Note: all course communications will take place over Slack. **Click here to join the course Slack.**

Teaching Team

Instructor: Adeline Lo

Office: CGIS Knafel 407

Contact: Slack/adelinelo@fas.harvard.edu

OH: 9-11AM Tu

TF: Shusei Eshima

Contact: Slack

Sections: 4:30-5:30PM, WTh (K450)

OH: 5:30-6:30PM, WTh (K450)

TF: Laura Morris

Contact: Slack

Sections: 2-3PM M (South S354); 12-1PM F (K109)

OH: 1-2 PM M (S354); 11AM-12PM F (K109) [*OH on **4/1 and 4/4 only** by appointment, reserved for project check-ins]*

CA: Noah Dasanaike

Contact: Slack

Sections: 9-10AM, M (K031); 10-11AM, F (K108)

OH: 10-11AM, M (K031); 11AM-12PM, F (K108)

CA: Rucha Joshi

Contact: Slack

Sections: 3-4PM, M (K450); 4-5PM, Th (K401)

OH: 4-5PM, M (K450); 5-6PM, Th (K401)

CA: Angie Shin

Contact: Slack

Sections: 5-6PM, W (K031)

OH: 6-7PM, W (K031)

Study Hall: 5-7PM, F

CA: Evelyn Cai

Contact: Slack

Sections: 4:30-5:30PM, Tue (K354)

OH: 5:30-6:30PM, Tue (K354)

Study Hall: 2-4PM, Th (K105)

CA: Yao Yu

Contact: Slack

Sections: 7-8PM, M (Harvard Hall 201)

OH: 8-9PM, M (Harvard Hall 201)

Study Hall: 9-11PM, M (Harvard Hall 201)

CA: Tahmid Ahmed

Contact: Slack

Sections: 4:30-5:30PM, W (S001)

OH: 5:30-6:30PM, W (S001)

Study Hall: 5-7PM, W (S001)

CA: Miroslav Bergam

Contact: Slack

Sections: 12-1PM, F (K050)

OH: 1-2PM, F (K050)

Study Hall: 2-4PM, F (K050)

* K: CGIS Knafel, S: CGIS South

Overview

What accounts for who votes and their choice of political candidates? Do ethnic groups vote in support of one another even in multiethnic societies? Do arrest patterns in US cities show evidence of racial profiling? This course will teach students how to address these and other social science questions by analyzing quantitative data. The course introduces basic principles of statistical inference and programming skills necessary for data analysis.
The goal is to provide students with the foundation necessary to analyze data in their own research as well as become critical consumers of statistical claims made in media, policy and academic research. We further motivate the course with the message of The New York Times article titled, â€œFor Todayâ€™s Graduate, Just One Word: Statistics.â€

Class goals

This course is primarily designed for undergraduate students in the social sciences. You will learn the statistical and computational principles necessary to conduct basic statistical tests and regression analyses of empirical data. You will be able to do so using the R programming language and present graphs and plots of your results and findings. It will require a lot of hard work for all of us to achieve that; however, the class is structured to provide you with the framework to achieve that in combination with hard work and reaching out when appropriate for extra support.

By the end of the semester, you will be able to:

- Critically read and interpret quantitative content of many articles in the quantitative social sciences
- Conduct, interpret, and communicate results from analysis using statistical tests and regression
- Explain the limitations of observational data for making causal claims, and begin to use existing strategies for attempting to make causal claims from observational data
- Write basic clean, reusable, and reliable R code
- Feel empowered working with data.

More specifically, the course covers basic probability, univariate inference, linear regression and its

applications in causal inference strategies. We will also provide an introduction to statistical programming in R.

Class and Section

Instruction for this course is conducted via two avenues: class and section/lab. Class lectures are twice a week and will typically focus on statistical material. Section meets once a week and will typically focus on practical problem solving and/or computational skills. Both are essential to the learning process.

Prerequisites

Formally, this course does not have prerequisites. **We recommend taking Tyler Simko's pre-semester workshop to familiarize yourself with outfitting your computer with R/RStudio and familiarizing yourself with packages, RMarkdown, Slack and Swirl, all of which feature extensively throughout the course.**

Materials

Computational tools

The best way, and often the only way, to learn about data analysis and new statistical procedures is by doing. We will therefore make extensive use of a flexible (open-source and free) statistical software program called R, RStudio, and a number of companion packages. Problem sets and section exercises will be completed in R Markdown. You will learn how to program in this class, if you do not know already. **Again, consider taking Tyler Simko's pre-semester workshop designed to help you hit the ground running for this course specifically.**

Readings

This course will use the following textbook:

- **Imai, Kosuke. 2018. Quantitative Social Science: An Introduction. https://tinyurl.com/W22-GOVT-50-1**

Suggested Readings for R

The following is available for free online or through the library and are excellent introductions to R in increasing order of difficulty.

- Matloff, Norman. 2011. The Art of R Programming: A Tour of Statistical Programming. No Starch Press.
- Wickham, Hadley. Advanced R (available free online!)

Swirl

You can and should practice R regularly as you interact with course materials; one way to inject extra practice that's already paired with our course content is through the usage of Swirl exercises (mini guided R practice!).

- Directions to download swirl exercises are also available at **https://github.com/kosukeimai/qss-swirl**

- A new interactive tutorial â€œ**qsslearnr**â€ that improves upon the qss-swirl has been made available by Matt Blackwell at **https://github.com/mattblackwell/qsslearnr**

Assignments

There are four types of assignments in this course:

1. **Preparing for class and participation**: Often for your classes there will be some reading you must do before class. We expect you to come to both class and sections 100% prepared; we do not assign much reading, but we assume you have read it. Your participation in class, section and on discussions on Slack as both a learner and a support to your colleagues is a part of your assessment for the

semester and good practice towards creating a positive learning environment!
2. **Section and labs:** Section is mandatory, and will almost always include a lab; please arrive prepared and ready to participate. Labs are due the following day from section and are graded on a complete/incomplete basis. They are a key way for you to experience guided practice for the learned materials an extra time before the point-graded weekly problem sets, so you have more comfort with the material.
3. **Weekly problem sets**: learning statistics and programming takes consistent practice. The problem sets are described below.
4. **Final project:** you will work in small groups to complete a final project.

Preparing for Class and Section/Participation

There are readings for each of the modules and topics within them through the semester; some of it will feature mathematical notation. We recognize that becoming familiar and comfortable with this type of language can be challenging the first time you do it --- and so it will be tempting to skip math that seems particularly daunting---don't do this! The math is often the meat of the statistical work, and part of your learning goals for this course is to become more accustomed and comfortable with reading and understanding mathematical notation. Read carefully and go line by line to make sure you understand. Read the required readings and any others that might evoke your interests as we progress through the semester. Engage with the readings: take notes, write down your questions, impressions and confusions, talk with your classmates and TFs, and post questions/answers on Slack. Actively assisting your classmates in class, section or Slack will constitute 5% of your final grade. Please bring your laptops to section, and feel free to also bring them to lectures.

Problem Sets

Learning by watching and not doing is hard when learning how to use any new tool, so in order to do more and practice with regularity you will have homework on a weekly basis. The assignments will be a mix of analytic problems, computer simulations and data analysis.

Assignments must be completed in R Markdown, which allows you to show both your answers and the code you used to arrive at them. You will need to submit to Gradescope. Once you upload the PDF file, you will see a list of the questions in the assignment and thumbnails of your file. For each assigned question, click the PDF page(s) that contains your answer. Don't worry if you don't know R Markdown, we will show you how it works in section with more detailed instructions before the first assignment is due.

Each week's homework will be made available on Canvas starting Tuesday immediately after class and is due Tuesday the following week (7 days later) immediately **before** class. Solutions will be made available through Canvas. Working through the problem sets includes looking at the solutions key so please remember to do this portion!

Problem sets are graded out of 0-10 points. We reserve the right to add bonus points for aesthetics including presentable graphs, clear code, nice formatting and well-written answers.

There will be 10 problem sets in total, constituting 50% of your grade. Your problem set with the lowest grade will be dropped when calculating your final grade. Late problem sets drop 3 points (out of the total 10) each late day, with a maximum of 2 late days, after which we will not accept the problem set anymore. We do not want to hold up the class and will not wait for everyone to submit their problem sets in order to post the solutions key. If you are turning in your problem set late, you are on your honor to *not look at the solutions key* before submitting your work.

Collaboration Policy

Unless otherwise stated, we encourage students to work together on the assignments, but you should write your own solutions (this includes code). That is, no copy-and-paste from other people's code. You would not copy-and-paste from someone's paper, and you should treat code the same way. However, we strongly suggest that you make a solo effort at all the problems before consulting others.

Final project

You will work in groups to ask a social science question of a dataset that will culminate in a final project due at the end of the semester. We will provide more details on the project later in the semester.

Grading

- Participation: 5%

- Problem sets: 10 total (lowest dropped), 50%
- Labs: 11 total, 10%
- Final project: 35%

The course will follow the standard grading scheme: A (100-94), A- (93-90), B+ = (89-87), B (86-83), B- (82-80), C+ (79-77), C (76-73), and C- (72-70).

How to learn in this course

If you find this course challenging, you are not alone. Statistics can be challenging and we cover a lot of ground. I have confidence in your abilities as smart and engaged researchers who can handle it. Below are some details on forms of support that we offer in this class.

Your primarily responsibilities in this class are to *work hard* and *communicate* with us about what you need. You cannot learn if you aren't putting in the time. We also can't help if we don't know there's a problem.

Resources for getting help

Below are a few main sources of support for this class.

1. Class and Section
    - We encourage you to be an active participant in class and section. Ask questions if you don't understand something that is happening.
2. Readings and Slides
    - If you are studying alone and hit something you don't understand, you should turn to the readings and slides. There will be a fair amount of material in the slides and they are intended to be used and reviewed multiple times, not just seen once during lecture.
3. Slack
    - We will be using a Slack workspace for communication in this class, but also as a source to post and answer questions about the material. You will not be required to post, but the system is designed to get you help quickly and efficiently from classmates, the TA and the professor. ***Unless the question is of a personal nature or completely specific to you, you should not email teaching staff***; **instead, you should post your questions on Slack**. We will be monitoring the discussion board, but we encourage you to help your classmates as well. Likely a significant amount of overlap will exist for both things people want to know more about and things people have just figured out.
4. TF office hours
    - Each of your TFs has office hours each week. TF office hours are often useful for getting help with new tricky material and problem sets.
5. Instructor office hours
    - My office hours are in my office listed at the top of the syllabus.
6. Problem set and lab keys
    - Soon after the problem sets are due, we will post the key. It may be tempting to immediately turn focus towards the next problem set, but if you were uncertain about anything in the problem set, I recommend you check the key to lock down core concepts. Some of the material builds directly on previous concepts! The same holds for lab keys.

Topics

Module I: Introduction

Week 1 Tuesday 1/25, Thursday 1/27
1. Introduction to the course
2. Introduction to causal inference, programming platforms
3. Lab 1
4. Problem Set 1 assigned (due following Tuesday).

Read before class Tuesday QSS Chapter 1 [available online] and try exercises in the chapter by end of week.

Swirl: INTRO1 INTRO 2 by end of week.

Module II: Causality

Week 2 Tuesday 2/1, Thursday 2/3

1. Randomized experiments
2. Lab 2
3. Problem Set 2 assigned.

Read before class Thursday 2/3: QSS Chapter 2.1-2.3 [available online].

Swirl: CAUSALITY1, CAUSALITY2

Week 3 Tuesday 2/8, Thursday 2/10
1. Randomized experiments
2. Observational studies
3. Lab 3
4. Problem Set 3 assigned.

Read before class Tuesday 2/8: QSS Chapter 2.4-2.5 [available online].

Read before class Thursday 2/10: QSS Chapter 2.6 [available online].

Swirl: MEASUREMENT1

Module III: Measurement

Week 4 Tuesday 2/15, Thursday 2/17
1. Survey sampling
2. Visualizing distributions
3. Lab 4
4. Problem Set 4 assigned.

Read before class Tuesday 2/15: QSS Chapter 3.1-3.4.

Read before class Thursday 2/17: QSS Chapter 3.5-3.6.

Swirl: MEASUREMENT2

Week 5 Tuesday 2/22, Thursday 2/24
1. Measuring ideology
2. Clustering
3. Lab 5
4. Problem Set 5 assigned.

Read before class Tuesday 2/22: QSS Chapter 3.7.

Module IV: Prediction

Week 6 Tuesday 3/1, Thursday 3/3
1. Election polls and predictions
2. Regression
3. Lab 6
4. Problem Set 6 assigned.

Read before class Tuesday 3/1: QSS Chapter 4.1.

Read before class Thursday 3/3: QSS Chapter 4.2.

Swirl: PREDICTION1, PREDICTION 2

Week 7 Tuesday 3/8, Thursday 3/10
1. Regression
2. Causal regressions
3. Lab 7
4. Problem Set 7 assigned.

Read before class Tuesday 3/8: QSS Chapter 4.3 (start).

Read before class Thursday 3/10: QSS Chapter 4.3 (finish).

Swirl: PREDICTION3

Module V: Probability

Week 8 Tuesday 3/22, Thursday 3/24
1. Probability and conditional probability
2. Bayes
3. Lab 8
4. Problem Set 8 assigned.

Read before class Tuesday 3/22: QSS Chapter 6.1-6.2.

Read before class Thursday 3/24: QSS Chapter 6.3.

Swirl: PROBABILITY1, PROBABILITY2

Week 9 Tuesday 3/29, Thursday 3/31
1. Random variables
2. Probability distributions
3. Large sample theorems
4. Lab 9
5. Problem Set 9 assigned.

Read before class Tuesday 3/29: QSS Chapter 6.4.

Read before class Thursday 3/31: QSS Chapter 6.4-6.5.

Module VI: Uncertainty

Week 10 Tuesday 4/5, Thursday 4/7
1. Estimation
2. Lab 10
5. Problem Set 10 assigned.

Read before class Tuesday 4/5: QSS Chapter 7.1.

Swirl: UNCERTAINTY1, UNCERTAINTY2

Week 11 Tuesday 4/12, Thursday 4/14
1. Hypothesis testing
2. Lab 11

Read before class Tuesday 4/12: QSS Chapter 7.2.

Swirl: UNCERTAINTY3

Week 12 Tuesday 4/19
1. Regression with uncertainty

Read before class Tuesday 4/19: QSS Chapter 7.3.