Gov 391J: Statistical Analysis in Political Science, I

Fall 2022 ∗ Mondays & Wednesdays, 10:00-11:30 am ∗ CBA 4.336

Connor T. Jerzak
Assistant Professor of Government
University of Texas at Austin

Social scientists seek to understand the social world, and this understanding is often critically based on the quantitative analysis of data. In this course, we introduce students to this mode of inquiry. We will explore classical approaches for bringing evidence to bear on empirical claims, examining their strengths, limitations, theoretical properties, and manner of implementation. Modeling strategies introduced will include probability models, linear regression, and some introductory causal inference approaches. Emphasis will also be placed on the effective communication of quantitative evidence. Therefore, we will emphasize throughout the course the importance of visualization methods, model sanity checks, and the qualitative meaning of the quantitative assumptions employed by various modeling strategies. By the end of the course, students should be able to tackle challenging empirical problems in their research with a variety of modeling strategies and to effectively communicate the implications of their analyses to their target audience.

A steel construction worker high up on the Empire State Building during construction (∼1931).

1 Contact Information

Instructor
NAME: Connor T. Jerzak
EMAIL: connor.jerzak@austin.utexas.edu
URL: ConnorJerzak.com
OFFICE: BAT 3.108
VIRTUAL OFFICE: utexas.zoom.us/my/jerzak
OFFICE HOURS: TBD tinyurl.com/CJerzakOH
or send me an email to make an appointment.
Feel free to drop by my office anytime as well:
As long as I'm there and the door is open, I'm happy to talk anytime.

Teaching Assistant
NAME: Matthew J. Martin
EMAIL: mjmartin@utexas.edu
URL: MattJamesMartin.github.io
OFFICE HOURS: Thursdays, 12:30-3:30 pm, BAT 1.118
WORKSHOPS: TBD

2 Prerequisites

By far the most important prerequisite for the course is a desire to learn the material and willingness to embrace the feeling of discomfort that can arise when learning new things. Otherwise, it would be helpful to have some working familiarity with

- probability theory
- data analysis using R
- LaTeX and Rmarkdown

Having said that, we will not assume existing knowledge and core ideas will be introduced from the foundations on up.

3 Instructional Tools

We will use a variety of tools to run this course.

- **Course website** (utexas.instructure.com/courses/1345560): Canvas will contain all course information related to the course. We will use the "Discussions" feature in Canvas to encourage opportunities for students to help each other learn by answering questions other students pose about the material and about problem setse.
- **Gradescope**: Gradescope is the online platform you will use to submit and receive feedback on your assignments. You can access our Gradescope page via Canvas.

4 Course Framework

The course has four main components: in-class lectures/discussions, section, weekly exercises done in pairs, an individual mid-term exam, and individual final exam. We believe it is important to have multiple ways in which students can be evaluated since some students may learn best with each modality (e.g., in group work, in solo work, in project-based work that will encompass some of the exercise questions).

Some guidelines.

- **Class Meetings**: In class meetings, key material will be introduced. We will typically begin class with a lecture. A break will occur at the class half-way point. After the break, we will usually have a more practical tutorial. We ask that students observe the following guidelines during class:

    1. *Be on time.* Late entry of students can be disruptive to the focus of students who arrived on time. There may be cases where, due to external circumstances, students may need to regularly arrive late; in that case, let the instructors know and we will think of a plan to minimize the degree of class disruption.
    2. *No computer or cell phone use during lecture components.* Computers can be a useful aide to learning in some situations, but evidence has accumulated that they can be a source of distraction for those around you in others. Therefore, use of laptops and cell phone during the lecture component of the course is not allowed.

- **Exercises**: Exercises will be done in partnership with a randomly selected peer pair. These pairs will be rotated every exercise. You and your partner will work together to answer the questions. Submissions are to be made to Gradescope by 5 pm Sunday. See next section for additional details.

- **TA workshops**: The TA will review critical course material in monthly workshops. We will be seeking feedback from students regarding what material to emphasize in these workshops so as to best assist your learning. More details will be released as the semester proceeds.

5 Course Requirements

The course grade is based on the following components:

- **Class participation** (10% of the course grade): The class participation portion of one's grade is important but can sometimes feel ill-defined. By "participation," we mean engagement broadly defined: Does a student attend class? Is the student engaged in class or section? Is the student helpful to others on the discussion page? Does the student treat other students with courtesy and respect in collaborations or discussions? The later aspect of the grade is not meant to be punitive; in your career, your success will be affected by your ability to work well with others, so this aspect is also a part of the course evaluation.

- **Exercises** (40% of the course grade): Students will be required to complete exercises on a weekly basis. You will be randomly assigned pairs. In these pairs, you will expected to support each other. The motivation for this pairing is to encourage students to (a) get to know a broad set of other students while (b) helping each other in your understanding of the material. The use of weekly exercises also encourages regular engagement with the material, something critical to learning. The following guidelines will apply to these exercises:

- *Collaboration policy.* Students are allowed to discuss the exercises with the instructional staff and any other students in class in addition to your assigned pair. After all, teaching one's peers can be helpful in learning the material better oneself. However, you and your assigned partner are required to write up your own answers. You are not allowed to copy the work of other teams. Copying others' work deprives you of the opportunity to learn the material; if you feel like you are struggling in the course, this information is very important for the instructional staff to know so we can support you. If you feel like you are struggling in the course, it is better to let us know sooner rather than later so we can identify impactful solutions.
- *Online help and office hours.* You are strongly encouraged to reach out to the instructional staff through and office hours about any questions you might have about the course materials. Students should also feel free to ask questions and answer the questions posed by others at , which will be considered in the overall evaluation of class participation.
- *Submission guidelines.* Answers to all questions should be incorporated into a single Rmarkdown (or LaTeX) file. The instructional staff will provide an Rmarkdown template. For the exercises, each student should submit the PDF file electronically to Gradescope. Once you upload the PDF file, you will see a list of the questions in the assignment and thumbnails of your file. For each assigned question, click the PDF page(s) that contains your answer. No late submission will be accepted unless you obtain a prior approval from the instructor.

- **Mid-term take-home exam** (20% of the course grade): There will be a mid-term take-home exam including some analytical and some empirical questions. The motivation for having an exam is that the existence of an individual evaluation event can be a good stimulus for students to hone in on learning the material in a more focused way than in the weekly exercises. You may not discuss the exam with anyone else. You are allowed to ask the instructional staff clarifying questions. You will have 24 hours to complete the exam once you've opened it. We will give more detailed instructions before the exam is released.
- **Final exam** (30% of the course grade): The final exam will be similar in nature and scope to the mid-term.

6 Course Materials

There is no single textbook for this course. This is by design: a given textbook tends to emphasize one perspective (reinforced in the notation), but it is useful to have a mental repository for multiple representations of the same thing. Each representation will have weakness in some situations and strengths in others. We will therefore draw from the following set of texts:

(Note that the texts we will draw most heavily on are freely available online; those which we will draw from less frequently will have that material made available via Canvas.)

ROS Gelman, Andrew, Jennifer Hill, and Aki Vehtari (2020). *Regression and Other Stories*. Cambridge University Press. https://avehtari.github.io/ROS-Examples/

FAS Aronow, Peter M. and Benjamin T. Miller (2019). *Foundations of Agnostic Statistics*. Cambridge University Press.

IP Blitzstein, J. K., & Hwang, J. (2015). *Introduction to probability*. Boca Raton, FL: Crc Press.

VIS Tufte, Edward R. (2001). *The Visual Display of Quantitative Information.* Graphics Press.

ECM Woodridge, Jeffrey (2013). *Introductory Econometrics: A Modern Approach.* 5th Edition. Cengage Learning.

HARM Angrist, Joshua D. and Jorn-Steffen Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

7 Course Plan

This course intends to provide an introduction to the following topics in probability, statistical inference, and quantitative social science.

Note that exercises are due every Sunday at 5 pm unless otherwise announced!

Introduction

Module 1 Introduction

Class 1 (Aug. 22) Overview of the course; why learn quantitative methods; goals of quantitative methods (statistical, predictive, causal, data processing)

Quantitative Modeling: Probability, Inference, Regression

Module 2 Modeling the World through Probability

Class 1 (Aug. 24) Random variables; essential concepts in probability (PMF, PDF, CDF, moments)

Reading
- FAS 1.1-1.2
- FAS 2.1

Class 2 (Aug. 29) Bayes' theorem, Simpson's paradox; conditional probability; joint random variables

Reading
- FAS 1.1-1.2 *(review)*
- FAS 1.3-1.4
- FAS 2.2-2.3

Class 3 (Aug. 31) Transforming random variables

Reading
- IP 8

Class 4 (Sept. 7) Simulating random variables, simulating social systems

Reading
- ROS 5

Interlude Data Visualization in Scientific Communication

Class 1 (Sept. 12) Discussing data visualization

Reading
- ROS 2
- VIS 1, 4-9 *(get the gist)*

Module 3 Learning about Probability from Data

Class 1 (Sept. 14) Learning about probabilities from data; random sampling; estimation

Reading – FAS 3.1

Class 2 (Sept. 19) Statistical inference: Limit theorems

Reading – FAS 3.2

Class 3 (Sept. 21) Statistical inference: Hypothesis testing

Reading – ROS 3.3-3.4

Class 4 (Sept. 26) Statistical inference with dependent data

Reading – ROS 3.5

Module 4 Doing, Understanding & Evaluating Regression

Class 1 (Sept. 28) Doing regression: regression as an optimization problem

Reading – ROS 6-8 (skim 6 if pressed for time)

Class 2 (Oct. 3) Understanding regression I: BLU-ness and asymptotic inference

Reading – HARM 3.1.3-3.1.4

– HARM 3.2

Class 3 (Oct. 5) Understanding regression 2: BLU-ness and asymptotic inference

– ECM 8

– ROS 22.3

Class 4 (Oct. 10) Non-parametric approaches to uncertainty in regression (and beyond)

– Review ROS 5.4

– Kreiss, J. P., & Paparoditis, E. (2011). Bootstrap methods for dependent data: A review. *Journal of the Korean Statistical Society*, 40(4), 357-378.

– Wehrens, R., Putter, H., & Buydens, L. M. (2000). The bootstrap: a tutorial. *Chemometrics and intelligent laboratory systems*, 54(1), 35-52. *(optional)*

Class 5 (Oct. 12) Evaluating regression: learning not to fool yourself; data transformations

Reading – ROS 12

– VIS 2-3 *(optional)*

Class 6 (Oct. 17) *No class due to mid-term*

Mid-term Released on Oct. 12

Mid-term Due by Oct. 19

Cumulating Quantitative Evidence... ... in the Wild

Module 5 Data Analysis in the Wild

Class 1 (Oct. 19) Poststratification

Reading – ROS 17.1-17.2

Class 2 (Oct. 24) Missing data

Reading – ROS 17.3-17.6

Class 3 (Oct. 26) When is regression appropriate for (trying) causal inference?

Reading – ROS 18

Class 4 (Nov. 2) Non-linearizing regression

Reading – FAS 4.3

Class 5 (Nov. 9) Time series modeling & panel data

Reading – HARM 5

Module 6 Cumulating Quantitative Evidence

Class 1 (Nov. 14) Pre-analysis work I: Pre-registration, replicability

Reading
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS medicine*, 2(8), e124.
- Harper, S. (2019). A future for observational epidemiology: clarity, credibility, transparency. *American journal of epidemiology*, 188(5), 840-845.
- P Simmons, J., D Nelson, L., & Simonsohn, U. (2021). Pre-registration: Why and how. *Journal of Consumer Psychology*, 31(1), 151-162.

Class 2 (Nov. 16) Pre-analysis work II: Power

Reading – ROS 16

Class 3 (Nov. 28) Placebo tests

Reading
- Eggers, A. C., Tuñón, G., & Dafoe, A. (2021). Placebo tests for causal inference. https://pelg.ucsd.edu/Eggers_2021.pdf
- FAS 7.2.8

Class 4 (Nov. 30) Post-analysis work: Connecting theory & empirics; interrogating robustness

Reading
- Alt, James. "Empirical Implications of Theoretical Models." *Political Analysis* 11.4 (2003).
- Lu, X., & White, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of econometrics*, 178, 194-206.

Conclusion

Module 7 Wrapping Up

Class 1 (Dec. 5) Review; residual topics; Q&A

– TBD

Final exam Released on Dec. 9 (tentative)

Final exam Due by Dec. 12 (tentative)

8 Other Resources

There are other resources that could be useful to students in different aspects of the course (or graduate school more generally), such as when giving presentations or when undertaking the creative process of formulating a research project.

On giving presentations:

* Winston, Patrick."How to Speak." youtube.com/watch?v=Unzc731iCUY

On the creativity:

* De Bono, Edward. *Laternal Thinking: Creativity Step by Step*. Harper Colophon.

On writing:

* Pinker, Stephen. *The Sense of Style: The Thinking Person's Guide to Writing in the 21st Century*. Penguin Books.
* Williams, Joseph and Joseph Bizup. *Style: Lessons in Clarity and Grace*. Pearson.

9 Arrangements

Students with arrangements through the UT Disability & Access Services Office should ensure that those arrangements are communicated to the instructor so we can set up a constructive environment for you.

10 Integrity

"The core values of The University of Texas at Austin are learning, discovery, freedom, leadership, individual opportunity, and responsibility. Each member of the University is expected to uphold these values through integrity, honesty, trust, fairness, and respect toward peers and community."

— UT Austin Code of Conduct

"As a student of The University of Texas at Austin, I shall abide by the core values of the University and uphold academic integrity."

— UT Austin Student Honor Code

11 Acknowledgements

I am grateful to Matthew Blackwell, Adel Daoud, Frances Hagopian, Kosuke Imai, Gary King, Michael Peress, Xiaolong Yang, and Xiang Zhou; the planning of the course builds on conversations or prior teaching experience with them.

12 Final Reflections

The instructors reserve the right to modify aspects of the course as it proceeds, both in response to student feedback and our own assessment of how to improve the course.