Political Science 2580

QUANTITATIVE RESEARCH METHODS I

Paul Testa	Olgahan Cat	Aidong (Eden) Li
Assistant Professor	Visiting Assistant Professor	Teaching Assistant
paul_testa@brown.edu	olgahan_cat@brown.edu	ruodan_xu@brown.edu

First Draft: August 29, 2017

This Draft: September 3, 2025

General Information

Class website https://pols2580.paultesta.org

Canvas https://canvas.brown.edu/courses/1100351

Where/When We meet Thursdays, 4:00–6:30 pm, in 111 Thayer Room 140.

Who **Instructors:** Paul Testa paul_testa@brown.edu

Olgahan Cat olgahan_cat@brown.edu

TA: Aidong (Eden) Li aidong_li@brown.edu

Office Hours **Paul's Office Hours** Thursdays's from **1:00 - 3:00 pm** at 111 Thayer St Room 339. If possible, please reserve a spot here

Sections TBD

Overview

This course provides an introduction into the basics of quantitative empirical research in the social sciences. We explore three main types of statistical inference–causal, descriptive, and predictive. Next will we develop some core tools for statistical inference from probability and statistics and finally, we will apply what we have learned to an empirical research project. Each class will be divided into two parts. The first part will consist of lecture and discussion around the week's assigned topics. The second part of class will involve applied data analysis using `R` an open source program available for download here. We will work through examples and provide an opportunity get started on that week's work together. All data and readings not listed among the required texts are available on the course website: `https:://pols2580.paultesta.org` and Canvas, as are additional supplemental materials, such as slides, notes, and code.

Why Should You Take This Course?

The simple answer is because you're required to for your degree. Why this course is a degree requirement is a more complicated question. The goal of any graduate program in political science is to give students the skills and knowledge they need to conduct novel and insightful research. Unfortunately (or perhaps fortunately, depending on your view on these things), there is no single, tried and true path to reach this end.

The methods you need to know depend on the questions you want to answer. For some students–for example, those whose work involves large-N, empirical quantitative research–this course will provide an introduction and foundation to the skills and methods they will use for the rest of their careers. For others–perhaps those whose research will rely more on carefully done case studies or close readings of important texts–the immediate benefits of this training may be less clear. A typical answer is that this course will allow you to be "conversant with your peers." A brief review of the *American Political Science Review* or *American Journal of Political Science* will show that your peers tend to talk a lot about the statistical analysis of quantitative data. This is not a value statement, but an observation. Much of the research that gets published in top journals relies on sophisticated, carefully done, empirical research.

The real reason we require training in quantitative empirical methods is because it represents a way of thinking and a mode of research that can benefit all scholars. The logic of models and statistical inference forces us to clarify our questions, clearly state our assumptions, and sharpen our research designs. No amount of statistical magic can (or should) save a bad research design. The goal of teaching things like hypothesis testing and linear regression isn't to produce pretty tables with lots of coefficients and asterisks (although you'll learn how do to that). The goal is to make you humble, to teach you how hard it is to make statements with certainty and offer causal claims about anything in political science. Even if you never "run a regression" again after this class, learning all that goes into producing a result and what that result does (and does not) tell you about the world, will be useful to you in your careers.

Academic Integrity

Neither the University nor I tolerate cheating or plagiarism. The Brown Writing Center defines plagiarism as "appropriating another person's ideas or words (spoken or written) without attributing those word or ideas to their true source." The consequences for plagiarism are often severe, and can include suspension or expulsion. This course will follow the guidelines in the Academic Code for determining what is and isn't plagiarism:

> In preparing assignments a student often needs or is required to employ outside sources of information or opinion. All such sources should be listed in the bibliography. Citations and footnote references are required for all specific facts that are not common knowledge and about which there is not general agreement. New discoveries or debatable opinions

> must be credited to the source, with specific references to edition and page even when the student restates the matter in his or her own words. Word-for-word inclusion of any part of someone else's written or oral sentence, even if only a phrase or sentence, requires citation in quotation marks and use of the appropriate conventions for attribution. Citations should normally include author, title, edition, and page. (Quotations longer than one sentence are generally indented from the text of the essay, without quotation marks, and identified by author, title, edition, and page.) Paraphrasing or summarizing the contents of another's work is not dishonest if the source or sources are clearly identified (author, title, edition, and page), but such paraphrasing does not constitute independent work and may be rejected by the instructor. Students who have questions about accurate and proper citation methods are expected to consult reference guides as well as course instructors.

We will discuss specific information about your written work in class in more detail, but if you are unsure of how to properly cite material, please ask for clarification. If you are having difficulty with writing or would like more information or assistance, consult the Writing Center, the Brown library and/or the Academic Code for more information.

Generative AI Generative AI is an increasingly present aspect of our lives. Some uses of resources like ChatGPT are acceptable in this course others are not. For example, in your replication projects, you may use Chat GPT to help you translate replication code from one programming language to another. In our labs, you may not use ChatGPT to answer questions (i.e. write code for you), but you can use it as a resource to troubleshoot a particular coding error. You may not use ChatGPT in your written assignments.

Community Standards

All students and the instructor must be respectful of others in the classroom. If you ever feel that the classroom environment is discouraging your participation or problematic in any way, please contact me, either directly through email, or anonymously through our periodic class surveys.

Academic Accommodations

Any student with a documented disability is welcome to contact me as early in the semester as possible so that we may arrange reasonable accommodations. As part of this process, please be in touch with Student Accessibility Services by calling 401-863-9588 or online

Diversity and Inclusion

This course is designed to support an inclusive learning environment where diverse perspectives are recognized, respected and seen as a source of strength. It is my intent to provide materials and activities that are respectful of various levels of diversity: mathematical background, previous computing skills, gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Toward that goal:

- If you have a name and/or set of pronouns that differ from those that appear in your official Brown records, please let me know!
- If there are things going on inside or outside of class that are affecting your performance in class, please don't hesitate to talk to me, provide anonymous feedback through our course survey, or contact one of Brown's Academic Deans.

Course Goals

In the bestiary of the social sciences, methodological training typically follows either the path of the tortoise or the hare. There's no right way to run this race. Going slow and steady ideally provides you with a foundation to learn the methods you need to know. The danger of this approach is that you can spend so much time up front doing proofs and problem sets that you lose sight of why you wanted to obtain this training in the first place. Similarly, ranging far and wide can provide an overview of the toolkit available you, but without a strong foundation in the motivations and assumptions behind these methods, there's a risk that you'll end up using a expensive table saw when a simple wrench would have sufficed.

This course aims to strike a middle ground. To continue the (belabored) animal metaphor, we'll start off as hedgehogs, focusing on knowing a few things well: inference (descriptive, statistical, and causal), linear models (as a tool for inference), and extensions and alternatives to the linear model (to facilitate better inferences). By the end of the course, we'll be ready to blossom (mutate?) into methodological foxes capable of learning the many things skills and methods needed for our research.

Requirements

To accomplish this metamorphosis, we'll need the following:

- Some math
- Some programming and computing skills
- Some general life skills

Math You either already know, or will learn, all the math you need to take this course.[1] We'll go over some key theorems of probability and statistics in class, emphasizing conceptual understanding (often illustrated via simulation) over formal proofs.[2]. Along the way, we'll need some calculus and linear algebra to make our lives easier, and so we'll briefly review this material together in class.

[1] This is not the same as all the math you need to know be a successful, methodologically sophisticated political scientist. But it's a start, and one that will hopefully help you figure out what additional training you'll need.

[2] We'll do the proofs as well, but your focus should be on making sure you understand concepts and implications rather than specific derivations

Computing Doing quantitative, empirical social science research requires working with data. Today, working with data requires a computer and statistical software. I assume that you have, or will acquire, a laptop that you will bring to class. In terms of software, there are many possible options. In this class,`R`.[3].

All the slides, notes, and assignments in this class are produced using a combination of `R Markdown`, a programming framework that allows us to combine our statistical code (written in `R`) with `Markdown`, a light-weight markup language that can be converted into a range of formats (e.g. html, pdf, Word document). We'll also learn how to create documents in LaTeX, the standard document preparation system for academic manuscripts. All of this will seem like a bit much at the beginning of the course, but I promise the benefits far outweigh the costs.

General Like any course, success in this class requires preparation, participation and perseverance. In terms of preparation, I expect that you will have done the readings and submitted your assignments on time (more on that below). In short, you'll get out of this class what you put in. In terms of participation, I expect that you will come to class eager to learn and engage with that week's topics. If you have a question, ask it. If you're getting an error, share it. In some ways, your job is to make errors. To paraphrase Joyce: people of genius make no mistakes. Their errors are volitional and are the portals of discovery. While this experience can be challenging and frustrating, it is also incredibly rewarding. I fully expect you persevere through the problems and difficulties that inevitably arise in this course, and will do everything I can to help in this process.

Course Structure and Policies

Readings There is one required textbook for the course:

Imai, K. (2022). ***Quantitative Social Science An Introduction in Tidyverse*****. Princeton University Press**

The primary textbook on which the course is structured. Most chapters are spread over multiple weeks. You should read this text with your laptop and R Studio open. Execute the code in the main text and ideally try to complete the assignments and exercises at the end of the chapter. **Approximate Cost: $48.00**[4]

Additional readings will be listed below and available to download on Canvas.

Class Broadly the structure for each class is as follows.

Before each class on Thursday, you will have done the assigned readings for the week and submitted the prior week's lab (more on that below) to Canvas **by 11:59 pm on the Wednesday before class**.

On Thursday, I'll post solutions to the prior assignment which you will review before class.

During class on Thursdays day, we'll review the prior weeks' work, discuss the current week's topic and then get started together on that week's lab.

[3]Available for free at `https://cran.r-project.org/`. Python is also increasingly common.
[4]Estimated from Amazon

Whatever you don't finish in class you will be expected to complete and submit to the Canvas by 11:59 pm on the following Wednesday after class.

Course slides, assignments, comments, and supplemental material will be posted to Canvas.

Labs The bulk of the work and learning you'll do in the course comes in the form of weekly labs in which you'll explore a given data set or paper using R. You'll be given an R Markdown document that will guide you through a set of exercises to teach concepts covered in the lectures and reading. You'll code in `R` and summaries of your findings in R Markdown. You will compile your document to produce a pdf, which you will **submit on Canvas by 11:59 pm on the following Monday after class.**

All work in this class **MUST BE SUBMITTED ONLINE VIA CANVAS.**

You are expected to work in collaboration with your peers. You may share code and discuss your results, but each of you must submit your own file.

Assignments In addition to weekly labs, you will have periodic assignments the goal of which is to help you stay on track for writing your final paper.

The timeline of assignments for your final paper is as follows:

Week 5: Drafting Research Questions
Due Thursday, October 9, 2025 at 4:00 pm on Canvas

Week 7: Developing your proposal
Due Thursday, October 23, 2025 at 4:00 pm on Canvas

Week 10: Initial analyses
Due Thursday, November 13, 2025 at 4:00 pm on Canvas

Week 13: Final Paper Draft
Due Thursday, November 20, 2025 at 11:59 pm on Canvas

Week 14: Slides and/or poster presentation of final paper
Due Sunday, December 7, 2024 at 11:59 pm on Canvas

Week 15: Final Papers DUE at 11:59 pm Sunday, December 14, 2025 on Canvas

Assignments and labs must be submitted on time to Canvas. No late work will be accepted without prior approval of the instructor or a note from the university.

Grades Your final grade for this course will be calculated as follows:

- **40% Weekly labs**
- **35% Final Paper**
- **10% Class involvement and participation**
- **5% Final paper assignments (not including draft)**
- **5% Final Paper Draft**

- **5% Final Presentations**

Labs, assignments excluding the pre-analysis plan, and presentations, will be graded Each weekly assignments will be graded roughly on a ✓+ (100, completed on time, acceptable), ✓(85, completed on time, passable), ✓- (0 not submitted on time, unacceptable). Your draft, presentations, and final paper will be graded on 100-point scales with rubrics provided beforehand.

Time This course meets 14 times over the semester, including the last class that will be held during reading period. Each week, you should expect to spend 2.5 hours per week in class (35 hours total); approximately 2 hours per week reading and 3 working on labs and reviewing slides and notes (70 hours total); approximately 15 hours on assignments for the final paper; approximately 20 hours researching, writing, and revising your final paper; and at least .5 hours meeting with me in person to discuss your work (Estimated Total Time: 140.5 hours)

Class Structure

Question: The motivating question for the class

Topics: Specific topics discussed and skills learned

Read: Readings to be completed before class on Wednesday. Generally a chapter from Imai (2017), with a supplemental reading or two.

Lab: A brief description of that week's lab often with a citation for the paper or data on which the lab is based. Skimming the paper or data set before class is probably a good idea, and may become mandatory if it seems like an issue.

Unless noted otherwise below, all labs are **due the Wednesday after the class in which they are assigned** and MUST be **submitted online to Canvas by 11:59 pm.** No exceptions.[5]

Assignment: A set of cumulative assignments to help you write your final paper for this course.

Unless noted otherwise below, all assignments are **due the Wednesday after the class in which they are assigned** and MUST be **submitted online to Canvas by 6:29 pm.**

ICYI: "In case you're interested..."[6] A set of related readings you may find interesting and useful. Occasionally, I'll also post links to supplemental notes on course related topics.

[5] I reserve the right to make exceptions to this no exception policy, but in principle, the expectation is that you will turn in your assignments on time, to Canvas, in the appropriate format.

[6] A convention/tic borrowed from DFW (**?**)[a][b]

[a] David Foster Wallace

[b] Who as this well-intentioned but probably ill-advised descent into self-referential footnotes might suggest held a major spot in your author's various vanity bookshelves over time, most of which, for what it's worth, were carefully curated before our present age of countless think pieces, Twitter performance art and David Brooks name checks If you're still reading this exceedingly long footnote, send me a picture of a cute animal and I'll give you a point of extra credit for the course. See reading the footnotes pays off.

Schedule

Note: This schedule is preliminary and subject to change. If you miss a class make sure you contact me or one of your colleagues to find out about changes in the lesson plans or assignments. Each week contains the following information:

1 — September 4, 2025— Introduction and Course Overview

Question: What am I getting myself into?

Reminder: Bring laptops if you have them. We'll take some time at the end of class to make sure everyone's setup with R, *R Studio*

Read: Imai (2022) Chapter 1

Topics: Course overview. Basics of R, *R Studio*, and LaTeX.

Lab: None

Assignment: None

ICYI: Garfinkel (1981) is a really nice discussion of the kind of questions we ask (or think we ask) as social scientists

King et al. (1994) is a classic piece on bridging the supposed qual-quant divide

2 — September 11, 2025— Data and Measurement

Question: How do we describe the world around us?

Topics: Levels of measurement; Measures of Central Tendency and Dispersion, Organizing describing and data

Read: Imai (2022) Chapters 1 and 3

Wickham et al. (2023)Chapter 5 and Chapter 12

Lab: Exploration of Covid-19 data

Wednesday, September 17, 2025 at 11:59 pm on Canvas

3 — September 18, 2025— Data Visualization

Question: How can vizualizations help us understand our data

Topics: The grammar of graphics

Read: Imai (2022) Chapters 1 and 3

Wickham et al. (2023) Chapter 3 `https://r4ds.had.co.nz/data-visualisation.html`

Lab: Exploration of Covid-19 data

Wednesday, September 24, 2025 at 11:59 pm on Canvas

4 — September 25, 2025— Causation I

Question: How do we know if X causes Y?

Topics: Potential outcomes and counterfactuals; The fundamental problem of causal inference; A "statistical" solution to that problem; The role of randomization

Read: Imai (2022) Chapter 2

Broockman and Kalla (2016) skim for understanding of basic question, data and design

Lab: Exploration of Broockman and Kalla (2016)

Wednesday, October 1, 2025 at 11:59 pm on Canvas

ICYI: Rubin (1974) a classic, oft-cited study illustrating the value of the potential outcomes (PO) framework

Holland (1986) useful summary of PO framework at the time with some philosophical and historical context. Cite for mantra of "no causation without manipulation"

Holland (2003) interesting discussion of "immutable" characteristics and their role in causal inference.

Pearl (2009), an alternative view of causality using graphs and structural models.

5 — October 2, 2025— Causation II

Question: How do we know if X causes Y without randomly assigning X?

Topics: Tools for drawing causal inferences from observational data

Read: Imai (2022) Chapter 2

Angrist and Pischke (2014) Chapter 4

Grumbach and Hill (2022) skim for understanding of basic question, data and design

Lab: Exploration of Grumbach and Hill (2022)

Wednesday, October 8, 2025 at 11:59 pm on Canvas

Assignment 1: Drafting Research Questions

Due Thursday, October 9, 2025 at 3:59 pm on Canvas

ICYI: Rosenbaum (2002) The King James Bible of causal inference from observation data. A bit technical

Rosenbaum (2010) The New International Version of causal inference from observation data? Still pretty technical but maybe more approachable

Rosenbaum (2017) The Gideon's Bible of causal inference from observation data? New book, but I think designed to be more approachable

Imbens and Rubin (2015) Another great resource

6 — October 9, 2025— Prediction I

Question: How do we make predictions?

Topics: Simple linear regression

Read: Imai (2022) Chapter 4

Lab: Exploration of the phenomena of Red-Covid with simple linear regression

Wednesday, October 15, 2025 at 11:59 pm on Canvas

ICYI: James et al. (2013) a free textbook with accompanying course that provides a great introduction to topics of machine learning.

7 — October 16, 2025— Prediction II

Question: How do we make predictions adjusting for potentially confounding factors?

Topics: Multiple regression

Read: Imai (2022) Chapter 4

Angrist and Pischke (2014) Chapter 2

Lab: Exploration of the phenomena of Red-Covid with multiplw linear regression

Wednesday, October 22, 2025 at 11:59 pm on Canvas

Assignment 2: Developing your proposal and identifiying a replication project

Thursday, October 23, 2025 at 3:59 pm on Canvas

ICYI: Achen (2003) Toward a New Political Methodology: Microfoundations and ART or why someone will always ask you why there are more then three predictors in your model.

8 — October 23, 2025— Probability I

Question: What do we mean by probability and how do we use it?

Topics: Axioms of probability; Conditional probability; Bayes Rule; Discrete and continuous probability distributions; Expectations, variance, and moments

Read: Imai (2022) Chapter 6 Levendusky (2009) Skim Chapters 2-3

Lab: Exploration of Meierrieks and Auer (2024)

Wednesday, October 29, 2025 at 11:59 pm on Canvas

ICYI: There's probably no substitute for taking a course or two in probability and statistics at the intro graduate or advanced undergraduate level. That said, these can be useful references

Wasserman (2011)

Hogg et al. (2013)

Freedman (2009)

9 — October 30, 2025— Probability II

Question: What do we mean by probability and how do we use it?

Topics: Conditional probability; Bayes Rule; Likelihoods; The Law of Large Numbers; The Central Limit Theorem; Likelihoods

Read: Imai (2022) Chapter 6 Levendusky (2009) Skim Chapters 2-3

Lab: Exploration of Meierrieks and Auer (2024)

Wednesday, November 5, 2025 at 11:59 pm on Canvas

10 — November 6, 2025— Uncertainty I

Question: How do we quantify uncertainty?

Topics: Asymptotic and simulation based approaches to sampling distributions, standard errors, and confidence intervals.

Read: Imai (2022) Chapter 7

Lab: Exploration of National Election Studies data

Wednesday, November 12, 2025 at 11:59 pm on Canvas

Assignment 3: Initial analyses

Thursday, November 13, 2025 at 3:59 pm on Canvas

ICYI: Fisher and Others (1935) worth reading at some point in your careers

11 — November 13, 2025— Uncertainty II

Question: How do we quantify uncertainty?

Topics: Asymptotic and permutation based approaches to hypothesis test and p-values

Read: Imai (2022) Chapter 7

Lab: Further exploration of National Election Studies data

Wednesday, November 19, 2025 at 11:59 pm on Canvas

ICYI: Bowers and Panagopoulos (2011) A nice introduction to randomization-based inference

12 — November 20, 2025— Uncertainty III

Question: How do we evaluate model performance

Topics: Inference on linear models. Multiple hypothesis testing.

Read: Imai (2022) Chapter 7

Lab: Applications to your final project

ICYI: Bretz et al. (2011) A great resource on the issues of testing multiple comparisons with applications in R

Not Due, Focus on your final papers

Assignment 4: Final Paper Draft

Thursday, November 20, 2025at 11:59 pm on Canvas

13 — November 27, 2025— Thanksgiving Recess — NO CLASS

Question: What's your favorite pie?

Topics: Dry vs Wet Brining. Advanced napkin folding. Do we have enough wine?

Read: Nissan (2009)

14 — December 4, 2025— Final Paper Workshops

Question: How are your projects going?

Topics: We'll trouble shoot any technical issues and make sure you're ready to present at least some initial results the week after

Read:

Lab: **Not due. Work on your final papers and presentations**

Assignment 5: Slides and/or poster presentation of final paper.

Due Sunday, December 8, 2025 at 11:59 pm on Canvas

15 — December 11, 2025— Presentations

Question: Who's afraid of public speaking?

Topics: You and your great ideas!

Read: A peer's draft

Lab: Presentation of initial findings

ICYI: Jesse Shapiro "How to Give an Applied Micro Talk"

Sunday, December 15, 2025— Final Papers DUE at 11:59 pm on Canvas

References

Achen, C. H. (2003). TOWARD A NEW POLITICAL METHODOLOGY: Microfoundations and ART.

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering 'metrics: The path from cause to effect*. Princeton University Press, Princeton, NJ.

Bowers, J. and Panagopoulos, C. (2011). Fisher's randomization mode of statistical inference, then and now. *Unpublished manuscript*.

Bretz, F., Hothorn, T., and Westfall, P. (2011). *Multiple comparisons using R*. CRC Press, Boca Raton.

Broockman, D. and Kalla, J. (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, 352(6282):220–224.

Fisher, R. A. and Others (1935). The design of experiments. *The design of experiments.*

Freedman, D. A. (2009). *Statistical models: theory and practice*. cambridge university press.

Garfinkel, A. (1981). *Forms of explanation: Rethinking the questions in social theory*. Yale University Press New Haven.

Grumbach, J. M. and Hill, C. (2022). Rock the Registration: Same Day Registration Increases Turnout of Young Voters. *The journal of politics*, 84(1):405–417.

Hogg, R. V., McKean, J. W., and Craig, A. T. (2013). *Introduction to mathematical statistics*. Pearson Education.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

Holland, P. W. (2003). Causation and race. *ETS Research Report Series*, 2003(1):i—-21.

Imai, K. (2017). *Quantitative Social Science: An Introduction*. Princeton, NJ: Princeton University Press.

Imai, K. (2022). *Quantitative Social Science An Introduction in Tidyverse*. Princeton University Press.

Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*. Springer.

King, G., Keohane, R. O., and Verba, S. (1994). *Designing social inquiry: Scientific inference in qualitative research*. Princeton university press.

Levendusky, M. (2009). *The partisan sort: how liberals became Democrats and conservatives became Republicans*. University of Chicago Press.

Meierrieks, D. and Auer, D. (2024). Bribes and bombs: The effect of corruption on terrorism. *The American political science review*, pages 1–17.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys*, 3(0):96–146.

Rosenbaum, P. R. (2002). Observational studies. In *Observational Studies*, pages 1–17. Springer.

Rosenbaum, P. R. (2010). Design of observational studies. *Springer series in statistics*.

Rosenbaum, P. R. (2017). *Observation and experiment : an introduction to causal inference*. Harvard University Press, Cambridge, MA.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5):688.

Wasserman, L. (2011). *All of Statistics*. Springer New York.

Wickham, H., Çetinkaya Rundel, M., and Grolemund, G. (2023). *R for data science*. " O'Reilly Media, Inc.".