Applied Quantitative Methods for the Social Sciences I

Program: Master in Social Sciences, Fall 2022
Room: 18.1.A04
Time: Mondays, 10:00–13:00

Contact Information

Instructor: Patrick Kraft, PhD
Office: 18.2.C.07
Email: patrickwilli.kraft@uc3m.es
Office Hours: Mondays, 14:00–16:00, 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 math camp 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 is one required textbook for the course:
Gailmard, Sean. 2014. *Statistical modeling and inference for social science*. Cambridge University Press

In addition, we are going to use the following resources that are available online:

- Math Prefresher for Political Scientists (King et al., 2020): `https://iqss.github.io/prefresher`
- The Plain Person's Guide to Plain Text Social Science: `http://plain-text.co`
- R for Data Science (Wickham and Grolemund, 2016): `https://r4ds.had.co.nz`
- Data Visualization (Healy, 2018): `https://socviz.co`
- Fundamentals of Data Visualization (Wilke, 2019): `https://clauswilke.com/dataviz/`
- Introduction to Modern Statistics: `https://openintro.org/book/ims/`

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	Core Readings
26/09	MathCamp: Linear Algebra Review	King et al. 2020, ch. 6
03/10	Course Logistics & Introduction to RStudio	King et al. 2020, chs. 7-10
10/10	Descriptive Statistics	Gailmard 2014, chs. 1 & 2
17/10	Introducing Linear Regression	Gailmard 2014, chs. 1 & 2
24/10	Data-Generating Processes	Gailmard 2014, ch. 3
07/11	Random Variables and Probability Distributions	Gailmard 2014, ch. 4
14/11	Expectation, Variance, Covariance	Gailmard 2014, ch. 5
21/11	Probability and Models	Gailmard 2014, ch. 6
28/11	Sampling, Sampling Distributions, and Central Limit Theorem	Gailmard 2014, ch. 7
05/12	Hypothesis Testing	Gailmard 2014, ch. 8
12/12	Interval/Point Estimation & Maximum Likelihood	Gailmard 2014, ch. 9
19/12	Causal Inference	Gailmard 2014, ch. 10

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. Whether you complete readings before the associated lecture session or after is up to you. You should complete them in relatively close to the class session associated with a given topic in the readings. Most of the class sessions will consist of lectures, and mostly on theory, concepts, and simple examples. In general, readings will cover more material than the associated lecture. Despite the lecture format the nature and goals of this class require active involvement, discussion, and participation from students.

IV 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

V Evaluation

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

1. **Two take-home exams ($20\% = 10 * 2\%$):** As part of this course, you will work on 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. During the semester, you may select two of your weekly problem sets to count as take-home exams by submitting your individual solutions via email before the start of our weekly lecture.
2. **Final exam (30%):** 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.

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.

Healy, Kieran. 2018. *Data visualization: a practical introduction*. Princeton University Press.

King, Gary et al. 2020. "The Harvard Math Prefresher for Political Scientists." https://iqss.github.io/prefresher/.

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, and Garrett Grolemund. 2016. *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media, Inc.

Wilke, Claus O. 2019. *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media.

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