Spring 2022 / Monday & Wednesday, 1:15-2:30pm, (updated 1/22/2022)
Carnegie Building 109 and Zoom: https://pomonacollege.zoom.us/j/84648108533

POL90: Statistics

Prof Wasow
Office: 201 Fisher Hall
omar.wasow@pomona.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
>
> "We want facts. Factor, factor, factor is the motto which ought to stand at the head of all statistical work." — Florence Nightingale
>
> "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.

Office Hours and Mentor Sessions

- Prof Wasow
  omar.wasow@pomona.edu
  Office hours: Thursday 3-5
  Calendly: https://calendly.com/owasow/office-hours

Mentor	Day	Time	Link
Amber Lee	Wednesday	7-9pm	https://pomonacollege.zoom.us/my/mentor.amba
Arturo	Thursday	8-10pm	https://pomonacollege.zoom.us/j/6689769205
Mohsin Butt	Friday	6-8pm	https://pomonacollege.zoom.us/j/84925266132

Prerequisites

This is an introductory course with no prerequisites. 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` (or whatever Q&A site Pomona recommends). Answering other student's questions is the primary component of the class participation grade.
- Students have six months of access to http://www.DataCamp.com

Grading

- Class and Piazza participation - 4%.
- Problem sets (best eight of ten) - 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 three) - 24%.
  As part of preparing you for original research, theses, data analysis jobs and graduate school, you will complete three Reports. 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, an intuitive data visualization and a clear explanation of the analysis and results in prose.

- Quizzes (best two of three) - 12%.
  There will be three short in-class quizzes primarily covering material from the textbook. The dates will be provided in advance.

- 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 POL90 Gradescope 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	Jan 17	Mon	Introduction and Overview	-
1	Jan 19	Wed	Introduction	-
2	Jan 24	Mon	Drawing Statistical Conclusions	1
2	Jan 26	Wed	Drawing Statistical Conclusions	1
3	Jan 31	Mon	Inference Using t-Distributions	2
3	Feb 2	Wed	Inference Using t-Distributions	2
4	Feb 7	Mon	Inference Using t-Distributions	2
4	Feb 9	Wed	Confidence Intervals	2
5	Feb 14	Mon	A Closer Look at Assumptions	3
5	Feb 16	Wed	A Closer Look at Assumptions	3
6	Feb 21	Mon	Alternatives to the t-Tools	4
6	Feb 23	Wed	Alternatives to the t-Tools	4
7	Feb 28	Mon	Comparison Among Several Samples	5
7	Mar 2	Wed	Comparison Among Several Samples	5
8	Mar 7	Mon	Simple Linear Regression	7
8	Mar 9	Wed	Simple Linear Regression	7
9	Mar 14	Mon	Spring Recess	-
9	Mar 16	Wed	Spring Recess	-
10	Mar 21	Mon	Regression by Calculation	8
10	Mar 23	Wed	Null hypothesis, R-squared	8
11	Mar 28	Mon	Multiple regression	9
11	Mar 30	Wed	Interaction terms	9
12	Apr 4	Mon	Logistic regression	20
12	Apr 6	Wed	Logistic regression	20
13	Apr 11	Mon	Missing data	Handout
13	Apr 13	Wed	Missing data	Handout
14	Apr 18	Mon	Matching	Handout
14	Apr 20	Wed	Matching	Handout
15	Apr 25	Mon	Causal inference: Panel data	Handout
15	Apr 27	Wed	Causal inference: Natural Experiments	Dunning
16	May 2	Mon	Review	
16	May 4	Wed	No class	

Problem Set, Report, Quiz & Final Schedule

Week	Date	Day	Assignment	Percent
1	Jan 21	Fri	-	
2	Jan 28	Fri	PS01	3
3	Feb 4	Fri	PS02	3
4	Feb 11	Fri	PS03	3
5	Feb 18	Fri	PS04	3
6	Feb 25	Fri	PS05	3
7	Mar 4	Fri	Report1	6
8	Mar 11	Fri	PS06	3
9	Mar 18	Fri	Spring break	
10	Mar 25	Fri	PS07	3
11	Apr 1	Fri	PS08	3
12	Apr 8	Fri	Report2	8
13	Apr 15	Fri	PS09	3
14	Apr 22	Fri	PS10	3
15	Apr 29	Fri	Report3	10
16	May 6	Fri		
	May 13	Fri	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 many computer science classes. Please take this guideline seriously. In the past, plagiarism cases can result in serious consequences.

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 either your own work or that of your team. 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 College's academic regulations, making explicit what constitutes a violation for this course.

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