Note Well: you are expected to have access to a computer loaded with STATA, at least version 10. Versions 11 and 12 are major upgrades, but the University computer systems will have version 10 for the immediate future. I will not require anything that cannot be done in STATA 10, but will describe some of the capabilities added in later versions.

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 17 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.)

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
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.

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 take the new RPAD XXX version of this course, which meets jointly with RPAD 688, has all the same homework, but also has a substantial final project, is letter graded, and counts for two credits.

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, on order from Mary Jane Books. 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 B. I expect you to ask—and respond to—questions.

There will be six 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. The homeworks are heavily front-loaded to the beginning of the semester.

Midterm (either Oct 10): 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 29: STATA syntax

Sept 5: No Class

Sept 12: Data
   a. Homework 1 due
   b. 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
   c. Handout on precision

Sept 19: Do Files
   a. Recommended: Long, chapter 3, chapter 4.1, chapter 4.4
   b. Homework 2 due

Sept 26: Retrieving and Storing Values
   a. Recommended: Long chapter 4.2
   b. Homework 3 due
Oct 3: Loops
   a. Recommended: Long chapter 4.3

Oct 10: Midterm

Oct 17: Matrices
   a. Homework 4 due
   b. Recommended: Long chapter 7.4

Oct 24: Storing and Displaying Results

Oct 31: Graphing
   a. Homework 5 due

Nov 7: Reshaping Data

Nov 14: Example from current research project
   a. Homework 6 due

Nov 21: Survey Data

Nov 28: No Class

Dec 5: Useful Tricks