10th GESIS Summer School in Survey Methodology

[2nd Virtual GESIS Summer School]

28 July – 20 August 2021

Syllabus for Short Course B:
Using Directed Acyclic Graphs for Causal & Statistical Inference

Lecturer: Julian Schuessler
E-mail: julians@ps.au.dk
Homepage: www.julianschuessler.net

Date: 28-30 July 2021
Time: 09:00-11:00 and 13:00-15:00
Time zone: CEST/CEDT, course starts on Wednesday at 09:00 am
Venue: Online via Zoom

About the Lecturer:

Julian Schuessler ist a post-doc at the Institute for Political Science, Aarhus University, Denmark. He defended his PhD on the use of causal graphs for causal and statistical inference problems in the social sciences at the University of Konstanz in 2020. In 2019, he received the American Statistical Association's "Causality in Statistics Education Award". Beyond methodological research, he is also interested in studying public opinion towards the EU, using both survey experiments and observational data.

Selected Publications:

- Schuessler, Julian, and Peter Selb. 2019. "Graphical Causal Models for Survey Inference." SocArXiv. doi:10.31235/osf.io/hbg3m.

Short Course Description:

This course uses causal graphs (or "directed acyclic graphs") as a remarkably simple, yet general and powerful framework to describe and discuss a large set of problems that empirical social scientists need to tackle. Is my question of interest descriptive or causal? How can I communicate my assumptions effectively to others, and can I test them? How can I tell correlation from causation? How do I choose control variables for my regression models? After discussing how DAGs can be used to answer these foundational questions, the course also covers basics of causal mediation, instrumental variables, nonresponse/selection bias (and adjustments for it), and (if time permits) panel data analysis from a "graphical" perspective.

Keywords:

causal graphs; causal inference; mediation; instrumental variables; nonresponse

Course Prerequisites:

Participants should be willing to learn and use formal reasoning and must have at least Bachelor-level knowledge of statistics. Basic knowledge of R is helpful.

Target Group:

Participants will find the course useful if:

- They are interested in causal questions and want to understand the assumptions associated with various "causal methods" better

- They are interested in non-causal question, want to use data suffering from nonresponse, and want to understand why causal assumptions are necessary in this case

Course and Learning Objectives:

By the end of the course participants will:

- Know how to use causal graphs to visualize causal assumptions, define quantities of interest, and to determine testability of assumptions via d-separation
- Know how to graphically determine identification of causal and descriptive quantities like average causal effects, causal interaction, effect heterogeneity, natural direct and indirect effects, and population distributions from data with nonresponse
- Know under what graphical assumptions instrumental variable and panel data analysis typically operate
- Will have some basic knowledge about how all of this relates to implementation in standard statistical software

Organizational Structure of the Course:

This short course throughout will change between short lecture-style inputs and individual or small-group hands-on exercises supervised by the lecturer and a teaching assistant (4hrs/day). Participants are encouraged to bring their own research ideas to develop them further using the material from the class. After each class, participants are expected to do some homework exercises (1hr/day). The lecturer will also be available for individual consultation in the afternoons (1hr/day).

Software and Hardware Requirements:

We will use base R for simulation and will briefly discuss some elements of the R packages "dagitty", "simcausal", "sensemakr", "mediation", "AER", "estimatr", "PanelMatch",

Long Course Description:

In recent years, "directed acyclic graphs" or "causal graphs" have become a standard tool in many disciplines to better understand causal inference. Causal graphs depict assumptions about the causal relationships between observed and unobserved variables that researchers study. They very easily illustrate how biases occur and could be handled in non-experimental research, including instrumental variables, causal mediation, or panel models, but can also be extended to better understand problems associated with missing data.

Graphs help us to reason qualitatively about whether, for example, a regression containing specific control variables will tell us anything interesting about causal effects, whether a variable qualifies as an instrumental variable, and whether we can tell apart direct from indirect effects given some data. These are the questions that most empirical researchers encounter on a daily basis and on which standard statistical or econometrical resources are often silent.

Specific questions about estimation, e.g. the choice of a linear or non-linear model or the computation of standard errors, only come second, "after the DAG has spoken". The course will discuss these questions at times, focussing on specific software solution, but will overwhelmingly focus on building up a qualitative and analytical understanding of DAGs. We will, however, discuss the software implementation and interpretation of various "sensitivity analyses" related to the methods that we cover. We will also make heavy use of simulating data in R to illustrate the methods we discuss. This does require just a little knowledge of R. Participants with no prior exposure to R are invited to work through the two videos mentioned under "Prepatory Reading".

The course will change between short lecture-style inputs and individual or small-group hands-on exercises supervised by the lecturer and a teaching assistant (4hrs/day). Participants are encouraged to bring their own research ideas to develop them further using the material from the class. After each class, participants are expected to do some homework exercises (1hr/day). The lecturer will also be available for individual consultation in the afternoons (1hr/day).

Note: The syllabus represents what is maximally possible in terms of topics; depending on participant's interests, some of it may be shortened or deepened.

Day-to-day Schedule and Literature:

Day	Topic(s)
1	**Introductions, interests and preferences of participants, self-assessments** **Descriptive, predictive, and causal questions** **Simulation in R** **Graph basics and d-separation** **Statistical control and Simpson's paradox** **Interventions and causality** **Back-door criterion and unobserved confounding** **Sensitivity analysis for unobserved confounding** **Post-treatment bias**
	<u>Compulsory reading (have to be read before the session):</u> ▪ Cinelli, Carlos, and Chad Hazlett. "Making sense of sensitivity: Extending omitted variable bias." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82.1 (2020): 39-67. – Only sections 1, 2, and 5! ▪ Play around with DAGitty: http://dagitty.net/dags.html <u>Suggested reading (suggested, yet do not have to be read before the session):</u> ▪ Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. Causal inference in statistics: A primer. John Wiley & Sons, 2016. Preface and Sections 1.1, 1.2.
2	**Causal interaction and effect heterogeneity** **Causal mediation: Linear graphs, controlled and natural effects, post-treatment confounding, sensitivity analyses** **Mediation, discrimination, and audit studies** **Instrumental variables in linear & nonparametric models** **Choosing control variables in IV models** **Sensitivity analyses for IV**
	<u>Compulsory reading:</u> ▪ Spenkuch, Jörg L., and Philipp Tillmann. "Elite influence? Religion and the electoral success of the Nazis." American Journal of Political Science 62.1 (2018): 19-36. – Only pages 19-28! <u>Suggested reading:</u> ▪ Imai, Kosuke, and Teppei Yamamoto. "Identification and sensitivity analysis for multiple causal mechanisms: Revisiting evidence from framing experiments." Political Analysis (2013): 141-171. ▪ Schuessler, Julian, Glynn, Adam N., and Rueda, Miguel. R. "Post-Instrument Bias". Working paper. (2021) ▪ Acharya, Avidit, Matthew Blackwell, and Maya Sen. "Analyzing causal mechanisms in survey experiments." Political Analysis 26.4 (2018): 357-378.
3	**Non-reponse/missing data and its consequences** **Adjustment for non-response from a graphical perspective** **Differences when interest is in descriptive or causal quantities** **Panel data and fixed effects** **Wrap-up**
	<u>Compulsory reading:</u> ▪ Schuessler, Julian, and Peter Selb. "Graphical causal models for survey inference." Working Paper (2021). ▪ Imai, Kosuke, and In Song Kim. "When should we use unit fixed effects regression models for causal inference with longitudinal data?." American Journal of Political Science 63.2 (2019): 467-490. - Only pages 467-474! <u>Suggested reading:</u> ▪ Elwert, Felix, and Christopher Winship. "Endogenous selection bias: The problem of conditioning on a collider variable." Annual review of sociology 40 (2014): 31-53. ▪ Knox, Dean, Will Lowe, and Jonathan Mummolo. "Administrative records mask racially biased policing." American Political Science Review 114.3 (2020): 619-637.

Preparatory Reading:

- Downloading R and RStudio: https://rstudio-education.github.io/hopr/starting.html
- For those who have never used R before (18 mins): https://www.youtube.com/watch?v=tvv4IA8PEzw
- Basics of simulation in R (15 mins): https://www.youtube.com/watch?v=tvv4IA8PEzw

Additional Recommended Literature:

Textbooks:

- (Aronow/Miller and Imai do not discuss graphs)
- Aronow, Peter M., and Benjamin T. Miller. Foundations of agnostic statistics. Cambridge University Press, 2019.
- Hernán MA, Robins JM. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC, 2020. https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/
- Imai, Kosuke. Quantitative social science: An introduction. Princeton University Press, 2018.
- Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. Causal inference in statistics: A primer. John Wiley & Sons, 2016.
- Peters, Jonas, Dominik Janzing, and Bernhard Schölkopf. Elements of causal inference: foundations and learning algorithms. The MIT Press, 2017.
- Shalizi, Cosma. "Advanced data analysis from an elementary point of view." (2021) http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/

Other books and papers:

- Chen, Bryant, and Judea Pearl. "Regression and causation: a critical examination of six econometrics textbooks." Real-World Economics Review, Issue 65 (2013): 2-20.
- Ogburn, Elizabeth L., and Tyler J. VanderWeele. "Causal diagrams for interference." Statistical science 29.4 (2014): 559-578.
- Pearl, Judea. Causality. Cambridge university press, 2009.
- Pearl, Judea. "Linear models: A useful “microscope” for causal analysis." Journal of Causal Inference 1.1 (2013): 155-170.
- Pearl, Judea. "Causes of effects and effects of causes." Sociological Methods & Research 44.1 (2015): 149-164.
- Pearl, Judea. "Conditioning on post-treatment variables." Journal of Causal Inference 3.1 (2015): 131-137.
- Pearl, Judea. "Generalizing experimental findings." Journal of causal inference 3.2 (2015): 259-266.
- Pearl, Judea. "Indirect Confounding and Causal Calculus (On three papers by Cox and Wermuth)." (2015).
- Pearl, Judea. "Theoretical impediments to machine learning with seven sparks from the causal revolution." arXiv preprint arXiv:1801.04016 (2018).