INSTITUTE OF LAW, POLITICS AND DEVELOPMENT

Scuola Superiore Sant'Anna

Data Analysis for the Social Sciences with R
AY 2024/25 - Semester II
20h - 2CFU

Prof. Kevin Koehler

Sant'Anna School of Advanced Studies
kevin.koehler@santannapisa.it

Course description

This course is an introduction to data analysis for the social sciences using R. It focuses on hands-on applications, rather than statistical or mathematical background. Basic knowledge of statistics is useful, but not strictly necessary. Interested students are encouraged to consider following Prof. Bressanelli's introductory class on quantitative approaches to gain a basic understanding.

In the first part of the course, students will gain familiarity with RStudio, a free user interface for R. In particular, we will cover basic issues of data management and data wrangling using the `tidyverse` family of tools, as well as data visualization using `ggplot` and the basics of R programming.

In the second part of the course, we will then move on to data analysis. We will learn how to implement linear and logistic regressions in R, as well as how to efficiently present results in graphical or tabular form. In order to apply these tools, we will read a published paper, critically discuss the research design, and then replicate the analysis. By the end of the course, participants will be able to independently implement basic data management and analysis tasks in R.

The textbook used in this class is Kennedy, Ryan and Philip D. Waggoner. 2021. *Introduction to R for Social Scientists: A Tidy Programming Approach*, CRC Press.

Prerequisites

- Install R (here) and RStudio (here) before the first class

Assessment

Participants are assessed in the following way:

1. Attendance and participation: 50%
2. Final research note: 50%

Overview of sessions

1 Introduction: Getting set up
2 Data management
3 Visualization
4 R programming
5 Exploring data
6 Linear regression
7 Replication I: U.S. Soft-Power and Foreign Policy Behavior (OLS)
8 Regression models for binary outcomes
9 Replication II: Does Counterbalancing Prevent Military Coups? (Logistic Regression)
10 Conclusion: What else can R do?

Detailed description

Session 1: Getting set up

- **Aim**: What is R and why should you care about it? What can you expect from this course (and what can't you)?
- **Readings**: Kennedy and Waggoner 2021, Introduction and Chapter 2.

Session 2: Data management

- **Aim**: Data management is a major part of quantitative work in the social sciences. In this session, we work with the `tidyverse` set of tools to master various aspects of data management and data wrangling.
- **Readings**: Kennedy and Waggoner 2021, Chapter 3.

Session 3: Visualization

- **Aim**: R includes a range of powerful tools for data visualization. This session introduces the grammar of graphics approach implemented in the `ggplot` package.
- **Readings**: Kennedy and Waggoner 2021, Chapter 4.

Session 4: R programming

- **Aim**: R is not only a software package for data analysis, but primarily a programming language. This session introduces the basics of R programming.
- **Readings**: Kennedy and Waggoner 2021, Chapter 5.

Session 5: Exploring data

- **Aim**: Exploring your data is an essential first step preceding data analysis. This session introduces tools and procedures for data exploration in R.
- **Readings**: Kennedy and Waggoner 2021, Chapter 6.

Session 6: Linear regression

- **Aim**: Understanding the logic of linear regression is essential for understanding more complex regression approaches. This session discusses the linear regression model and its implementation in R.
- **Readings**: Kennedy and Waggoner 2021, Chapter 7 (specifically focus on section 7.5 Ordinary Least Squares Regression).

Session 7: Replication I - U.S. Soft-Power Matter and Foreign Policy Behavior (OLS)

- **Aim**: In this session we discuss the research design and replicate the analysis in an article by Benjamin Goldsmith and Yusaku Horiuchi (2012) published in *World Politics*. The article examines whether U.S. 'soft-power' matters for the foreign policy behavior of other states.

- **Readings**: Goldsmith, Benjamin E. and Yusaku Horiuchi. 2012. "In Search of Soft Power: Does Foreign Public Opinion Matter for US Foreign Policy?" *World Politics*, 64 (3): 555-585. doi:10.1017/S0043887112000123
- Replication data can be found here

Session 8: Regression models for binary outcomes

- **Aim**: Many outcomes we are interested in as social scientists are binary: Whether an event does or does not occur (for example, civil war onset, military coups, revolutions, etc). This session discusses how the linear regression approach discussed before can be generalized to deal with such outcomes.
- **Readings**: Kennedy and Waggoner 2021, Chapter 7 (specifically focus on section 7.6 Binary Response Models).

Session 9: Replication II - Does Counterbalancing Prevent Military Coups? (Logistic Regression)

- **Aim**: In this session we discuss the research design and replicate the analysis in an article by Erica de Bruin (2017) published in the *Journal of Conflict Resolution*. The article examines whether counterbalancing is an effective strategy to prevent military coups.
- **Readings**: De Bruin, Erica. 2018. "Preventing Coups d'état: How Counterbalancing Works." Journal of Conflict Resolution 62 (7): 1433–58. doi: 10.1177/0022002717692652.
- Replication data can be found at the journal website.

Session 10: What else can R do?

- **Aim**: In this session I will give a brief overview of what R can do beyond regression analysis.
- **Readings**: TBD