RPAD 688: Statistical Program Workshop (1 credit course)

Fall 2012

Instructor: Stephen Weinberg

Wednesdays, 5:45 to 7:00

Note Well: you are expected to have access to a computer loaded with STATA, at least version 11.

The unofficial title of this course is “stuff Dr. Weinberg had to learn the hard way that he wished someone had taught him.” I’ve been working with STATA for 16 years now, and I’m still learning better ways to take advantage of the vast power this program offers.

Why learn how to program? Why not simply do everything using the various pull-down menus and buttons that STATA (or, God help you, Excel) offers?

If you want to do empirical work, you need to learn how to program, for at least five key reasons:

1) Some data sources expect you to be able to do some rudimentary programming in order to unpack the data;
2) You need to be able to manipulate your data and perform analyses that are not part of the standard pre-programmed cookbook;
3) Programming is a LOT faster than doing things interactively, once you know what you’re doing;
4) You need to be able to document exactly what you did, from the first moment you got your data until you wrote your paper;
5) You need to be able to modify what you did, so that if you change your mind later or get new data, you can simply modify your code and re-run it instead of having to start all over again.

This is a 1-credit course, which means that we meet for only 1 hour a week (well, one hour and fifteen minutes), and there’s a lot less homework than I’d assign in a 4-credit course. Nonetheless, doing some homework is essential if you are to gain proficiency.

The workload is heavily front-loaded. The most important material, and most of the homeworks, comes before the midterm. In the second half of the class, there will be much less homework (and no final exam). There is no homework in the last couple of weeks, when I know that you’ll be inundated with final projects and papers and exams in your other courses.
Grading

Participation: 8%
Homeworks: 60%
Midterm: 32%

The class is pass/fail. Each element will be graded on a scale of check-plus (20 points), check (17 points), check-minus (14 points), half (10 points), and zero (0 points). To pass, you need an average score of 14.

Participation: Showing up is worth a check-minus. To do better, you’ll need to, you know, participate.

Homework: There will be six short homework assignments, worth 10% of your grade each. For example, I may take some code from one of my projects, create a bunch of errors in it, and have you debug it. Or I may take some of my code, remove the comments, and ask you to explain some of the lines.

Midterm (Oct 24): I am a firm believer in the value of exams as devices for forcing you to figure out if you actually know something or not, and as a way to force you to pull the material together in your mind. I know that most of your courses will have a great deal of work at the end of the term, and the core material for the course comes in the first half, so I’ve decided to use a midterm instead of a final.

Collaboration

Working in groups is encouraged, provided that

1) Everyone takes a first pass on the assignment on their own;
2) Everyone prepares their own assignment to hand in, from scratch; and
3) You acknowledge who you worked with at the top of your assignment.

August 29: STATA syntax

Sept 5: Data structures

Sept 12: Do Files
   a. Homework 1 due

Sept 19: Analysis and Matrices
   a. Homework 2 due

Sept 26: NO CLASS

Oct 3: Loops
   a. Homework 3 due
Oct 10: Parallel Lists; String Variables
   a. Homework 4 due

Oct 17: Categorical Variables

Oct 24: Midterm

Oct 31: Generating Predicted Values

Nov 7: Graphing
   a. Homework 5 due

Nov 14: Survey Data
   a. Homework 6 due

Nov 21: NO CLASS

Nov 28: Reshaping Data

Dec 5: Useful Tricks