POLI SCI 403: Probability and Statistics

Fall 2024

Instructor: Gustavo Diaz
Time and Place: Tues/Thurs 11:00am – 12:20pm, room TBD
Office Hours: TBD
GitHub Repository: TBD

Teaching Assistant: TBD

Course Overview

This is the first course in the required two-quarter quantitative methods sequence for graduate students in the Department of Political Science. The course focuses on statistical inference in quantitative social science. Topics include probability theory, statistical inference from random samples, estimation theory, linear regression, maximum likelihood estimation, and nonparametric identification.

Learning Objectives

The goal of statistical inference is to use the data we have to learn about something for which we do not have data. That connection requires making assumptions. This course aims at introducing tools and developing skills to do so with as few assumptions as possible.

By the end of the course, you will be able to apply statistical methods to conduct your own analyses, explain statistical tools and concepts in your own words, evaluate the credibility of applied and methodological work, and continue learning more advanced methods.

Prerequisites

There are no formal requirements to take this course other than enrollment in the Political Science PhD program or express approval to enroll.

This course does not assume prior training on statistics or quantitative methods beyond a tenuous grasp of basic concepts in algebra and calculus covered US high schools or math camp. For example, you know that integrals can be used to calculate the area under a curve,

but you do not need to remember how to do integration by hand. I do assume you know how computer file systems work.

I expect you to participate actively, productively, and respectfully in our meetings. Some of the material addresses complicated concepts or uses math extensively. I do not expect you to understand every single equation for this course, but I do expect you to read carefully enough that you would understand every equation if you chose to revisit the material after taking this course.

Requirements

Reading

The main textbook we will follow is:

> **Textbook**
>
> Aronow, P.M. and Benjamin T. Miller. 2019. *Foundations of Agnostic Statistics*. Cambridge University Press

The rest of the syllabus refers to this book as AM. You can purchase a physical or digital copy directly from the publisher, although it's usually cheaper in storefronts like Amazon.

The book tends to err on the side of brevity and mathematical rigor. Much of our class discussion, assignments, and additional reading will involve untangling and applying the topics in AM. Additional readings should be available through university library subscriptions or distributed promptly otherwise. The digital version of the syllabus will have URLs for additional readings in the Schedule section.

The final section of this syllabus includes recommended readings that may be useful to complement current or future learning.

Computing

We will use R and RStudio to work on lab assignments and the final research note. The advantage of R is that it is free and open source, meaning that you will be able to apply everything you learn in this course anywhere else. The disadvantage is a somewhat steep learning curve. I believe the investment is worthwhile for anyone working with data or in data-adjacent careers.

You can install R and RStudio on your personal computer, which is the preferred workflow. You can use this link for installation instructions on Windows and MacOS. See this link for installation instructions on Chromebooks, which is a bit more involved.

You can also use Posit Cloud to access RStudio from any web browser. A free account should be sufficient for the purposes of this course and has the advantage of letting you access your work across devices.

If you ever need more computing resources than what a personal computer or a free Posit Cloud account allows, you should consider requesting access to the Quest Analytic Nodes from Northwestern IT. I do not anticipate this to be relevant for this course, but it may be useful for future reference.

Evaluation

Your final grade in this course will depend on the following:

- Participation
- Lab assignments (8 total, due Mondays 8 AM?)
- Research note (due December 11)

Participation

Lab assignments

Research note

Grading

Schedule

Week 1 (September 24/26): Preliminaries

Reading:

Lab 0:

Week 2 (October 1/3): Probability theory

Reading:

Lab 1:

Week 3 (October 8/10): Summarizing distributions

Reading:

Lab 2:

Week 4 (October 15/17): Random samples

Reading:

Lab 3:

Week 5 (October 22/24): Regression

Reading:

Lab 4:

Week 6 (October 29/31): Parametric models

Reading:

Lab 5:

Week 7 (November 5/7): Missing data

Reading:

Lab 6:

Week 9 (November 12/14): Causal inference I

Reading:

Lab 7:

Week 10 (November 19/21): Causal inference II

Reading:

Lab 8:

Week 11 (November 26): Flex week/future directions?

Reading:

Week 12 (December 3/5): Research note presentations

Academic Integrity

Accessibility

Generative AI

Recommended readings

Statistics

Computing