RPAD 688: Statistical Program Workshop (1 credit course)

Fall 2011

Instructor: Stephen Weinberg

Tuesdays, 5:45 to 6:45 [NOTE THE CHANGE!!!!]

Draper 313a

Note Well: you are expected to have access to a computer loaded with STATA, preferably version 11, but version 10 or 12 is acceptable.

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, and there’s a lot less homework than in a 4-credit course. Nonetheless, doing some homework is essential if you are to gain proficiency.

Grading

Participation: 8%

Homeworks: 64%

Midterm: 28%
Participation: showing up is worth a B. I expect you to ask—and respond to—questions.

There will be eight short homework assignments, worth 8% 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 25): 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 30: STATA syntax
   STATA overview
   Elements of the command line
   Some basic commands
   reading help files

Sept 6: Data structures
   How data are organized
   Different types of variables
   Commands for describing and manipulating data
   Merging datasets

Sept 13: Analysis
   Commands for doing analyses
   Commands for running tests
   Return and Ereturn lists
   Homework 1 due

Sept 20: Do files and Macros
   Basic mechanics of do files
   Local macros
   Passing arguments
   Quietly, Set More Off, and Capture Drop
   Homework 2 due
Sept 27: Looping and Matrices
   Looping and programmer’s if
   Matrices
   Homework 3 due

Oct 4: Strings
   String functions
   Homework 4 due

Oct 11: NO CLASS (there’s a talk about accountable care organizations that I want to go to)
   To Be Scheduled: an optional hour to go over some more examples from my own research

Oct 18: Graphics
   Dialog boxes
   Twoway graphs
   Graphics options
   Homework 5 due

Oct 25: MIDTERM

Nov 1: Useful Tricks
   Temporary variables and files
   Explicit subscripting
   Version 11’s new factor analysis syntax
   Version 11’s new margins command
   Homework 6 due

Nov 8: Survey Data I: theory of complex survey design
   Weighting
   Clustering
   Stratification
   Homework 7 due

Nov 15: Survey Data II: analyzing survey data in STATA
   Svyset

Nov 22: NO CLASS

Nov 29: Organizing for the Long Haul
   Homework 8 due

Dec 6: EndNote tips