

PLSC 500: STATISTICS

Fall 2016

Course Personnel:

- Instructor: Alex Coppock alex.coppock@yale.edu 87 Trumbull Street (ISPS), Office D233. Office Hours: Tuesdays 9am - 12pm. Please email to schedule appointments, otherwise you might have to wait.
- Teaching Assistant: Jonathon Baron jonathon.baron@yale.edu Office Hours: Wednesdays, 3pm-5pm, RKZ 104.
- Teaching Assistant: Stephen Feagin stephen.feagin@yale.edu Office Hours: Mondays, 5:30pm-7:30pm, RKZ 104.

Course Meeting Times:

- Lecture: Tuesdays and Thursdays 1:30pm to 2:45pm in ISPS Room A001.
- Section: Fridays 10:30am to 11:20am in RKZ Room 102
- All course meetings are like a Liz Lemon party – **mandatory**.

Objectives: PSLC 500 is the first course in the graduate-level statistical methods sequence for political science students. It is nominally an introduction to statistics and linear regression with special emphasis on the nonparametric analysis of real-world data. We also have loftier goals. We hope to inspire:

1. An intuition for what data can and can't tell us about the world.
2. A love of code.
3. A habit of creating beautiful, reproducible documents.

Prerequisites: While mathematics training at the level of the math camp is preferred, all students are encouraged to attend. If you are uncertain about whether or not you are prepared for the course, please feel free to contact the instructor.

Textbooks

1. Aronow, Peter. M. and Miller, Benjamin T. (2016) *The Foundations of Agnostic Statistics*. (This book is in progress. A PDF (subject to change) will be placed on canvas.yale.edu.)
2. Goldberger, A. S. (1991). *A Course in Econometrics*. Cambridge, MA: Harvard University Press.
3. Additional weekly readings will be placed on canvas.yale.edu.

Software: We will be using the open-source statistical software R. While other statistical software packages such as SPSS, Stata, or even Excel can be used for many statistical operations R has many advantages. First, (with apologies to Python) it is the programming language of choice of many (most?) data scientists and statisticians. Second, it makes writing loops and functions very easy, tasks that are nearly impossible in Excel. Third, there is a large community of developers who have contributed a huge number of add-ons for R that you will find invaluable. Finally, it's free, and always will be, which is not true of other software. In addition to R, please also download and install RStudio, the top-of-the-line script editor.

- Download R here: www.r-project.org
- Download RStudio here: www.rstudio.org

Problem Sets: This course will involve a relatively heavy workload, and students should be aware that this class will require sustained, serious effort all throughout the term. The weekly problem sets will be a mix of mathematical demonstrations, simulations, and data analysis. The documents that you turn in must satisfy the following criteria:

- It must be in .pdf format.
- Mathematical expressions must be nicely formatted.
- Tables must be nicely formatted.
- Figures must be well-labeled and placed in-text.
- The code you use and what it produces must be in the same place.
- It must have your name on it [important]
- All students must write up their problem sets individually. However, you may work in groups of up to three, though you are not required to work in groups at all. Please indicate the names of the other students you worked with that week. Do not “share” members across groups. Do not copy and paste the answers across group members.

This list of requirements is restrictive. They basically mean that you can't use Microsoft Word to do your homework. While this policy seems obnoxious now, I *promise* that you will be happier later in your graduate career when you've already practiced making professional documents that will signal your high type. Practically speaking, you will use `rmarkdown`. `rmarkdown` is a way of weaving together text, code, and output in a single document. These documents are reproducible and transparent.¹ Your TAs are both good at `rmarkdown` and stand ready to help you. Please do not resist learning R and `rmarkdown`.

Grading Policy: Problem Sets (30%), Midterm Exam (30%), Final Exam (30%), Class Participation (10%).

Academic Honesty:

To ensure that you do not accidentally violate Yale's academic honesty policies, please review these sites:

- Academic Honesty: <http://bit.ly/2a6uTC5>
- Understanding and Avoiding Plagiarism: <http://bit.ly/29VnoN1>

I would like to emphasize that it is a violation of the honesty policy to:

- Copy another student's problem set, just changing a few words here and there. Collaboration is encouraged, but at some point relying too much on your partner becomes a violation of academic integrity. Most cases are clear-cut; for cases that are ambiguous, ask.
- Copy and paste whole blocks of code from your partner that you didn't have a hand in writing.

¹If you want to use \LaTeX + `knitr`, that's OK too.

- Copy whole sentences from the internet.

It is *not* a violation of the honesty policy to:

- Copy code from websites like stackoverflow or other online forums. This is not cheating, it's learning. Part of what makes it learning is that understanding code off the internet well enough to use it usually means that you at least sort of understand it. If you do copy such code, please include a link to the forum or site where you obtained the code in the comments. This is good practice anyway, as you will often forget where code came from!
- Discuss the problem sets with your partners and compare answers.

Acknowledgments:

This syllabus draws heavily from Peter Aronow's syllabus for this course.