by Miguel Hernán: https://www.edx.org/course/causal-diagrams-draw-assumptions-harvardx-ph559x

DAGitty: Drawing and analyzing causal diagrams (DAGs) by Johannes Textor:
http://www.dagitty.net/

Causal inference book by Miguel Hernán and James Robins:
https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/

Causal mediation analysis by Tyler VanderWeele:
https://www.hsph.harvard.edu/tyler-vanderweele/tools-and-tutorials/

Introduction to causal inference by Maya Petersen and Laura Balzer:
http://www.ucbbiostat.com/

Judea Pearl's blog: http://causality.cs.ucla.edu/blog/