GESIS Summer School in Survey Methodology 2023

Syllabus for course:
“Causal Inference Using Survey Data”

Lecturers: Heinz Leitgöb  
Tobias Wolbring
E-mail: heinz.leitgoeb@uni-leipzig.de  
tobias.wolbring@fau.de
Website: http://sozweb.sozphil.uni-leipzig.de/de/personen.html  
https://www.wirtschaftssoziologie.rw.fau.eu/person/tobias-wolbring/

Date: 07 – 11 August 2023
Time: Mo: 10:00-17:00 | Tu-Thu: 09:00 – 16:30 | Fr: 09:00 – 16:00
Venue: On-site at GESIS Cologne

About the Lecturers:

**Heinz Leitgöb** is an Interim Professor of Quantitative Research Methods at the Institute of Sociology, Leipzig University, Germany. He is also associated at the Goethe University of Frankfurt, Germany. He chairs the methods section of the German Sociological Association and the European Working Group on Quantitative Methods in Criminology of the European Society of Criminology. Besides causal inference, his research interests cover the topics of rare events modeling, measurement invariance, survey mode effects, digitalization, and quantitative criminology.

**Tobias Wolbring** is a Professor of Empirical Economic Sociology at School of Business, Economics, and Society at FAU Erlangen-Nürnberg, Germany. He is particularly focussed on applying statistical methods for panel data and on running different kinds of experiments to draw causal inference. His research covers a wide range of topics including labour market behaviour, discrimination, social inequality effects, the validity of students’ evaluations of teaching, and the determinants of subjective well-being.

Selected Publications:


Course Description:
This course will introduce participants to the concepts and methods of causal inference and causal modeling in the social sciences. It will highlight the relevance of research design, analytical methods and their systematic combination to optimize the validity of causal inferences drawn from empirical studies. Participants will learn the key principles and techniques of causal inference, including potential outcomes, counterfactuals, and causal graphs, and will get to know the experimental approach to causality. Building on existing knowledge concerning linear regression modeling and research design, the course will then cover key methods of causal modeling using survey data, such as fixed effects panel models, matching, difference-in-differences, regression discontinuity, and instrumental variables. Throughout the course, participants will apply these concepts and methods in hands-on sessions to real-world examples in the social sciences. The application will be conducted with the statistical software package Stata. The course will also touch upon advanced topics such as effect modification, reverse causality, measurement issues, and data quality. By the end of the course, participants will have the skills and knowledge to design, conduct, and interpret causal inference studies in the social sciences. They will be able to engage with the contemporary literature of causal inference and identify state-of-the-art methods which might be most relevant to their specific research question.

Keywords:
causal inference, observational studies, panel data, experiments, causal identification

Course Prerequisites:
▪ Knowledge of basic statistical concepts, including the principles of linear and binary logistic regression.
▪ Background in statistical software, preferably Stata.
▪ Basic understanding of designing quantitative studies.

Target Group:
Participants will find the course useful if they:
▪ have a background in the social, behavioral or economic sciences (economists, political scientists, sociologists, criminologists, psychologists, etc.)
▪ are interested in methods for causal inference based on experimental and/or observational data, especially panel data
▪ have a firm knowledge in linear regression modelling
▪ are motivated to apply the concepts and statistical approaches in hands-on sessions.

Course and Learning Objectives:
By the end of the course participants will:
▪ Have a good understanding of the potential outcome framework, causal diagrams, and the counterfactual way of thinking.
▪ Be capable of designing their own studies to derive causal estimates in observational settings.
▪ Acquire an in-depth understanding of and the skills to carry out five family of methods: fixed effects models, matching, difference-in-differences, instrumental variables, and regression discontinuity design.
▪ Become familiar with interdisciplinary applications of the methods covered by the course.
▪ Be able to engage the contemporary literature of causal inference and identify state-of-the-art methods which might be most relevant to their specific research question.
Organizational Structure of the Course:
The course will be split into a three-hour morning and a three-hour afternoon session, including coffee breaks. In order to secure a close link between the learning and the application of contents, we will switch between lecture format (~50%) and hands-on exercises, tutorials, or lab sessions (~50%) in a flexible way. In addition to shorter exercises, a selected number of more in-depth assignments will be provided which participants solve in groups of 2-3. These include the application of causal inference methods to estimate effects based on existing datasets using Stata. Lecturers will be available for individual consultations to support work on group assignments and to facilitate discussions within groups.

Software and Hardware Requirements:
Participants will need to bring a laptop computer with a recent version of Stata (13 or higher) installed to successfully participate in this course. Stata short term licenses will be provided by GESIS for the duration of the course if needed.

Day-to-day Schedule and Literature:

<table>
<thead>
<tr>
<th>Day</th>
<th>Topic(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><strong>Causality, Counterfactuals and Experiments</strong>&lt;br&gt;10:00 – 13:00</td>
</tr>
<tr>
<td></td>
<td>▪ General introduction to causality and causal inference (foundations in the philosophy of science, perspectives, conditioning, selection, endogeneity, …)</td>
</tr>
<tr>
<td></td>
<td>▪ Counterfactual model and the fundamental problem of causal inference</td>
</tr>
<tr>
<td></td>
<td>▪ Different types of estimands: ATE, ATT, ATC</td>
</tr>
<tr>
<td></td>
<td>▪ What is your estimand?</td>
</tr>
<tr>
<td></td>
<td>14:00 – 17:00</td>
</tr>
<tr>
<td></td>
<td>▪ Directed Acyclic Graphs (DAGs) I: Foundations</td>
</tr>
<tr>
<td></td>
<td>▪ Different types of experiments and their link to the counterfactual model</td>
</tr>
<tr>
<td></td>
<td>▪ Factorial survey experiments</td>
</tr>
<tr>
<td>2</td>
<td><strong>Experiments and Cross-sectional Data</strong>&lt;br&gt;9:00 – 12:00</td>
</tr>
<tr>
<td></td>
<td>▪ Natural experiments in a survey context</td>
</tr>
<tr>
<td></td>
<td>▪ Stata Application: Factorial Survey Experiments &amp; Natural Experiments</td>
</tr>
<tr>
<td></td>
<td>13:30 – 16:30</td>
</tr>
<tr>
<td></td>
<td>▪ Regression, matching &amp; inverse probability weighting</td>
</tr>
<tr>
<td></td>
<td>▪ Directed Acyclic Graphs (DAGs) II: Confounders, colliders &amp; mediators</td>
</tr>
<tr>
<td></td>
<td>▪ Stata Application: regression, matching &amp; inverse probability weighting</td>
</tr>
<tr>
<td>3</td>
<td><strong>Longitudinal Data</strong>&lt;br&gt;9:00 – 12:00</td>
</tr>
<tr>
<td></td>
<td>▪ Longitudinal (in particular panel) designs</td>
</tr>
<tr>
<td></td>
<td>▪ Handling panel data</td>
</tr>
<tr>
<td></td>
<td>▪ First-differences, random and fixed effects regression</td>
</tr>
<tr>
<td></td>
<td>13:30 – 16:30</td>
</tr>
<tr>
<td></td>
<td>▪ Diff-in-diff-estimation</td>
</tr>
<tr>
<td></td>
<td>▪ Tailoring control groups (sample restrictions, matching, …)</td>
</tr>
<tr>
<td></td>
<td>▪ Stata Application: First-differences, random, fixed effects regression, Diff-in-diff-estimation</td>
</tr>
</tbody>
</table>
### Temporal Dynamics, Heterogeneity, and Data Quality

**9:00 – 12:00**

- Impact Dummies
- Heterogeneous treatment effects (standard tools, more advanced tools)
- Methods to address: reverse causality & simultaneity: Instruments & ML-SEM

**13:30 – 16:30**

- Stata Application: Impact Dummies, Heterogeneous Effects & IV estimation
- Impact of data quality on causal inference I (e.g., missing values, panel mortality, reactivity)

### Advanced Topics

**9:00 – 12:00**

- Impact of data quality on causal inference II (e.g., missing values, panel mortality, reactivity)
- Measurement issues (graph theory; measurement invariance; changing measures)

**13:00 – 16:00**

- Stata Application: Data Quality & Measurement Issues
- Further advanced topics
- Wrap-up & Outlook

---

### Preparatory Reading:

For some basic background on causality:

To review material on statistics, please consult this excellent overview:

To familiarise yourself with the statistical software Stata, you can use the following textbook:

### Additional Recommended Literature:

Oaks: Sage.
Press.