# Applied Causal Inference with R

Felix Hartmann
Email: f.hartmann@hu-berlin.de
Chair of Comparative Political Behaviour, Department of Social Science, Humboldt-Universität zu Berlin
Summer Semester 2021

## General Information

We will be distributing readings and assignments via Moodle at least one week in advance.

*Where/When*

*Office Hours* Monday 2-3 pm: https://hu-berlin.zoom.us/j/63534457667

## Overview

The seminar will combine methodological sessions introducing the design-based approach to causal inference with substantive sessions where we going to apply the ideas using the statistical software R. The course provides an introduction to the design-based approach to causal inference that will be used to evaluate whether citizens are able to hold politicians accountable. Topics include (1) randomised experiments, (2) matching, (3) regression, (4) difference-in-differences, and (5) regression discontinuity designs. The course encourages students to think about the assumption necessary to make causal claims, to become a critical consumer of causal claims in the social sciences, and equip them to conduct their own research.

Prior knowledge of hypothesis testing and linear regression is required, knowledge of the statistical software R is an advantage.

## Learning Outcomes

First, students will understand the potential outcomes framework, and the key assumptions underlying causal inference, and will be able to choose appropriate methods for a variety of research questions posing different identification challenges. Second, students will be apply their knowledge to empirical papers using the statistical software R.

## Assessment

The assessment consists of problem sets (40%) and a research design essay (60%), in which students will outline how they would address a causal research question of their choice using methods introduced in class. The research design should be structured like a pre-analysis plan (PAP) and include a short literature review, hypotheses, research design, and estimation.

## Prerequisites

Linear Regression, hypothesis testing

*Involvement* Participation includes coming to class; turning in assignments on time; thinking and caring about the material and expressing your thoughts respectfully and succinctly in class. As much as possible, we will be working in groups during the class meetings. This work will require that you have done the assigned reading in advance and that you are an active collaborator.

*Problem Sets* For the first part of the seminar students should submit problem sets as a PDF document via Moodle on the Tuesday after each seminar. I recommend using Rmarkdown to prepare your problem sets, though you can use whatever you like as long as you compile a PDF. All students must write up their problem sets individually. However, you may work in groups of up to three (though you are not required to work in groups at all). Please indicate at the top of your homework the names of the other students you worked with that week. Don't "share" members across groups. Do not copy and paste the answers across group members. All problem sets are self-assessed. A solution set will be provided on the course website and the activity will be discussed in class. Hand in the **problem sets via Moodle the next Wednesday after class**.

*Final Paper* The central assignment for the class is a research design paper in a form of a pre-analysis plan. The final paper should be divided into seven sections: 1) Introduction; 2) Theory and Hypothesis; 3) Identification 4) Estimand and Estimation Strategy; 5) Implications; 7) . I'll provide more information on this as we go along. Hand in research design paper by **August 31st, 2021** via Moodle.)

*Computing* For some exercises, we will be using R in class. Please install R (http://www.r-project.org) and R-Studio (https://rstudio.com/products/rstudio/download/#download) on your computers before the first class session. As you work on your papers, you will also learn to write about data analysis in a way that sounds and looks professional by using either Word, OpenOffice, or Wordperfect, or a typesetting system like LaTeX, to produce documents that are suitable for correspondence, collaboration, publication, and reproduction. I recommend R-Markdown. No paper will be accepted without a code appendix or reproduction archive attached. No paper will be accepted unless it is in PDF. No paper will be accepted with cut and pasted computer output in the place of well presented and replicable figures and tables.

## Books

*Required* Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press

Gerber, A. S. and Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton

*Applied with R* Irizarry, R. A. (2019). *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press

*Additional* Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton University Press, Princeton

## Schedule

### Week 1 (Apr. 16): Introduction

**Before Class:** Make sure you have a satisfactory setup for statistical computing in R.

**Topics:** Overview of the course.

**Reading:** Keele (2015)

### Week 2 (Apr. 23): Introduction to R and RStudio

**Problem Set** Go through chapter 1, 2, and 40 of: https://rafalab.github.io/dsbook/getting-started.html

### Week 3 ( Apr. 30.): Causal Inference

**Topics:** Causality as counterfactuals, Potential outcomes, Identification and estimation, Causal estimands

**Reading:** Angrist and Pischke, 2014, Ch. 1: p. 1-17, Gerber and Green, 2012, Ch. 1-2.6: p. 1-39

### Week 4 (May 7.): Applied Causal Inference

**Problem Set:** Potential Outcomes, ATE vs. ATT vs ATC

### Week 5 (May 14): Randomized Experiments

**Topics:** Identification of Causal Effects under Randomization; Covariate adjustment

**Reading:** Angrist and Pischke, 2014, p. 17-33; Gerber and Green, 2012, Ch. 2.7-3: p. 39-86

### Week 6 (May 21): Randomized Experiments

**Problem Set:** Applied (with R): Go through chapter 4 of: https://rafalab.github.io/dsbook/getting-started.html, Experimental Design

### Week 7 (May. 28): Selection on Observables and Matching

**Topics:** Identification under Selection on Observables

**Reading:** Angrist and Pischke (2014, p. 47-79); Gelman and Hill (2007), Sekhon, 2009

### Week 8 (Jun. 4): Applied Selection on Observables and Matching

**Problem Set:** Applied (with R): Olken (2007)

### Week 9 (June. 11): Instrumental Variables

**Topics:** Identification: Using Exogenous Variation in Treatment Intake Given by Instruments

**Reading:** Angrist and Pischke (2014, Ch. 3: p. 98-139), Gerber and Green (2012, Ch. 5-6: p. 131-67 & 173-205 )

### Week 10 (June. 18): Instrumental Variables

**Problem Set:** Applied (with R): Acemoglu et al. (2001)

### Week 11 (June. 25): Regression Discontinuity Design

**Topics:** Identification, Estimation, Falsification Checks

**Reading:** Angrist and Pischke (2014, Ch. 4: p. 147-75 )

### Week 12 (July. 2): Regression Discontinuity Design

**Problem Set:** replicate Titiunik (2011)

### Week 13 (July. 9): Difference-in-Differences, Fixed Effects

**Topics:** Identification, Estimation, Falsification tests

**Reading:** Angrist and Pischke, 2014, Ch. 5: p. 178-204

### Week 14 (July. 16): Difference-in-Differences

**Problem Set:** Applied (with R): Dube and Vargas (2013)

# I References


Acemoglu, D., Johnson, S., and Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American economic review*, 91(5):1369–1401.

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press.

Dube, O. and Vargas, J. F. (2013). Commodity price shocks and civil conflict: Evidence from colombia. *The review of economic studies*, 80(4):1384–1421.

Gelman, A. and Hill, J. (2007). Causal inference using regression on the treatment variable.

Gerber, A. S. and Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton University Press, Princeton.

Irizarry, R. A. (2019). *Introduction to data science: Data analysis and prediction algorithms with R*. CRC Press.

Keele, L. (2015). The statistics of causal inference: A view from political methodology. *Political Analysis*, 23(3):313–335.

Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in indonesia. *Journal of political Economy*, 115(2):200–249.

Sekhon, J. S. (2009). Opiates for the matches: Matching methods for causal inference. *Annual Review of Political Science*, 12:487–508.

Titiunik, R. (2011). Incumbency advantage in brazil: Evidence from municipal mayor elections (under revision). *University of Michigan*.