

## DATA IN POLITICS II: FRONTIERS AND APPLICATIONS

This draft: April 11, 2021 (Check Sakai for the latest)

Poli381, Spring 2021

Classroom: The Cloud

Meeting time: Tuesdays and Thursdays, 2:00 to 3:15

Professor Timothy Ryan; [tjr@email.unc.edu](mailto:tjr@email.unc.edu); 919-962-0403

Office: Hamilton 307 (but really Zoom)

Office Hours: Wednesdays, 1:30-4:30 pm, or by appointment. See Sakai for a link to reserve a timeslot during the default period, or email me for an appointment outside that period.

Final exam: Thursday, May 13 at noon.

Course website: Sakai

The Information Revolution has dramatically expanded the volume of information we have about the world around us. Social scientific analytical skills are transforming many sectors—business, journalism, law, public policy, health care, and finance, to name but a few—and are more valued now than ever. Poli381 is designed to provide students who have developed an initial familiarity with statistical concepts and skills with further tools they need to be leaders in data-driven sectors. More specifically, the learning objectives are for students:

- 1) To develop comfort using the linear model to make inferences in a variety of social scientific settings.
- 2) To learn statistical concepts used to characterize the degree of uncertainty in estimates and inferences.
- 3) To increase their comfort understanding quantitative procedures used in formal technical reports (including, but not limited to, academic journals in the social sciences).
- 4) To comprehend the limitations of quantitative inference.

Students will leave this class with a solid foundation to employ quantitative techniques in the private sector, as well as in more advanced academic settings (such as work on a senior honors thesis). The target audience for this course is undergraduate students with interest in the social sciences (not only Political Science), who want to use quantitative approaches to solve important problems and develop marketable analytical skills.

The prerequisite for this course is *Poli281: Data in Politics I: An Introduction*, or equivalent experience in the R statistical language. Please speak to the instructor if you are uncertain whether or not you have the necessary foundation.

This course fulfills the Quantitative Intensive (QI) requirement in UNC's Making Connections curriculum. It also counts as an SS (Social and Behavioral Sciences) course. It also counts as a course in research methods (required for completing the Political Science Bachelor's degree).

### Requirements and Evaluation

Course assessment is broken down as follows.

ASSIGNMENTS (50%) There will be six main assignments due as noted in the schedule below. Assignment 0, due very early in the semester, is graded pass/fail. You need to do it, but it does not affect your final course average. Otherwise, assignments are weighted equally. Written assignments are due at 6pm on the days indicated.

In addition to these “main” assignments, I also assign occasional small homework assignments, such as to finish up something we couldn’t quite finish in class. When such things come up, they count toward your participation grade, or for extra credit, depending on the circumstances.

FINAL PROJECT (35%, broken up as follows) The class has a capstone final project for which students, working in groups, conduct and present an original data analysis on an existing dataset. This project consists of a paper (30%) and a presentation (5%). A separate document will specify final project requirements more completely.

PARTICIPATION (15%) Your participation grade has two components. The first component is preparation for (and actual attendance in) class. At each class meeting, you will give yourself 0 to 3 points, depending on how prepared you are for class and how much you intend to participate. (0 = absent from class; 1 = attended class but did not prepare; 2 = attended class but preparation was partial or rushed; 3 = attended class, prepared with care, *and commit to being proactive in asking questions and contributing to discussion.*) The semester-long score generated by these reports is advisory to the instructor-assigned participation grade—I can adjust self-reports that are inconsistent with what I see in class—but I take them seriously.

The second component is participation in our class’s online Piazza discussion forum. (See Sakai for a link.) On Piazza, you can benefit your participation grade **either** by posting your own questions or by providing thoughtful answers to other students’ questions. (The website keeps track of your activity.) These statistics, too, are advisory to the summary participation grade.

Because attendance is a course requirement (see the Undergraduate Bulletin) and critical to having a lively conversation, more than three unexcused absences can have a negative effect on your grade—potentially a drastic one—outside the scheme above. (That is, it can affect more than just your 15% participation grade.) **If you have missed more than three classes, it is your responsibility to come and talk to me about what’s going on.**

Note that, the weighting scheme above notwithstanding, all assignments must be submitted. If you have not submitted an assignment, you are at risk of receiving a course grade of I (Incomplete) until it is complete.

## Course Texts

This course has one required textbook:

- Imai, Kosuke. 2017. *Quantitative Social Science: An Introduction*. Princeton: Princeton University Press.

It also has two textbooks that are recommended, but not required (certain key selections will be made available on Sakai). The second of these is free, open-access online. (At least for now.)

- Pollock, Philip H. and Barry C. Edwards, *The Essentials of Political Analysis*. New York: Sage University Press.

- Hanck, Christoph et al. *Introduction to Econometrics with R*. Available online at <https://www.econometrics-with-r.org/>

In addition, there are software requirements. Students must download and install R, a free statistical program available at <http://cran.r-project.org/>, as well as RStudio (also free), which is available at [www.rstudio.com](http://www.rstudio.com).

## Other Policies

*Technology.* The class is taught on Zoom, so obviously you can use your computers. That said, please understand how important it is to me that you eliminate distractions and be mentally “present.” It makes such a difference for the quality of discussion, enjoyability of the class, and how much we all learn. Barring some personal emergency, I promise to close my email program and mute all notifications during class, and I expect you to do the same.

*Email.* I usually respond to emails within 24 hours. However, please limit your use of email to issues that are private, or at least specific to you. For matters that are not private and where other students might want to see my response, please use Piazza (see above). Note that Piazza permits anonymous posting. (The post will be anonymous to other students, but not to me.)

I typically will not use email to repeat information that was missed because of an absence; I’ll direct you to correspond with a classmate.

*Cooperation and academic integrity.* In a class setting, cooperative work has both benefits and pitfalls. Peers learn a lot by explaining things to each other. But it can also be easy to stumble into a passive mindset where you’re not really *assimilating* the concepts. To strike a balance, I will designate some activities and assignments (or parts thereof) as being Cooperative, and others as Individual. It is critical that you attend to this distinction, as completing individual work cooperatively would be a breach of academic integrity.

By its nature, this class has an extra matter we need to address. While discussion with other people is permitted and encouraged for work designated as cooperative, there is a distinction between discussing a problem and copying someone else’s work. (Writing computer code is an especially tempting activity for which to copy work.) Students can discuss problem-solving strategies, clarify concepts, and point out mistakes—but ultimately each person must generate his or her own path to the solution. *In our class, copying and pasting another person’s computer code is potentially tantamount to plagiarism. **Even for work designated as cooperative, you must write your code individually.*** Unless I have given you explicit permission for some special reason, do not do it. Copied code is surprisingly easy to detect (there is software designed to detect it). Be assured that if I identify a case of cheating or plagiarism, I will handle it 100% “by the book.”

*Students with disabilities.* If you think you need an accommodation for a disability, please let me know. Some aspects of the course and its assignments may be modified to facilitate your success. I will work with the Office of Accessibility Resources and Services to determine appropriate accommodations. I will treat any information you provide as confidential. Barring unusual circumstances, I require notice of a need for accommodation within the *first two weeks* of the semester.

*Grade grievances.* Requests for regrades have a time window. They cannot be submitted until at least 48 hours have passed since the assignment was returned (a cool-down period), and then they will only be

accepted within three weeks of an assignment being returned (a statute of limitations). To request a regrade, you must submit a written memo (two pages max) explaining what aspect of your original grade you think was in error.

*Absence on critical days.* Generally, absences on exam days will only be excused for reasons of religious observance, illness, or family emergencies. All requests for an excused exam absence must be submitted and approved by email (for documentation). The request must come as far in advance of the absence as possible if the absence is foreseen, or as soon after as possible if it is not foreseen. Per university policy, only your academic advisor can provide an official final exam excused absence.

Additionally, this course includes several class periods focused on workshopping final projects. Absences on these days are especially disruptive, since they interfere with your group's ability to work on the final project together. For this reason, absences on days marked with an asterisk below require special permission, or your individual final project grade will automatically receive a 5-point penalty.

*Schedule changes:* This course has only been taught once before (and in a semester interrupted by the Covid-19 pandemic). There will likely still be a few miscalculations about how long each topic and activity will take. Please be ready for the schedule below to change, though the workload (the number of assignments) will not. Additionally, I will try especially hard to avoid changes to any of the "critical days" (see above).

## Course Schedule

### *Part 1: The Friendly, Flexible Linear Model*

January 19: Course Introduction

- Read the syllabus

January 21: Linear Regression: A Fresh Pass

- Give an *initial* read to Imai, chapter 4 and Pollock, chapter 8. This is a lot of material and you should not expect to absorb most of it on this read. Rather, I want you to be familiar with which topics are covered where, so you can refer to the appropriate sections of the text as they come up below. Think of this reading assignment as brushing off the dust and building up a mental table of contents you will use for the next few weeks.

January 26: Multivariate Regression: Incorporating Categories

- Review relevant sections of Imai and Pollock. (Especially Pollock 260-263.)
- Assignment 0 due

January 28: Multivariate Regression: Collinearity and Statistical Control

- Review relevant sections of Imai and Pollock

February 2: Multivariate Regression: Using Residuals and Predicted Values

- Review relevant sections of Imai and Pollock

February 4: Multivariate Regression: Transformations of the IV

- Review relevant sections of Imai and Pollock

February 9: Multivariate Regression: Prediction vs. Causality

- Review relevant sections of Imai and Pollock
- Read “[Prediction vs. Causation in Regression Analysis](#).” (Short blog post by Paul Allison.)
- Read the Assignment 1 article assigned to your group

February 11: Flex Day

February 16: University-designated Wellness Day (No class)

February 18: Multivariate Regression: Interaction Terms

- Review relevant sections of Imai and Pollock

February 23: Press Secretary Presentations 1.1 (aka Assignment 1)\*

February 25: Press Secretary Presentations 1.2 (aka Assignment 1)\*

*Part 2: Characterizing Uncertainty*

March 2: Interaction Terms 2.

March 4: Probability 1: Conditional Probability

- Imai, pp. 242-265. Read pp. 242-253 for context and to get a sense of relevant terms, but give greatest attention to pp. 254-265, which covers the more important concepts.

March 9: Probability 2: PDFs and CDFs

- Imai, pp. 277-292
- Assignment 2 due

March 11: University-designated Wellness Day (No class)

March 16: Probability 3 Expectations and Large Sample Theorems

- Imai, pp. 292-306

March 18: Uncertainty 1: Standard Errors

- Pollock, pp. 167-183 (Sakai)
- Imai, pp. 314-326

March 23: Uncertainty 2: Confidence Intervals

- Pollock, pp. 184-194 (Sakai)
- Imai, pp. 326-332

March 25: Uncertainty 3: Hypothesis tests

- Pollock, pp. 199-215
- Imai, pp. 342-464; 370-389. (Skim 363-369. Useful stuff but I don't expect to dwell on it.)

March 30: Group work on Assignment 4 and Red Team Final Projects\*

*Part III: Applied Quantitative Analysis*

April 1: Press Secretary Presentations 2.1 (aka Assignment 4)\*

April 6: Press Secretary Presentations 2.2 (aka Assignment 4)\*

April 8: How to reshape data

- [Reshaping Your Data with Tidy](#)
- Assignment 3 due

April 13: Red Team Workshopping\*

April 15: When the Linear Model Breaks—and How to Maybe Possibly Fix It (Pt. 1)

- [7 Classical Assumptions of Ordinary Least Squares \(OLS\) Regression](#)

April 20: Red Team Workshopping\*

April 22: Flex Day: Activities to be determined

- Assignment 5 due

April 27: Red Team Workshopping\*

April 28: Upload first draft of your Red Team report to Sakai, and circulate it to other Red Team groups, by 6pm.

April 29: Red Team Workshopping\*

May 4: Review and conclusions

- Upload comment on other teams' reports to Sakai by 6pm. You should also send the comment document to the other teams directly.

May 13 at noon: Final exam

- Final draft of Red Team report, along with final analytical files, due at noon.