Gov 1000/2000/2000e/Stat E-190
Introduction to Quantitative Methods I

Xiang Zhou

Fall 2019

E-mail: xiang_zhou@fas.harvard.edu
Office Hours: Monday 3-5pm
Office: CGIS K309

Web: scholar.harvard.edu/xzhou
Class Hours: Tue 9:30am-11:45am
Class Room: CGIS S153

TF: Jacob Brown
Office hours: CGIS K108, Fri 3-5pm
Email: jrbrown@g.harvard.edu

TF: Margaret Schwenzfeier
Office hours: CGIS K108, Tue 4-6pm
email: schwenzfeier@g.harvard.edu

Course Description

How can we detect voting irregularities? What causes individuals to vote? In what sense (if any) does democracy (or trade) reduce the probability of war? Quantitative political scientists address these questions and many others by using statistical methods that are informed by theories in political science and the social sciences more generally. In this course, we provide an introduction to the tools used in basic quantitative social science research. The first part of the course cover introductory univariate statistics, while the remainder of the course focuses on linear regression models. Furthermore, the principles learned in this course provide a foundation for the future study of more advanced topics in quantitative political methodology. We will cover both the theoretical and computational aspects of statistics, proving important theorems and learning how to analyze real data. While the tools of statistical inference are worth studying in their own right, another goal of this course is to provide graduate students (and some undergraduates) with the necessary skills to critically read, interpret, and replicate the quantitative content of many political science articles. As such, the statistical methods covered in this course will be presented within the context of a number of articles. Throughout the term, we will reanalyze the data and revisit the conclusions from various prominent papers in the social sciences.

Who should take which course number?

We have designed the class with a great deal of flexibility in mind and have various course numbers that correspond to students with different backgrounds. Note that all sections of the course will use the R statistical computing environment.

Gov 2000 This is the default course number for all graduate students who will be doing any empirical research in political science or the social sciences more generally. Even if you think you're going to use only qualitative methods in your research, you should still take this course to give yourself a solid understanding of quantitative methods. This course will teach you to be flexible data analysts, capable of tailoring standard methods to the unique situation of each task. You will control the tools, not the other way around. You will learn to write and adjust

code to replicate and critique results from the literature. You will also learn how to work with basic statistical theory, working with proofs of canonical results and building the foundation for working at a higher methodological level. This level will require some calculus and matrix algebra knowledge at the level of the Harvard Gov Math Prefresher (see prerequisites section below).

Gov 2000e This course number is designed for those students who plan to do absolutely no empirical work in political science. Students in this section will focus on the analysis and critique of methods and empirical work. You will still do some data analysis, but the coding aspect of the class will be less emphasized. In its place, you will be expected to produce a higher and more competent level of analysis/criticism in all assignments (no free lunches).

Gov 1000 This is the default course number for undergraduate students and roughly covers the same material as Gov 2000 with special tailoring for Gov undergraduates. This means fewer technical computational or mathematical questions on assignments.

Stat E-190 This is the course number for Harvard Extension School students. Those taking the course for graduate-level credit will do work corresponding to the Gov 2000 level while those taking it for undergraduate-level credit will do work corresponding to the Gov 1000 level. Lectures and sections will be taped and made available to all students within 1 to 2 days.

Prerequisites

The most important prerequisite is a willingness to work hard on possibly unfamiliar material. Statistical methods is like a language and it will take time and dedication to master its vocabulary, its grammar, and its idioms. This presents a challenge for us as instructors to give you the best intuition and a challenge for you as a student to work hard to internalize that intuition.

Formally, the prerequisites vary for different types of students. For graduate students in the Government Department there are no course prerequisites except the completion of the Math Prefresher (or the equivalent). For other graduate students, undergraduate students, and Extension School students, the prerequisite for all course numbers is a basic course in statistics such as Gov 50, Gov E-1005, Stat E-100, or the equivalent. For graduate-level enrollment (either in Gov 2000 or Stat E-190), some previous experience with probability, calculus, and matrix algebra is strongly recommended. Working knowledge of basic algebra is assumed for all course numbers.

For any student who meets the prerequisites yet is concerned with his or her preparedness for the course, we strongly encourage the following in advance of the semester. First, we recommend reading and working through the exercises in David Freedman, Robert Pisani, and Roger Purves, Statistics, 2007 (any of the older editions should suffice as well). Next, we encourage familiarization with the R for the section of the course the student intends on taking. Moreover, if the student plans on typesetting problem set answers in LATEX, familiarity with the LATEX markup language would be helpful. Resources on R and LATEX are available under the "Resources" tab on the class website. Finally, it may be helpful to review the material from the Government Department Math Prefresher, available here: http://projects.iq.harvard.edu/prefresher/home.

Course Details

Pre-class Readings

In each week, there are readings for topics covered in the lecture. **All readings should be finished before attending the lecture**. Take notes, re-derive expressions, write down your impressions and confusions, and bring your questions to the classroom or to Piazza(especially for Extension school

students). All of your classes should be pushing your research forward and you will be more creative the more you actively read.

Lectures

Lectures will be held weekly and will cover the broad theoretical topics of the course. In addition, we will work through example problems, computation in R, and canonical or insightful proofs of key results. Lectures will be taped and made available to both the extension school students and those in the College and GSAS.

Homeworks

Methods are tools and it isn't very instructive to read a lot about hammers or watch someone else wield a hammer. You need to get your hands on a hammer or two. Thus, in this course, you will have homeworks on a weekly basis. They will be a mix of analytic problems, computer simulations, and data analysis. In general, the homework will be released on Tuesday (after the lecture) and due at 2pm on the next Wednesday. Solutions will be posted after section. These homeworks should be typed and well-formatted, with tables and figures incorporated into the text. We will grade on a $(+; \checkmark; -)$ basis (including half grades between these categories). No late homework will be accepted except in the case of a documented emergency.

Midterm

The midterm will be a checkout exam, that should only take a few hours to complete, and only involves short analytical problems. You will have five hours to complete exam, but it should take less time than this. This exam will be available for checkout one week after we finish the material on univariate statistics, and it is designed to ensure that all students understand the foundational material before we move to regression. Both FAS and Extension School students will upload the completed exam to the course Canvas site. The tentative schedule for the midterm is that it will be distributed on roughly October 8th and due back on October 11th. *This schedule is subject to change depending how fast our course is moving.*

Take-home Final

The take-home final exam will be handed out on Thursday, December 5. It will be due at 5:00pm on Thursday, December 12. The take-home final primarily involves data analysis and interpretation. Note that the format and goals for the take-home exam are very different from the format and goals for the midterm exam. Both FAS and Extension School students will upload the completed exam to the course Canvas site.

Collaboration Policy

We encourage students to work together on the homework assignments, but you must write your own solutions (this includes computer code), and you must write the names of your collaborators on your assignment. I also strongly suggest that you make a solo effort at all the problems before consulting others. The midterm and the final will be very difficult if you have no experience working on your own. **No** collaboration is allowed on either the midterm or the final exam.

Section

Section will be held every week from the second week (CGIS K050, Thu, 6-7pm and 7-8pm). Section will cover various topics, including review of class materials and help with computing issues.

Jacob and Connor will run section and can give more details. Attendance is *very strongly* encouraged.

Questions about Course Materials

In this course, we will use an online discussion board called Piazza. This is a question-and-answer platform that is easy to use and designed to get you answers to questions quickly. We encourage you to use the Piazza Q&A board when asking questions about lectures, problem sets, and other class materials outside of section and office hours. You can sign up to the Piazza course page either directly from the below address or the link posted on the Canvas course website (there are also free Piazza apps for Android and iOS devices):

https://piazza.com/harvard/fall2019/gov100020002000estate190

Using Piazza will allow you to see and learn from questions others have. Both TFs and I will regularly check the board and answer questions posted, although everyone else is also encouraged to contribute to the discussion. Your respectful and constructive participation on the forum will count toward your class participation grade. Do not email your questions directly to TFs or me (unless they are of a personal nature) — we will not answer them!

Office Hours and Availability

My office hours are 3-5pm Mondays. The office hours for the TFs are posted above and will be held in the CGIS Cafe, known as the Fisher Family Commons. If you have questions about the course material, computational issues, or other course-related issues please do not hesitate to set up an appointment with either any of us.

If you have a general question, you can also post it on Piazza. This is almost always the fastest way to get an answer. However, you can email me directly if you have a question of personal nature.

Participation

Ten percent of the grade will be awarded for class participation, quality of presentation on the homework, and reading comments. Posting questions on Piazza about the assigned reading or the lecture notes will count towards class participation. These comments and questions provide feedback for tailoring the lecture to the needs of the students in the course.

Grading

- weekly homework assignments (50% of final grade)
- a midterm exam (10% of final grade)
- cumulative take-home final exam (30% of final grade)
- participation (10% of final grade).

Required Books

The following textbooks are required for this course:

- Imai, Kosuke. 2018. Quantitative Social Science: An Introduction. Princeton University Press.

- Fox, John. 2016. Applied Regression Analysis and Generalized Linear Models, 3rd Edition. SAGE Publications, Inc.

We will also use the following materials (*available on Canvas*)

- Bertsekas & Tsitsiklis Lecture Notes, chapters 2-3, 7
- Gov Math Prefresher, chapter 6

Computing

We will use R in this class, which you can download for free at http://www.r-project.org. R is open source and available on all major platforms (including Solaris, so no excuses). You can find a virtually endless set of resources for R on the internet, including this Getting Started With R page. We recommend using RStudio, an editor and development environment for R. If you are completely new to R, we recommend that you walk through the programming part of the Gov Math Prefresher(booklet available on Canvas), which provides an excellent introduction to R.

The HES-specific policy on Academic Integrity

You are responsible for understanding Harvard Extension School policies on academic integrity (https://www.extension.harvard.edu/resources-policies/student-conduct/academic-integrity) and how to use sources responsibly. Not knowing the rules, misunderstanding the rules, running out of time, submitting the wrong draft, or being overwhelmed with multiple demands are not acceptable excuses. There are no excuses for failure to uphold academic integrity. To support your learning about academic citation rules, please visit the Harvard Extension School Tips to Avoid Plagiarism (https://www.extension.harvard.edu/resources-policies/resources/tips-avoid-plagiarism), where you'll find links to the Harvard Guide to Using Sources and two free online 15-minute tutorials to test your knowledge of academic citation policy. The tutorials are anonymous open-learning tools.

The HES-specific policy on Accessibility

The Extension School is committed to providing an accessible academic community. The Accessibility Office offers a variety of accommodations and services to students with documented disabilities. Please visit here for more information.

Preliminary Schedule

The following is an anticipated schedule of course topics. The plan is to cover one topic per week, but we will go as fast as needed to make sure that everyone is understanding the material. Check the Canvas site to know what we will be covering the in an upcoming lecture.

Week 01, 09/03: Overview

- Course details and requirements
- What are the goals of the course?
- Basic descriptive statistics
- Random variables

Week 02, 09/10: Probability Distributions

- Probability distributions
- Cumulative distribution functions
- Summarizing probability distributions
- Important distributions
- Simulating from random variables

Pre-class Reading

- Imai, Ch. 6.1-6.3
- Bertsekas & Tsitsiklis, Ch. 2.1-2.4 & 3.1-3.3

Week 03, 09/17: Multiple Random Variables

- Joint and conditional distributions
- Covariance, correlation, and independence
- Laws of total expectation

Pre-class Reading

- Bertsekas & Tsitsiklis, Ch. 2.5-2.7, 3.4-3.5

Week 04, 09/24: Conditional Variance & Sample Means

- Law of total variance
- Distribution of the sample mean
- Useful inequalities
- Law of Large Numbers
- Central Limit Theorem

Pre-class Reading

- Imai, Ch. 6.4
- Bertsekas & Tsitsiklis, Ch. 5.1-5.4

Week 05, 10/01: Estimation and Statistical Inference

- Populations, samples, statistical models
- Point estimation
- Properties of estimators
- Confidence intervals

Pre-class Reading

- Imai, Ch. 7.1

Week 06, 10/08: Midterm Review

Week 07, 10/15: Hypothesis Testing

- Hypothesis testing
- p values
- Power analysis

Pre-class Reading

- Imai, Ch. 7.2

Week 08, 10/22: What is Regression?

- Conditional expectation function
- Ordinary least squares
- Mechanics of least squares

Pre-class Reading

- Fox, Ch. 2, Ch. 5.1

Week 09, 10/29: Simple Linear Regression

- Assumptions of the linear model
- Properties of least squares
- Gauss-Markov Theorem
- Inference with regression

Pre-class Reading

- Fox, Ch. 6.1

Week 10, 11/05: Multiple Regression I: Motivation and Matrix Algebra

- Motivation
- Matrix algebra

Pre-class Reading

- Gov Math Prefresher, Ch. 6

Week 11, 11/12: Multiple Regression II: Statistical Inference

- Projection matrices
- Mechanics of multiple linear regression
- Inference in a multiple linear regression model
- Omitted Variable Bias

Pre-class Reading

- Fox, Ch. 5.2, Ch. 6.2-6.3

Week 12, 11/19: More on Multiple Regression

- Regression Anatomy Formula
- Inference on Expected Outcomes
- Dummy variables and F-test
- Interaction Effects

Pre-class Reading

- Fox, Ch. 7.1-7.4

Week 13, 11/26: Diagnosing and Fixing Problems

- Detecting Outliers
- Nonlinear CEF
- Heteroskedastic Errors

- Dependent Data

Pre-class Reading

- Fox, Ch. 11, Ch. 12.1-12.3

Week 14, 12/03: Final Review