Spring 2021 / Monday & Wednesday, 2:30-3:20pm, (updated 1/28/2021)
Zoom: https://princeton.zoom.us/j/96410924045

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
  Office hours: http://bit.ly/wasow_cal
  Zoom: https://princeton.zoom.us/j/5503454968

- · Maria Aranzazu Rodriguez
  Uribe
  mauribe@princeton.edu
  Office Hours: TBD

- · Claire H. Willeck
  cwilleck@princeton.edu
  Office Hours: TBD

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 due dates via assignment folders at the POL346 Canvas site.

· 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 1	Mon	Introduction and Overview	1
1	Feb 3	Wed	Drawing Statistical Conclusions	1
2	Feb 8	Mon	Inference Using t-Distributions	2
2	Feb 10	Wed	Inference Using t-Distributions	2
3	Feb 15	Mon	Inference Using t-Distributions	2
3	Feb 17	Wed	A Closer Look at Assumptions	3
4	Feb 22	Mon	A Closer Look at Assumptions	3
4	Feb 24	Wed	Alternatives to the t-Tools	4
5	Mar 1	Mon	Comparison Among Several Samples	5
5	Mar 3	Wed	Comparison Among Several Samples	5
6	Mar 8	Mon	Simple Linear Regression	7
6	Mar 10	Wed	Simple Linear Regression	7
	Mar 15	Mon	Spring Recess	
7	Mar 17	Wed	Regression by Calculation	8
8	Mar 22	Mon	Null hypothesis, R-squared	8
8	Mar 24	Wed	Multiple regression	9
9	Mar 29	Mon	Interaction terms	9
9	Mar 31	Wed	Logistic regression	10
10	Apr 5	Mon	Binary and categorical models	20
10	Apr 7	Wed	Missing data	20
11	Apr 12	Mon	Causal inference: Matching	Handout
11	Apr 14	Wed	Causal inference: Matching	Handout
12	Apr 19	Mon	Causal inference: Panel data	Handout
12	Apr 21	Wed	Causal inference: Natural Experiments	Dunning
13	Apr 26	Mon	Review	
13	Apr 28	Wed	No class	

Precepts

Section	Preceptor	Day	Time	Room
P01	Willeck	Th	10:00 am - 10:50 am	TBD
P02	Willeck	Th	11:00 am - 11:50 am	TBD
P03	Willeck	Th	01:30 pm - 02:20 pm	TBD
P04	Uribe	F	11:00 am - 11:50 am	TBD
P05	Uribe	F	12:30 pm - 01:20 pm	TBD
P06	Uribe	F	01:30 pm - 02:20 pm	TBD
P07	Wasow	W	03:30 pm - 04:20 pm	TBD

Problem Set, Report & Final Schedule

Week	Date	Day	Assignment	Percent
1	Feb 5	Fri	PS1	3
2	Feb 12	Fri	PS2	3
3	Feb 19	Fri	PS3	3
4	Feb 23	Fri	Report1	6
5	Feb 26	Fri	PS4	3
6	Mar 5	Fri	PS5	3
7	Mar 12	Fri	Report2	6
	Mar 19	Fri	Spring Break	
8	Mar 26	Fri	PS6	3
9	Apr 2	Fri	PS7	3
10	Apr 9	Fri	Report3	10
11	Apr 16	Fri	PS8	3
12	Apr 23	Fri	PS9	3
13	Apr 30	Fri	Report4	10
	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 seriously. 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.