BAMBERG
GRADUATE SCHOOL
OF SOCIAL SCIENCES

The Statistics of Causal Inference

Instructor: Professor Dr Elias Dinas, European University Institute

Course outline

> Whenever, looking at my watch, I see the hand has reached the figure X, I hear the bells beginning to ring in the church close by. But from the fact that the watch hands point to ten whenever the bells begin to ring, I have not the right to infer that the position of the hands of my watch is the cause of the vibration of the bells.
>
> Leo Tolstoy, War and Peace, trans. Constance Garnett
> (New York: Modern Library Classics, 2002), p. 939.

Do hospitals make people healthier? Is it a problem that more people die in hospitals than in bars? Does an additional year of schooling increase future earnings? Do parties that enter the parliament enjoy vote gains in subsequent elections? The answers to these questions (and many others which affect our daily life) involve the identification and measurement of causal links: an old problem in philosophy and statistics. To address this problem we either use experiments or try to mimic them by collecting information on potential factors that may affect both treatment assignment and potential outcomes. Customary ways of doing this in the past entailed the specification of sophisticated versions of multivariate regressions. However, it is by now well understood that causality can only be dealt with during the design, not during the estimation process. The goal of this course is to familiarize participants with the logic of casual inference, the underlying theory behind it and introduce research methods that help us approach experimental benchmarks with observational data. Particular emphasis is placed on how to use these methods in order to address historical questions. Why did Protestant countries flourish more than catholic ones? How did the European migrants integrate in the US throughout the 20th century? What is the effect of Nazi territorial control on patterns of resistance during the WWII? While discussing these and other applications, we will also shed light on how to think about archival work under the prism of causal inference. Hence,

this will be a much applied course, which aims at providing participants with ideas for strong research designs in their own work.

The objective is to learn how statistical methods can help us to draw causal claims about phenomena of interest. Participants will be introduced into an authoritative framework of causal inference in social sciences, i.e. the potential outcomes framework. By the end of the course, students will be in position to:

1. critically read and evaluate statements about causal relationships based on some analysis of data;
2. apply a variety of design-based easy-to-implement methods that will help them draw causal inferences in their own research.
3. think about archival data under the logic of causal inference. Either explicitly or implicitly, the goal of most empirical research is to interpret causally the co-occurrence of interesting phenomena.

Addressing causality, however, has been notoriously difficult without the luxury of experimental data. This course will introduce you to methods that allow you to make convincing causal claims without working with experimental data. In the first part of the course, we will look at three such designs:

1. Difference-in-Differences estimation;
2. Instrumental Variables and;
3. Regression Discontinuity Design

For every method, the following structure will be employed: first, a running example will provide the motivation and intuition. We will then proceed with the formal identification derivation and finally we will focus on estimation strategies and robustness checks. For each method there will be a hands-on lab section, where we will apply these methods with real data.

We will not have the time to cover two topics, which I would love to talk about: attrition and bounds. Do ask me about this towards the second part of the course.

Timetable

Day 1

- Introduction to the Potential Outcomes Framework
  - Motivation, examples, discussion. We will see examples of the fundamental problem of causal inference. Introduction to the potential outcomes framework. We will derive the causal quantities of interest.
  - How experiments solve the fundamental problem of causal inference. The logic of randomization and why it works.

Day 2

- Randomization given by nature: Examples of Natural Experiments
- Instrumental Variables: Intuition, Identification & Estimation, Wald Estimator, 2SLS Estimator

Day 3

- Regression Discontinuity Design: Motivation, Identification, Estimation Strategies.
- Testing for Sorting, Robustness Checks, Examples, Applications.

Day 4

- The Fuzzy RD
- The Local Randomization Framework
- Extensions: Identification away from the Cut-off point & Randomization-Based Inference in the RD design.
- Lab Session: Regression Discontinuity Design

Day 5

- Difference-In-Differences: Motivation, examples, identification, estimation.
- Threads to validity and examples; Gentle introduction into Staggered DiD

About the Trainer

Elias Dinas holds the Swiss Chair in Federalism, Democracy and International Governance. He holds a PhD in Political Science from the European University Institute (2010) and his research interests include the dynamics of political socialization, the downstream effects of institutional interventions and the legacy of authoritarian rule on the ideological predispositions of citizens in new democracies. He has also a keen interest in research methodology. His work has been published, among others, in the American Political Science Review, the American Journal of Political Science, the Journal of Politics, and Political Analysis and mentioned in The Economist, the Atlantic and the New York Times.