

Research Design and Causal Analysis with R

Data Science Summer School

July 21, 2021, 10 AM – 2:15 PM CET

Online – Zoom link TBA

Julian Schuessler
Post-Doc
Aarhus University
Department of Political Science
julians@ps.au.dk
www.julianschuessler.net

Most empirical researchers and decision makers are interested in causal questions:

- Does exposure to violence change people's attitude towards peace?
- How does information about migration impact on people's political preferences?
- Are AI algorithms racially biased?
- Does foreign aid improve human rights?

In recent years, methodological research on causal inference based on directed acyclic graphs (DAGs) and potential outcomes has fundamentally altered and expanded our knowledge on how to answer such questions.

In this workshop, we will use DAGs to highlight crucial assumptions for answering such questions and to explain sources of bias, and we will discuss the use of various R packages for sensitivity analysis.

We will first discuss what differentiates causal from other kind of questions. We will then discuss the graphical (DAG-based) “back-door criterion” that allows us to judge whether statistical adjustment for control variables is sufficient to estimate causal effects. We will discuss the nature of unobserved confounding and how one can use sensitivity analysis in R to probe whether estimates are plausibly causal. We will then look deeper into post-treatment bias, that is, bias introduced by controlling for “too many” variables that are influenced by the independent variable of interest.

Problems with post-treatment variables lead us to the (valid) statistical analysis of causal mechanism. We discuss the definition and interpretation of direct and indirect effects, under what circumstances such effects can be estimated by adjusting for mediators, and how sensitivity analysis can be used to assess bias in these cases.

Finally, we consider instrumental variables as a prominent approach to estimate causal effects when there is unobserved confounding. We discuss graphical conditions for valid instrumental variables, basic estimation, and sensitivity analysis.

Depending on participants' interests and if time permits, we can also discuss modern approaches to panel analysis or the causal foundations of missing data adjustments.

Much of the workshop will be “pen-and-paper-” and discussion-based, but some (very) basic knowledge of R and linear regression is required.

Background literature:

Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.

Pearl, Judea. "Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution." *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. 2018.

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.

Montgomery, Jacob M., Brendan Nyhan, and Michelle Torres. "How conditioning on posttreatment variables can ruin your experiment and what to do about it." *American Journal of Political Science* 62.3 (2018): 760-775.

Imai, Kosuke, et al. "Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies." *American Political Science Review* (2011): 765-789.

Schuessler, Julian, Glynn, Adam N., and Miguel R. Rueda. "Post-Instrument Bias." Working Paper.

Schuessler, Julian, and Peter Selb. "Graphical causal models for survey inference." Working Paper. <https://osf.io/preprints/socarxiv/hbg3m/>