Spring 2020 / Monday & Wednesday, 1:30-2:20 pm, Zoom (updated 4/7/2020)

POL346: Applied Quantative Analysis

Prof Wasow
Office: 201 Fisher Hall
owasow@princeton.edu

> "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write." — H.G. Wells
>
> "The sexy job in the next ten years will be statisticians...The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill." — Hal Varian, Chief Economist, Google, *McKinsey Insights*, 1/09
>
> "For Today's Graduate, Just One Word: Statistics" — *NYT*, 6/8/09

Course Description

In a world awash in data, how can we distinguish signals from noise? This course focuses on developing an intuition for statistics and applying it through data analysis, regression models and a final project. We will wrestle with what makes a good research question, play with data to see how statistical methods can help us make sense of real world concerns, and work at communicating quantitative findings clearly to broad audiences. Particular attention will be paid to applying these techniques in Junior Papers and Senior Theses. Coursework involves using the R statistical platform.

Overview

Broadly, the goal of this course is to develop your statistical literacy so that you can generate, interpret, convey and critique statistical findings. In particular, we will focus on developing an intuition for statistics, learning to apply statistical tools and practicing how to communicate meaningful statistical insights. Learning statistics is a lot like learning a language in that it requires lots of practice. Toward that end, the coursework emphasizes working with real data and developing skills that can be applied to your own research and future careers.

Instructors

- Prof Wasow
  owasow@princeton.edu
  201 Fisher Hall
  http://bit.ly/wasow_cal
- Beatriz H. Barros
  bbarros@princeton.edu
  Fisher Hall B01
  Office Hours: 3-5 on Monday, Fisher B16
- David Ribar
  dribar@princeton.edu
  Fisher Hall B22
  Office Hours: Wednesdays 10-12
- Federico H. Tiberti
  ftiberti@princeton.edu
  Fisher Hall B01
  Office Hours: Mondays 10-12.

Prerequisites

This course is a continuation of POL345. If you have not taken POL345, but have taken an equivalent in Econ, COS or ORFE, please see me on the first day. This course will require you to use algebra but not calculus or matrix algebra.

Textbook

- For the first two-thirds of the semester, we will primarily use *The Statistical Sleuth: A Course in Methods of Data Analysis*, by Fred Ramsey and Daniel Schafer. You may use either the second or third edition. Used editions are fine (and often much cheaper).

Additional References

- Our intro to the `tidyverse` is *R for Data Science*, by Garrett Grolemund and Hadley Wickham, at http://r4ds.had.co.nz.
- Our intro to the `infer` package is *Modern Dive*, by Chester Ismay and Albert Y. Kim, at https://moderndive.com.
- For discussing data visualization, we recommend these two books:
  - *Data Visualization: A practical introduction* by Kieran Healy https://socviz.co
  - *Fundamentals of Data Visualization* by Claus O. Wilke https://serialmentor.com/dataviz/
- For the last third of the semester we will reference several other more advanced texts including:
  - Marginal effects plots: "Making the Most of Statistical Analyses: Improving Interpretation and Presentation," by Gary King, Michael Tomz, and Jason Wittenberg, at https://web.stanford.edu/~tomz/pubs/ajps00.pdf
  - Missing data: "Missing-data imputation," by Andrew Gelman and Jennifer Hill, Chap 25 in *Data Analysis Using Regression and Multilevel/Hierarchical Models* at http://www.stat.columbia.edu/~gelman/arm/missing.pdf.
  - Matching: "Matching methods for causal inference: A review and a look forward," by Elizabeth A. Stuart, at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2943670/.
  - Natural experiments: *Natural Experiments in the Social Sciences: A Design-Based Approach*, by Thad Dunning.
  - Text analysis: *Text Mining with R*, by Julia Silge and David Robinson, at https://www.tidytextmining.com.

Other resources

- Students are encouraged to ask and answer questions on `Piazza`. Participation on `Piazza` counts towards the class participation grade.
- Students have six months of access to http://www.DataCamp.com

Grading

- Class precept and Piazza participation - 4%.
- Problem sets (complete eight of nine) - 24%
  There will be nine problem sets through the semester of which you should complete eight. The questions will primarily come from *The Statistical Sleuth*. Problem sets should be submitted online with a compiled pdf that shows R code (we will discuss this further in class).
- Reports (complete four) - 36%.
  As part of preparing you for JPs, Senior Theses and the class final, there will be four report assignments. These will be assigned approximately three weeks in advance and will require an original analysis of data. Generally, each report should include a hypothesis, a statistical test of the hypothesis and an intuitive data visualization.
- Take home final - 36%
  There will be an open-book, take-home final exam. Students should not discuss its contents until after the submission deadline.

Grading Policies

- Deadlines: Unless otherwise stated, assignments should be submitted online by 11:59pm on Friday (Problem Sets) or Monday (Reports) via assignment folders at the POL346 Blackboard.
- Grading: All assignments will be graded on a 100 point scale.
- Lateness: Assignments will face a full grade-level penalty (e.g., A to B) for each day late without a prior extension.
- Accommodation: If you cannot make a deadline due to extracurricular activities, email me at least one week in advance to make alternative arrangements. If you have a personal problem that precludes you from completing coursework on time, please send me an email immediately. A doctor's note, or note from a Dean, may be requested.
- Format: For problem sets, submit your write-up, code and any plots or tables in a single document with all relevant code visible. For reports, submit separate pdf and Rmd files. Unless otherwise specified, reports should avoid showing code.

Class Schedule (will be revised)

Week	Date	Day	Title	Chapter
1	Feb 3	Mon	Introduction and Overview	1
1	Feb 5	Wed	Drawing Statistical Conclusions	1
2	Feb 10	Mon	Inference Using t-Distributions	2
2	Feb 12	Wed	Inference Using t-Distributions	2
3	Feb 17	Mon	Inference Using t-Distributions	2
3	Feb 19	Wed	A Closer Look at Assumptions	3
4	Feb 24	Mon	A Closer Look at Assumptions	3
4	Feb 26	Wed	Alternatives to the t-Tools	4
5	Mar 2	Mon	Comparison Among Several Samples	5
5	Mar 4	Wed	Comparison Among Several Samples	5
6	Mar 9	Mon	Simple Linear Regression	7
6	Mar 11	Wed	Simple Linear Regression	7
	Mar 16	Mon	Spring Recess	
	Mar 18	Wed	Spring Recess	
7	Mar 23	Mon	Simple Linear Regression	8
7	Mar 25	Wed	Null hypothesis, R-squared	9
8	Mar 30	Mon	Multiple regression	9
8	Apr 1	Wed	Interaction terms	10
9	Apr 6	Mon	Logistic regression	20
9	Apr 8	Wed	Binary and categorical models	20
10	Apr 13	Mon	Missing data	Handout
10	Apr 15	Wed	Causal inference: Matching	Handout
11	Apr 20	Mon	Causal inference: Matching	Handout
11	Apr 22	Wed	Causal inference: Panel data	Dunning
12	Apr 27	Mon	Causal inference: Natural Experiments	
12	Apr 29	Wed	Review	

Precepts

Section	Day	Time	Room	Preceptor
P01	Wed	2:30PM - 3:20PM	Peyton Hall 140	Beatriz H. Barros
P01A	Wed	2:30PM - 3:20PM	Friend Center 009	Federico Tiberti
P02	Wed	3:30PM - 4:20PM	McCosh Hall 26	Prof Wasow
P03	Wed	7:30PM - 8:20PM	Robertson Hall 008	Federico Tiberti
P04	Thurs	10:00AM - 10:50AM	Robertson Hall 008	David Ribar
P05	Thurs	11:00AM - 11:50AM	Robertson Hall 008	David Ribar

Problem Set, Report & Final Schedule

Week	Date	Day	Assignment	Percent
2	Feb 10	Mon	PS1	3
3	Feb 17	Mon	PS2	3
4	Feb 24	Mon	PS3	3
4	Feb 28	Fri	Report1	6
5	Mar 2	Mon	PS4	3
6	Mar 9	Mon	PS5	3
	Mar 16	Mon	Spring Break	
7	Mar 30	Mon	Report2	8
8	Apr 1	Wed	PS6	3
9	Apr 8	Wed	PS7	3
10	Apr 15	Wed	Report3	10
11	Apr 22	Wed	PS8	3
12	Apr 29	Wed	PS9	3
13	May 4	Mon	Report4	12
	May 18	Mon	Final	36

Collaboration

Problem sets for this course present opportunities for students to discuss questions and collaborate to find a solution together. At the same time, as with any class that includes analytical exercises and computer programming, there is a clear distinction between permissible collaboration and unacceptable copying or plagiarism. This course will follow a modified version of the guidelines used for computer science classes here at Princeton. Please take this guideline seri-

ously. In the past, plagiarism cases typically resulted in one-year suspension from Princeton.

Programming necessitates that you reach your own understanding of the problem and discover a path to its solution. During this time, discussions with other people (whether via the Internet or in person) are permitted and encouraged. However, when the time comes to write code that solves the problem, such discussions (except with course staff members) are no longer appropriate: the code must be your own work. For each assignment, please list the names of any individuals with whom you collaborated.

Do not, under any circumstances, copy another person's code. Incorporating someone else's code into your program in any form is a violation of academic regulations. Abetting plagiarism or unauthorized collaboration by sharing your code is also prohibited. Sharing code in digital form is an especially egregious violation: do not e-mail your code to anyone.

Novices often have the misconception that copying and mechanically transforming a program (by rearranging independent code, renaming variables, or similar operations) makes it something different. Actually, identifying plagiarized source code is easier than you might think. For example, computer software exists to detect plagiarism.

This policy supplements the University's academic regulations, making explicit what constitutes a violation for this course. Princeton Rights, Rules, Responsibilities handbook asserts:

> The only adequate defense for a student accused of an academic violation is that the work in question does not, in fact, constitute a violation. Neither the defense that the student was ignorant of the regulations concerning academic violations nor the defense that the student was under pressure at the time the violation was committed is considered an adequate defense.

If you have any questions about these matters, please consult me.