Applied Quantitative Methods for the Social Sciences I

Universidad Carlos III de Madrid

Program: Master in Social Sciences, Fall 2023
Room: 18.1.A04
Time: Mondays, 10:00–13:00

Contact Information

Instructor: Patrick Kraft, PhD
Office: 18.2.A19
Email: patrickwilli.kraft@uc3m.es
Office Hours: Mondays, 14:30–16:00, or by appointment.

I Overview

Course Description

This is a first course on statistical inference and modeling for use in social science research. It covers the theory of statistical inference, essential concepts in statistical modeling, justifications for and problems with common statistical procedures, and how to apply procedures to empirical social science data to draw conclusions relevant to positive social theory. We will pay particular attention to the motivation for statistical inference and modeling from the standpoint of social science. Lectures and reading will primarily cover theory and simple examples. Problem sets will cover both simple theoretical extensions and applications of tools we develop to real data. The topics we will discuss here include:

1. Programming with R
2. R Markdown & LaTeX
3. Probability theory
4. Statistical inference
5. Introduction to OLS
6. Data visualization.

Prerequisites

Students should have completed our intro course "Mathematics for Social Sciences and Basic Statistics" or its equivalent. Students should have a working knowledge of arithmetic, algebra, and elementary calculus. The course is suitable for students with a large range of prior exposure to statistics and mathematics. All students capable of gaining admission to our MA program can fully succeed in this class regardless of prior technical preparation other than the required skills listed above.

II Textbook and Required Material

There are two required textbooks for the course:

- Gailmard, Sean. 2014. *Statistical modeling and inference for social science*. Cambridge University Press
- Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for data science: import, tidy, transform, visualize, and model data*. 2nd ed. O'Reilly Media, Inc.
  Free online version available at `https://r4ds.hadley.nz/`

We will be using RStudio for the programming portion of the course. You can get started by installing R and RStudio on your computer. Next, you can work through RStudio's primers, a set of interactive tutorials that will help you familiarize yourself with basic programming concepts and R.

III Schedule

Date	Topics	Gailmard 2014	Wickham et al. 2023
11/09	Course Logistics & Introduction		
18/09	Descriptive Statistics	chs. 1 & 2	ch. 1
25/09	Introducing Linear Regression	chs. 1 & 2	chs. 2 & 3
02/10	Data-Generating Processes	ch. 3	chs. 4 & 5
09/10	Probability Distributions	ch. 4	chs. 6 & 7
16/10	Expectation, Variance, Covariance	ch. 5	chs. 8 & 9
23/10	Probability and Models	ch. 6	chs. 10 – 12
30/10	Sampling Distributions and the CLT	ch. 7	chs. 13 – 16
06/11	Hypothesis Testing	ch. 8	chs. 17 – 20
13/11	Maximum Likelihood Estimation	ch. 9	chs. 26 – 28
20/11	Causal Inference	ch. 10	chs. 29 & 30
27/11	Review Session		

Note: Schedule may be subject to change depending on our progress during the semester.

Note on Assigned Readings: Additional required and/or recommended readings will be available on Aula Global. You should complete the readings associated with a given topic before each class session.

IV Evaluation

Your final grade will be determined based on the following three components:

1. **Problem sets ($10\% = 5 * 2\%$):** As part of this course, you will work on bi-weekly problem sets. I strongly encourage you to work in groups and discuss each question with your peers. However, each student must write up and submit their own original solution.
2. **Final exam (40%):** The final exam will test you on all the material covered throughout the semester. It will focus on the theoretical questions related to characterizing and analyzing data-generating processes. The assignments in Gailmard (2014) are a great resource to prepare for this exam.
3. **Research project (50%):** At the end of the semester, you are expected to submit a research project that utilizes the methods and techniques covered throughout the course. While you have to incorporate an original data analysis using R, you are free to choose any topic and/or data source you find interesting (and it may overlap with your other substantive coursework). Further details will be discussed in class.

V Additional Resources

As we work through the course material, some of you may want additional information on the underlying mathematical concepts, while others want to dig deeper into programming. Here is a list of additional textbooks that you might find helpful in either case:

	Mathematics / Statistics	*Programming / R*
Recommended	Angrist and Pischke (2008) Wooldridge (2013)	Verzani (2014) Urdinez and Cruz (2020)
Optional	DeGroot and Schervish (2012) Casella and Berger (2021)	Teetor (2011) Fox and Weisberg (2018)

There are countless other resources available online, but I want to highlight two great sets of YouTube videos in case you want to learn more about specific topics covered in our course. You'll find links to these videos on Aula Global as well:

- Danielle Navarro's introductory videos on R, RMarkdown, and the tidyverse: https://www.youtube.com/c/DanielleNavarro77/playlists.
- Andrew Heiss's lecture videos on Causal Inference and Data Visualization: https://www.youtube.com/c/AndrewHeiss/playlists.
- Gary King's lecture videos on quantitative social science methods: https://www.youtube.com/channel/UCtrwX29xpuWc9y0-0PKrHSQ/playlists https://projects.iq.harvard.edu/gov2001
- Lastly, I have taught courses similar to this one at the University of Wisconsin-Milwaukee in the past. You can find my old lecture videos—which cover a lot of the same topics—here: https://www.youtube.com/channel/UCmXfZJxXiwypm7f0vnKO_DA/playlists

VI Acknowledgements

I have adapted the ideas and language from the work of several educators for this syllabus and the course material. For example, I have borrowed liberally from other courses on social science research methods and statistics, as taught by Sean Gailmard, Kosuke Imai, Gary King, Michael Peress, Thomas Gschwend, and others. I appreciate their contributions to the discipline and thank all educators who make their teaching material available to others. To pay it forward, I will share my own material with anyone who is interested.

References

Angrist, Joshua D, and Jörn-Steffen Pischke. 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Casella, George, and Roger L Berger. 2021. *Statistical inference*. Cengage Learning.

DeGroot, Morris H, and Mark J Schervish. 2012. *Probability and statistics*. Pearson Education.

Fox, John, and Sanford Weisberg. 2018. *An R companion to applied regression*. 3 ed. Sage Publications.

Gailmard, Sean. 2014. *Statistical modeling and inference for social science*. Cambridge University Press.

Teetor, Paul. 2011. *R cookbook*. O'Reilly Media, Inc.

Urdinez, Francisco, and Andres Cruz. 2020. *R for Political Data Science: A Practical Guide*. CRC Press.

Verzani, John. 2014. *Using R for introductory statistics*. CRC Press.

Wickham, Hadley, Mine Çetinkaya-Rundel, and Garrett Grolemund. 2023. *R for data science: import, tidy, transform, visualize, and model data*. 2nd ed. O'Reilly Media, Inc.

Wooldridge, Jeffrey M. 2013. *Introductory econometrics: a modern approach*. Cengage Learning.