RPAD 688: STATA Programming Workshop (1 credit course)

Fall 2013

Thursday, 2-3:15

Catalog #: 7335

Instructor: Stephen Weinberg

Draper 023

Note Well: you are expected to have access to a computer loaded with STATA, at least version 12. The University classrooms and libraries have STATA 13 installed.

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 18 years now, and I’m still learning better ways to take advantage of the vast power this program offers. (I’m also still learning ways to make dumb mistakes.)

This course is not about statistics per se, but about how to write programs to help you do statistics in STATA.

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. As a concession to the fact that people have seemed to find the workload rather heavy in the past, I have dropped one of the homeworks. I have also slowed the syllabus down a bit to spend more time on loop control macros, which has historically
been the hardest topic. (As a result, we will not get to cover weighting data or using STATA’s survey data features.)

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 few weeks, when I know that you’ll be inundated with final projects and papers and exams in your other courses.

Note Well: this course does not have a final project to cement the programming concepts we cover. The course was designed with doctoral students in mind, on the presumption that the “final project” would be the dissertation. The class has recently attracted the interest of MPA students, particularly given a demand for STATA in the job market. These students may find it beneficial to do an independent study with me in the spring consisting of a final programming project.

This term, I am experimenting with posting up to thirty minutes of additional lecture each week as online videos.

Prerequisites
None. If you have computer programming experience, you will find substantial parts of this course boring, but will at least learn useful tricks and warnings about STATA.

Where to Go Next?
RPAD 705 (pre-requisites: doctoral student or permission of Prof Deloughery) is a doctoral statistics class that uses basic STATA methods

RPAD 725 (pre-requisites: RPAD 705, 688) is a doctoral statistics class that applies these tools and teaches advanced STATA methods (such as the margins command).

Texts
Recommended: Long, 2009, The Workflow of Data Analysis Using STATA, STATA Press. This text is a useful supplement to the course but is not required.

Note: Mitchell, A Visual Guide to STATA Graphics, Third Edition, is highly recommended for when you start using this stuff in practice, but will not be useful for this class. We’ll learn some basics of STATA graphing; Mitchell walks through a very wide range of graphing options, with pictures.

Grading

Participation: 8%

Homeworks: 60%

Midterm: 32%
Participation: showing up is worth a check minus. I expect you to ask—and respond to—questions.

There will be five homework assignments, worth 12% 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. I like to include a section asking you to interpret STATA help files. The homeworks are heavily front-loaded to the beginning of the semester.

Midterm: 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.

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.

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.

Schedule

Note: Readings and homeworks are due on the day listed. The Long readings are optional, in that all material needed for homeworks and the midterm will be given in class or in video supplements. But the readings are highly recommended.

August 28: STATA syntax

Sept 4: Data
   a. Recommended: Long, Preface, the page titled “A word about fonts, files, commands, and examples,” chapter 1, chapter 2, chapter 3.1, chapter 6.3, Appendix A.1
   b. Handout on precision

Sept 11: Do Files
   a. Homework 1 due
   b. Recommended: Long, chapter 3, chapter 4.1, chapter 4.4
Sept 18: Retrieving Stored Results
   a. Recommended: Long chapter 4.2
   b. Homework 2 due

Sept 25: no class

Oct 2: Matrices
   a. Homework 3 due

Oct 9: Loops
   a. Recommended: Long chapter 4.3

Oct 16: Midterm

Oct 23: Nested Loops
   a. Recommended: Long 7.4

Oct 30: Graphing
   a. Homework 4 due

Nov 6: no class (watch videos on looping with graphs)

Nov 13: Merging datasets
   a. Recommended: Long 6.6
   b. Homework 5 due

Nov 20: the Reshape command

Nov 27: No Class

Dec 4: the Collapse command