

PSC 508
Basic Statistics
645-8445

Jim Battista
jbattist@buffalo.edu
Office hours: F 11–1; otherwise email me

Introduction

Methods classes are generally about last on people’s list of Big Fun Things, somewhere after root canal or watching *Battlefield Earth* with no beer. They involve math, and most of you probably picked poli-sci in part because it doesn’t require differential equations and such, and they’re otherwise often difficult or at least tricky and fussy. So it’s worth looking at why we’re requiring you to take this, and then more, and why people will pay attention to your methodological training when it’s off to the job hunt.

You’re not learning stats because it’s a rite of passage that we went through and are making you jump through too. It’s not about memorizing page after page of picky little formulas. Memorizing formulas, instead of learning concepts, is actually counterproductive.

You’re not learning stats just so that you can read journals with statistical tables in them and make up your own obfuscatory tables to baffle innocent readers. It’s not (usually) about constructing a private language that others can’t read.

You’re learning stats as part of a broader program of making inferences about the real world, and, importantly, getting them right. We can’t do experiments, at least not very often, so we have to rely on the experiments the world presents us with, and analyzing them means (sometimes) doing statistical inference. The difference between political science and political journalism is that we’re supposedly doing science; that we have an answer to the question ”How would you know if you were wrong?” Statistical inference about the real world is part of knowing whether you’re right or wrong about whatever it is you’re arguing.

Statistics are tricky and fussy because doing inference, because getting it right, is inherently a tricky, fussy, and careful process.

The good news is that statistics are not as difficult as you probably think. The mathematical notation can be inscrutable until you get used to it, but then so is something written in Spanish. The concepts themselves are usually not that difficult to hold in your head and work with, and you’ll rarely be required to do more than messy algebra.

The purpose of (this version of) this class is explicitly to prepare you for further work in Intermediate Statistics and beyond in a series of courses emphasizing inference. In the real world, you’ll be doing most of your statistical work on computer on the basis of articles, workshops, and books you’ve read. To that end, this course will concentrate on understanding what’s going on, on what the different things that you’re asking the computer to do for you really are, and on being able to make enough sense of what an article about using meta-tectonic rhombized regression to be able to set up a run and interpret it. That doesn’t mean we won’t do math problems; we will. Or at least you will. It means that the math problems will focus on helping you understand what’s going on, on what variances and hypothesis tests and confidence intervals are (and aren’t), and how they work. Likewise, it means getting down into the nitty-gritty of the definitions and equations enough that you start being (relatively) comfortable manipulating the math under them.

Requirements and grading

- Problem sets every week
 - Work on them alone; talk about them together after they're turned in
 - Bring –TWO– copies of your problem set. One to turn in to me, another to take notes on as we go over it.
 - SHOW YOUR WORK. The problem set will usually just tell you the final answer, so I expect just about everyone to get the vast majority of the questions right. I want to see if you know *why* you got that answer. If you give me answers that are just “3.85,” expect me to respond with chimpanzee-level angry shrieking.
 - Be neat. I may be generally informal, but your output will be formal, especially when we get to turning in results from statistical software. Then, I will expect neatly formatted tables, with the proper information in them, suitable for journal publication. Ya gotta start learning sometime. If you just cut and paste the output from your statistical software into your word processor, you will know what it is to be roasted in the depths of the Slor that day, I can tell you.
- A takehome midterm.
- A takehome final exam.
- Problems sets and each test are each roughly 1/3 of the course grade, subject to change at my whim.

Readings

- Wonnacott and Wonnacott, *Introductory Statistics*
- Hamilton, *Statistics with Stata*. Not strictly required, but you'll want it unless you're either a Stata guru already, or deeply invested in some other package.
- Occasional journal articles or working papers.

Schedule (subject to change at my whim)

31 August	Introduction, descriptive statistics W&W, Chs. 1, 2
7 September	NO CLASS
14 September	Probability, Bayes' Rule W&W, Ch. 3
21 September	Probability distributions W&W, Ch. 4
28 September	NO CLASS
5 October	Sampling W&W, Chs. 6, 7
12 October	Confidence intervals W&W, Ch. 8
19 October	Hypothesis testing W&W, Ch. 9
26 October	More hypothesis testing W&W, Chs. 9, 17 TAKEHOME MIDTERM THIS WEEK covering topics through 26 October
2 November	Two random variables, covariance, correlation W&W, Ch. 5
9 November	Regression W&W, Chs. 11, 12
16 November	Multiple regression W&W, Ch. 13; Gary King, "How Not to Lie With Statistics"
23 November	Dummies W&W, Ch. 14, maybe more
30 November	Interactions W&W, Ch. 14, maybe more
7 December	Diagnosing simple problems W&W, Chs. 14, 15

As well, there is the possibility of having an optional EXTRA BONUS!!! day during finals week.