Political Science 2580

INTRODUCTION TO QUANTITATIVE RESEARCH METHODS

Paul Testa
Assistant Professor
paul_testa@brown.edu

First Draft: August 29, 2017
This Draft: September 3, 2019

General Information

Canvas https://canvas.brown.edu/courses/1079715

Where/When We meet Mondays, 6:30–9:00 pm, in 111 Thayer Room 140.

Office Hours **Paul's Office Hours** Friday, 1–3 pm in 111 Thayer St Room 339. If you know in advance that you want to come to officehours, please email me to reserve a 20 minute slot.

Sections: To be discussed in class.

Overview

This course provides an introduction into the basics of quantitative, empirical research in the social sciences. We will begin exploring 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, 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 at https://cran.r-project.org/. 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, as are additional supplemental materials, such as slides, notes, 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 their 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 improve our studies. 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 research.

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.

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 LaTeX, a free, open-source typesetting software available here: https://www.latex-project.org/get/. **All of your assignments and papers for this class will also be submitted as pdfs created using R Markdown and LaTeX** It's a short but steep learning curve, the benefits of which (pretty documents, nicely formatted tables and figures, easy integration with citation managers) far outweigh the costs (finicky syntax that can be insanely frustrating at first). We will download and learn LaTeX together in class and I will provide ample TeX support throughout the course.

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. Our errors are volitional and 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. Students in need of academic accommodations should contact me via email to set up a time to meet and determine reasonable

[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

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

accommodations.

Course Structure and Policies

Readings There are two required textbooks for the course (**Estimated cost: ∼$70.00**):

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

Angrist, J. D. and Pischke, J.-S. (2013). Mastering Metrics. *Journal of Chemical Information and Modeling*, 53(9):1689–1699

A supplemental text with slightly more rigorous treatment of some statistical concepts. **Approximate Cost: $22.00**

Additional readings will be listed below and available to download on Canvas. You may also consider purchasing Gerber and Green (2012). We'll read selections, but it is an excellent reference and well worth your money (**Approximate Cost: $42.00**)

Class Broadly the structure for each class is as follows. Before each class on Wednesday, 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 Saturday before class**. On Sunday, I'll post solutions to the prior assignment. During class on Monday day, we'll review the prior weeks' work, discuss the current week's topic and then get started together on that week's lab. Whatever you don't finish in class you will be expected to complete and submit to the Canvas by 11:590 pm on the Saturday after class. Course slides, assignments, comments, and supplemental material will be posted to Canvas.

Math Throughout the course, we will devote part of lecture to reviewing some of the underlying math behind each weeks' topics. Sections that week, will provide an opportunity to work through these concepts in more detail through optional problems sets. The problem sets will also provide an opportunity to practice your programming skills.

The topics covered are:

- Linear Algebra
- Functions and operations
- Limits
- Differential Calculus
- Integral Calculus

[4] Estimated from Amazon

You'll find notes for these topics and more at `http://paultesta.org/mathcamp` (Currently a work in progress)

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 Saturday after class.**[5]

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. All assignments are due the Monday after the class with which they are associated.

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

Week 4: Drafting Research Questions
Due Monday, October 7, 2019 at 6:29 pm on Canvas

Week 7: Developing your proposal
Due Monday, October 26, 2019 at 6:29 pm on Canvas

Week 10: Pre-analysis Plan
Due Monday, November 25, 2019 at 6:29 pm on Canvas

Week 13: Final Paper Draft
Due Saturday, December 7, 2019 at 6:29 pm on Canvas

Week 14: Slides and/or poster presentation of final paper
Due Thursday, December 12, 2019 at 11:59 pm on Canvas

Week 15: Final Papers DUE at 11:59 pm Wednesday, December 18, 2019 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:

- **10% Class involvement and participation**
- **35% Weekly labs**
- **5% Assignments not including draft**
- **5% Final Paper Draft**

[5] The lab for week before Thanksgiving will be due the Wednesday we come back from Fall Break.

- **5% Final Presentations**
- **40% Final Papers**

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 Tuesday. 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 Monday after the class in which they are assigned** and MUST be **submitted online to Canvas by 5:00 pm.** No exceptions.[6]

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 Monday after the class in which they are assigned** and MUST be **submitted online to Canvas by 9:00 pm.**

ICYI: "In case you're interested..."[7] A set of related readings you may find interesting and useful. Occasionally, I'll also post links to supplemental notes on course related topics.

[6] 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.

[7] A convention/tic borrowed from DFW (2010)[a][b]

[a] David Foster Wallace

[b] Who as this well-intentioned but probably ill-advised descent into self-referential footnotes might

Math: Links to relevant mathematical concepts for that week.

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 9, 2019— 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* and LaTeX

Read: Imai (2017) Chapter 1

King et al. (1994) Chapter 1

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

Lab: None

Assignment: Download templates from Canvas for Pre-Analysis Plan, Final Paper, and Presentations. Update and compile these documents. Upload updated documents to canvas. (We'll start these in class)

Due Saturday, September 14, 2019 at 5:00 pm on Canvas

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

2 — September 16, 2019— Data and Measurement

Question: How do we describe the world around us?

Topics: Levels of measurement; Organizing describing and visualizing data

Read: Imai (2017) Chapters 1 and 3

Wickham (2014) `http://r4ds.had.co.nz/tidy-data.html`

Lab: Exploration of Quality of Goverment Data QoG data

Due Saturday, September 21, 2019 at 11:59 pm on Canvas

---

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

Math: Linear Algebra I and Functions

3 — September 23, 2019— 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 (2017) Chapter 2

Angrist and Pischke (2013) Chapter 1

Findley et al. (2013) skim for understanding of basic question, data and design

Lab: Exploration of Findley et al. (2013)

Due Saturday, September 28, 2019 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.

4 — September 30, 2019— Causation II

We'll need to reschedule this class.

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

Topics: Tools for drawing causal inferences from observational data

Read: Imai (2017) Chapter 2

Angrist and Pischke (2013) Chapter 4

Ferwerda and Miller (2014) skim for understanding of basic question, data and design

Lab: Exploration of Ferwerda and Miller (2014)

Due Saturday, October 5, 2019 at 11:59 pm on Canvas

Assignment 1: Drafting Research Questions

Due Monday, October 7, 2019 at 6:29 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

5 — October 7, 2019— Prediction I

Question: How do we make predictions?

Topics: Simple linear regression

Read: Imai (2017) Chapter 4

Angrist and Pischke (2013) Chapter 2

Lab: Exploration of Taubman Poll Data

Due Saturday, October 12, 2019 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.

Math: Limits and Differential Calculus

6 — October 14, 2019— Indigenous Peoples' Day. No University exercises

7 — October 21, 2019— Prediction II

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

Topics: Multiple regression

Read: Imai (2017) Chapter 4

Angrist and Pischke (2013) Chapter 2

Lab: Exploration of Taubman Poll Data

Due Saturday, October 26, 2019 at 11:59 pm on Canvas

Assignment 2: Developing your proposal and identifiying datasets

Due Monday, October 28, 2019 at 6:29 pm on Canvas

ICYI: Achen (2002) 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.

Math: Linear Algebra II

8 — October 28, 2019— 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 (2017) Chapter 6 Levendusky (2009) Skim Chapters 2-3

Lab: Exploration of Levendusky (2009) Chapters 2-3

Due Saturday, November 2, 2019 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 and Craig (1995)

Freedman (2005)

Math: Integral Calculus

9 — November 4, 2019— 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 (2017) Chapter 6 Levendusky (2009) Skim Chapters 2-3

Lab: Exploration of Levendusky (2009)

Due Saturday, November 9, 2019 at 11:59 pm on Canvas

10 — November 11, 2019— Uncertainty I

Question: How do we quantify uncertainty?

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

Read: Imai (2017) Chapter 7

Lab: Exploration revisiting Findley et al. (2013)

Due Saturday, November 16, 2019 at 11:59 pm on Canvas

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

11 — November 18, 2019— Uncertainty II

Question: How do we quantify uncertainty?

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

Read: Imai (2017) Chapter 7

Lab: Exploration revisiting Ferwerda and Miller (2014)

Due Saturday, November 23, 2019 at 11:59 pm on Canvas

Assignment 3: Pre-Analysis Plan

Due Monday, November 25, 2019 at 6:29 pm on Canvas

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

12 — November 25, 2019— Uncertainty III

Question: How do we quantify uncertainty?

Topics: Inference on linear models; Testing multiple hypotheses

Read: Imai (2017) Chapter 7

Lab: Applications to your final project

Due Monday, December 2, 2019 at 6:30 pm on Canvas

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

13 — December 2, 2019— Explorations

Question: What do you want to know?

Topics: TBD. We'll use this and the following week to work on your final papers and reviewing core concepts. An optional lab will provide some tools and topics of interest to you based on your work.

Read: TBD

Lab: TBD. Possible topics include: generalized linear models, instrumental variables, regression discontinuities, matching, panel data, hierarchical models, time series analysis, text as data, spatial analysis.

Not Due, Focus on your final papers

Assignment 4: Final Paper Draft

Due Saturday, December 7, 2019 at 11:59 pm on Canvas

14 — December 9, 2019— Explorations

Question: What do you want to know?

Topics: TBD. An optional lab will provide some tools and topics of interest to you based on your work.

Read: TBD

Lab: TBD. See topics above

Not due. Work on your final papers and presentations

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

Due Thursday, December 12, 2019 at 11:59 pm on Canvas I need to get the posters to the printer for...

15 — December 16, 2019— 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: Spirling (2013) "Giving a (Job) Talk: Notes from the Field"

Jesse Shapiro "How to Give an Applied Micro Talk"

Wednesday, December 18, 2019— Final Papers DUE at 11:59 pm on Canvas

References

Achen, C. H. (2002). T OWARD A N EW P OLITICAL M ETHODOLOGY : Microfoundations and ART. *Annual Review of Political Science*, 5(1):423–450.

Angrist, J. D. and Pischke, J.-S. (2013). Mastering Metrics. *Journal of Chemical Information and Modeling*, 53(9):1689–1699.

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.

Ferwerda, J. and Miller, N. L. (2014). Political Devolution and Resistance to Foreign Rule: A Natural Experiment. *American Political Science Review*, 108(03):642–660.

Findley, M. G., Nielson, D. L., and Sharman, J. C. (2013). Using Field Experiments in International Relations: A Randomized Study of Anonymous Incorporation 1. *International Organization*, 67(04):657–693.

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

Freedman, D. (2005). Statistical models: Theory and practice. *books.google.com*.

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

Gerber, A. S. and Green, D. P. (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Hogg, R. V. and Craig, A. T. (1995). *Introduction to Mathematical Statistics*, volume 14.

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.

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.

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.

Spirling, A. (2013). Giving a (Job) Talk: Notes from the Field | The Political Methodologist.

Wallace, D. F. (2010). *Everything and more : a compact history of Infinity*. W.W. Norton.

Wasserman, L. (2011). *All of Statistics*, volume 102.

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10).