Gov 50: Introduction to Political Science Research Methods

Matthew Blackwell

Office: CGIS K305
Office Hours: Wed., 2–4pm (or by appointment)
mblackwell@gov.harvard.edu
http://www.mattblackwell.org

Head TF: Michael Olson
email: michaelolson@g.harvard.edu

TF: Hanno Hilbig
email: hhilbig@g.harvard.edu

General Information

How can we measure racial discrimination in job hiring? What is the best way to predict election outcomes? What factors drive the onset of civil wars? Is it possible to determine what members of Congress are more or less liberal given their voting record? These are just a few of the numerous question that social scientists are tackling with quantitative data. Beyond academia, companies and non-profits have invested heavily in data science techniques to learn about their users, platforms, and programs. Data scientists at these institutions are essentially applied social scientists and employ many of the same techniques you will learn in this course.

What will you learn in this course? Our goal is to give you the ability to understand, explain, and perform social science research, with a special focus on data analysis and causal reasoning. You will be able to read and understand the methodology of most academic articles in the social sciences, but more importantly you will have a foot in the

door of the data science world. The ability to collect and analyze data in a sophisticated manner is becoming a crucial skill set for the modern job market across industries. Finally, you will obtain data literacy that will help you be a critical consumer of evidence for the rest of your life.

Who should take this course?

Gov 50 is designed primarily for Government concentrators and fulfills the Methods requirement for these students. Students from other departments who are interested in quantitative social science are welcome as well. You should be prepared to dedicate time outside of class to keep up with the material even beyond the time dedicated to assignments and exam review. Furthermore, you should feel comfortable with engaging in real-life data analysis using statistical software. We will guide you through that process, but it may be unfamiliar and therefore challenging. You should especially take this course if you plan to use quantitative data at all in a senior thesis. All of these statements applies equally to extension schools students.

Finally, note that Gov 50 is not the only way to satisfy the Methods requirement for the Government concentration. Statistics 100 and Statistics 104 both satisfy the requirement and different focuses in terms of material and statistical computing. You may consider shopping each of the courses to see which of them will be the best fit for you. I think we'll have the most fun in Gov 50, though.

Prerequisites

The most important prerequisite is a willingness to work hard on possibly unfamiliar material. Beyond that, we will assume a basic familiarity with high-school algebra and a working knowledge of computers. If you are unfamiliar with downloading and installing software programs on your Mac or PC, you may want to allocate additional time to make sure those aspects of the course go smoothly. Previous experience with statistical methods and computing (Stata, R, SAS, SPSS) is helpful, but not required. You can always get in touch with the teaching staff for additional help on these issues.

Course Details

Lectures

Lectures will be held twice weekly and will cover the broad theoretical topics of the course. In addition, we will work through example problems and show how to use R to analyze data. Lectures will be taped (for use in an Extension school section) and made available 24-48 hours after lecture ends.

Sections

Every student will be assigned to a section of course, the time and location of which are to be determined. There will be a taped section for extension school students. These sections will focus on reviewing material from class that is useful for the homeworks and exams. These section meetings are *crucial* for your understanding of the material in this course.

Grading

- Datacamp computer assignments (10% of final grade)

- four homework assignments (20% of final grade)

- two take-home midterm exams (40% of final grade)

- final group project (20% of final grade)

- participation (10% of final grade).

Datacamp

We will assign short weekly R-based computing exercises through https://www.datacamp.com. These will be interactive and guided lessons through computing topics covered that week, and they will be graded based on completion. These will generally be due Thursdays in weeks when no homework is due.

Reading

The reading for the course will be available online through the Perusall app. You will be able to rent or purchase the required Imai book on that platform and we will post any additional readings there as well. Perusall will allow you to annotate the reading when you are confused, have a question, or want to make a comment. You will also see the annotations of other students and can respond to their questions. All of this will allow us to better tailor lectures and sections to where students are having difficultly. Annotations will count toward the participation grade.

Homeworks

Only reading about data science is about as instructive as reading a lot about hammers or watching 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 usually be focused on data analysis in general and will often involve a real dataset. There will

be 4 throughout the semester. See the schedule below for dates when these homeworks will be distributed and due back.

Midterms

There will be two in-class midterm exams on **Tuesday, October 9th** and **Thursday, November 15th**. We will hold exam review sessions in place of regular lectures in the lecture before these exams.

Final Group Project

The final project for the course will be a data analysis project that students taking the course for undergraduate credit will complete in groups of 3-4. Students will find a dataset of interest, state an interesting research question about that data, and answering this question using that data. **The final group project report will be due Monday, December 10th at 11:59pm ET.** More details will be announced as the semester continues.

Participation

Ten percent of the grade will be awarded for class participation both in-class, on Canvas, and on Perusall. This includes asking substantive questions about the course material, answering other students' questions about the material, or providing additional resources about the topics covered in class from other sources (like YouTube). Full credit will be subject to review of the quality of the posts, but 5-10 substantive posts throughout the semester should be sufficient to earn full credit on this portion of the grade.

Regrading Policy

If you feel there has been an error in the grading of one your assignment, you may request (in writing) a regrade of the assignment. A member of the teaching staff will regrade the entire assignment, not just the part you are disputing. Therefore, your regrade might increase or decrease the overall grade on the assignment.

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 midterms will be very difficult

if you have no experience working on your own. **There is no collaboration allowed on the midterm exams.**

Course Canvas Site & Discussion Board

We will be using Canvas to host the course website this year. You can find the site at the following URL: https://canvas.harvard.edu/courses/43384. On Canvas you will find a Discussion Board for class-related discussion. The quicker you begin asking questions on Canvas, the quicker you'll benefit from the collective knowledge of your classmates and instructors. This is an ideal forum for posting questions regarding the course material and/or computing. I encourage students to reply to each other's questions, and a student's respectful and constructive participation on Canvas will count toward his/her class participation grade.

Office Hours and Availability

My office hours are 2-4pm Wednesdays. 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 Canvas. This is almost always the fastest way to get an answer. However, you can also email me directly at mblackwell@gov.harvard.edu. If the question is of general interest, I will forward the question and my answer to the class. Make sure to tell me explicitly in your email if you would like to stay anonymous.

Required Books

The following textbook is **required** for this course:

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

You should purchase this book through Perusall, which you can access through the Canvas site. There you have the option to purchase the e-book for $50 or rent the book for 180 days for $25.

Optional Books

- Diez, David M., Christopher D. Barr, and Mine Çetinkaya-Rundel. 2015. *Open-Intro Statistics.* 3rd edition. https://www.openintro.org/

· Freedman, David, Pisani, Robert, and Purves, Roger. 2007. *Statistics*. W.W. Norton & Company. 4th edition.

Computing

We'll use R in this class to conduct data analysis. R is free, open source, and available on all major platforms (including Solaris, so no excuses). RStudio (also free) is a graphical interface to R that is widely used to work with the R language. You can install R and RStudio on your own computer (and you should!), but we will also provide you with an account on a cloud-based version of RStudio that you can run through your web browser (Chrome, Safari, Internet Explorer, Firefox, etc). This will minimize some of the pain of setting up RStudio and allow us to easily provide you with script templates for homework assignments.

You can find a virtually endless set of resources for R on the internet. For beginners, there are several web-based tutorials including one from DataCamp. In these, you will be able to learn the basic syntax of R. We'll post more R resources on Canvas.

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.

We will also have a Google Calendar with up-to-date information on HWs, office hours, lectures, sections, and more here: https://goo.gl/5JxdJc

Class Number	Lecture Date	Topic	Reading	Notes	DataCamp
1	9/4/2018	Introduction			
2	9/6/2018	Intro to R & RStudio	QSS 1.1–1.4		DC1
3	9/11/2018	Causality: Introduction	QSS 2.1–2.3		
4	9/13/2018	Causality: Randomized Studies	QSS 2.4	HW 1 out	DC2
5	9/18/2018	Causality: Observational Studies	QSS 2.5		
6	9/20/2018	Measurement: Descriptive Statistics	QSS 2.6, 3.1–3.2	HW 1 due	
7	9/25/2018	Measurement: Visualization	QSS 3.3		
8	9/27/2018	Measurement: Survey Sampling	QSS 3.4	HW 2 out	DC3
9	10/2/2018	Measurement: Bivariate Statistics	QSS 3.5–3.6		
10	10/4/2018	**Review Session**		HW 2 due	
11	10/9/2018	**No lecture (exam)**			
12	10/11/2018	Prediction: Election Prediction	QSS 4.1		
13	10/16/2018	Prediction: Regression I	QSS 4.2.1–4.2.3		
14	10/18/2018	Prediction: Regression II	QSS 4.2.4–4.2.6	HW 3 out	DC4
15	10/23/2018	Prediction: Regression & Causality I	QSS 4.3.1–4.3.2		
16	10/25/2018	Prediction: Regression & Causality II	QSS 4.3.3–4.3.4	HW 3 due	
17	10/30/2018	Probability: Introduction	QSS 6.1–6.2		
18	11/1/2018	Probability: Random Variables	QSS 6.3	HW 4 out	DC5
19	11/6/2018	Inference: Large Samples	QSS 6.4		
20	11/8/2018	Inference: Survey Estimation	QSS 7.1	HW 4 due	
21	11/13/2018	**Review Session**			
22	11/15/2018	**No lecture (exam)**			
23	11/20/2018	Inference: Experimental Estimation	QSS 7.1		
24	11/27/2018	Inference: Hypothesis Testing I	QSS 7.2		
25	11/29/2018	Inference: Hypothesis Testing II	QSS 7.2		DC 6
26	12/4/2018	Inference: Linear Regression	QSS 7.3		