

## **SOCY7709: QUANTITATIVE DATA MANAGEMENT**

Instructor: Natasha Sarkisian  
Email: [natasha@sarkisian.net](mailto:natasha@sarkisian.net)  
Phone: (617) 755-3178  
Office: McGuinn 417

Class time: Wednesdays 2-4:20 PM  
Class location: O'Neill 245  
Office hours: By appointment  
Webpage: <http://www.sarkisian.net/socy7709/>

### **COURSE DESCRIPTION**

The goals of this course are to develop the skills necessary to prepare complex datasets for analysis using an efficient, straightforward, error-free, and well documented data management process. The major topics of the course will include reading datasets into Stata, combining and reshaping datasets, saving and storing them, cleaning, recoding, and labeling the data, dealing with skip patterns and missing data, organizing the workflow, and basic programming in Stata.

This course is designed for graduate students with a prior background in statistics at the level of SOCY7703: Multivariate Statistics (or its equivalent). This means that students should have considerable experience with at least one statistical software package (e.g., SPSS, SAS, Stata, etc.). We will be using Stata for all the analyses throughout the course. No previous Stata experience is necessary: I will provide an introduction to Stata in the beginning of the course and guide you throughout the course. For your assignments, you can use Stata on Citrix: see <http://apps.bc.edu>. Some additional helpful resources for learning Stata include <http://www.ats.ucla.edu/stat/stata/> as well as Stata forum, Statalist: <http://www.statalist.org/>.

### **COURSE POLICIES**

**In-Class Activities:** For each data management topic covered in the course, I will give a lecture involving a demonstration in Stata, and you will get a chance to follow along and do all the necessary steps in Stata on your laptop. After that, you will work on applying these data management steps to a different dataset or do exercises, either individually or in groups.

**Communication:** The course is based on an interactive relationship between the instructor and students, as well as on collaboration among the students. You are strongly encouraged to ask questions and discuss the material in class. I also would like to stress that you are always welcome to come and see me with any additional questions. Email is the best way to get in touch with me to get a quick question answered or to set up an appointment to discuss something at length. Also, please check your email regularly: I will let you know by email when course notes are posted on the website and send other announcements from time to time.

**Feedback:** I would like to know how I could make this course experience as useful and interesting as possible. Therefore, every class in the end of class I will ask you to send me an email with at least one sentence of reaction to that class meeting, indicating what you learned, or something you liked or did not like, found clear or too simplistic, or found confusing and in need of further (or better) explanation. You may also submit comments on the course in general. Please be honest in your comments – if something is unclear or doesn't work for you, I really do want to know about that and will not penalize you in any way!

Coursework: Throughout the course, you are expected to do all your coursework on time. Ordinarily, no late assignments will be accepted, unless you have a valid (and documented) excuse.

Academic Integrity: It is your obligation to be fully aware of the Boston College policies on academic honesty. ANY violation may subject the offender to severe penalty, including course failure. If you are not familiar with the Boston College policy on academic honesty, see: <http://www.bc.edu/offices/stserv/academic/integrity.html>

Disability Accommodation: If you are a student with a documented disability seeking reasonable accommodations in this course, please contact Kathy Duggan, (617) 552-8093, [dugganka@bc.edu](mailto:dugganka@bc.edu), at the Connors Family Learning Center regarding learning disabilities and ADHD, or Paulette Durrett, (617) 552-3470, [paulette.durrett@bc.edu](mailto:paulette.durrett@bc.edu), in the Disability Services Office regarding all other types of disabilities, including temporary disabilities. Advance notice and appropriate documentation are required for accommodations.

### **COURSE REQUIREMENTS AND GRADING**

Two required books will be available at the Boston College bookstore:

1. Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.
2. Baum, Christopher F. 2016. *An Introduction to Stata Programming*, 2<sup>nd</sup> edition. Stata Press.

Other required readings will be available on electronic reserve in the library:

<http://www.bc.edu/reserves>

There will be six assignments in this course; each will require that you conduct specific data management steps using a large, complex dataset (e.g., read a dataset into Stata; recode and label all the variables needed for a specific analysis; merge and reshape the data; conduct multiple imputation to deal with missing data, etc.). I will provide the datasets for all the assignments.

If, however, you have a research project that involves a large, complex dataset, you can obtain permission from me to use that dataset for your assignments instead. Please contact me in the beginning of the course if you intend to use a different dataset and send me the information about the dataset and your project.

All assignments will be submitted electronically. Small files can be sent by email; any large files should be submitted using Google Drive, Dropbox or another file sharing website. Each assignment will be 15% of your grade; in-class participation will be another 10% of the grade.

| <b>Assignment</b>                               | <b>Due date</b> |
|---|-----------------|
| Assignment 1: Reading and Writing Datasets      | September 21    |
| Assignment 2: Recoding and Labeling             | October 12      |
| Assignment 3: Automating Your Work              | October 26      |
| Assignment 4: Merging, Reshaping, and Subgroups | November 9      |
| Assignment 5: Missing Data                      | November 30     |
| Assignment 6: Programming                       | December 14     |

## COURSE OUTLINE

### **August 31. Introduction to Data Management using Stata.**

### **September 7. Reading and Writing Datasets.**

Ch. 1, 2 from Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.

### **September 14. Data Cleaning.**

Ch. 3 from Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.

### **September 21. Creating and Changing Variables.**

Ch. 5 from Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.

**\*\*\*Assignment 1 (Reading and Writing Datasets) due 9PM\*\*\***

### **September 28. Labeling Variables, Values, Datasets.**

Ch. 4 from Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.

### **October 5. Planning the Data Management Workflow.**

Ch. 3, 5 from Long, J. Scott. 2009. *The Workflow of Data Analysis Using Data*. Stata Press.

### **October 12. Automating Your Work.**

Ch. 9 from Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.

**\*\*\*Assignment 2 (Recoding and Labeling) due 9PM\*\*\***

### **October 19. Combining and Reshaping Data Files.**

Ch. 6, 8 from Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.

### **October 26. Processing Observations Across Subgroups.**

Ch. 7 from Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. Stata Press.

**\*\*\*Assignment 3 (Automating Your Work) due 9PM\*\*\***

### **November 2. Dealing with Missing Data.**

Allison, Paul D. 2009. "Missing Data." Chapter 4 in *The SAGE Handbook of Quantitative Methods in Psychology*, edited by Roger E. Millsap and Alberto Maydeu-Olivares. Sage Publications. <http://www.statisticalhorizons.com/wp-content/uploads/2012/01/Milsap-Allison.pdf>

Johnson, David R., and Rebekah Young. 2011. "Toward Best Practices in Analyzing Datasets with Missing Data: Comparisons and Recommendations." *Journal of Marriage and Family* 73:926-945.

**November 9. Do File Programming.**

Ch. 4, 6, 7 from Baum, Christopher F. 2016. An Introduction to Stata Programming, 2<sup>nd</sup> ed. Stata Press.

**\*\*\*Assignment 4 (Merging, Reshaping, and Subgroups) due 9PM\*\*\***

**November 16. Do File Programming (continued).**

Ch. 8, 9, 10 from Baum, Christopher F. 2016. An Introduction to Stata Programming, 2<sup>nd</sup> ed. Stata Press.

**November 23: No class, Thanksgiving break**

**November 30. Ado File Programming.**

Ch. 11 from Baum, Christopher F. 2016. An Introduction to Stata Programming, 2<sup>nd</sup> ed. Stata Press.

**\*\*\*Assignment 5 (Missing Data) due 9PM\*\*\***

**December 7. Ado File Programming (continued).**

Ch. 12 from Baum, Christopher F. 2016. An Introduction to Stata Programming, 2<sup>nd</sup> ed. Stata Press.

**\*\*\*Assignment 6 (Programming) due December 14 at 9PM\*\*\***