SOCY7709: QUANTITATIVE DATA MANAGEMENT

Instructor: Natasha Sarkisian   Class time: Wednesdays 2:30-4:50 PM
Email: natasha@sarkisian.net   Class location: O'Neill 245
Phone: (617) 755-3178   Office hours: By appointment
Office: McGuinn 417   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 software 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 topic, I will give a lecture involving a demonstration in Stata, and you will get a chance to follow along and do all the necessary data management steps in Stata on your laptop. After that, you will work on applying these 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. Please check your email regularly: I will send out announcements from time to time.

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, at the Connors Family Learning Center regarding learning disabilities and ADHD, or Paulette Durrett, (617) 552-3470, 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.

Feedback: I would like to know how I could make this course experience as useful and interesting as possible. Therefore, every class session, in the end of class I will ask you to go to this link: https://goo.gl/forms/GhvgZ15aZ7A02m452 (also available as QR code below) and submit your feedback, indicating what you learned, as well as specifying what 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!

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, 2nd edition. Stata Press.

Other required readings will be available on electronic reserve in the library:
http://www.bc.edu/reserves

There will be five 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 20% of your grade.

Assignment	Due date
Assignment 1: Reading and Writing Datasets	February 13
Assignment 2: Recoding and Labeling	March 20
Assignment 3: Merging and Reshaping	April 10
Assignment 4: Subgroups, Macros, Loops	April 24
Assignment 5: Missing Data	May 13

COURSE OUTLINE

January 16. Introduction to Data Management using Stata.

January 23: No class (Monday schedule)

January 30. Reading and Writing Datasets.
Ch. 1, 2 from Mitchell, Michael N. 2010. Data Management Using Stata: A Practical Handbook.
Stata Press.

February 6. Data Cleaning.
Ch. 3 from Mitchell, Michael N. 2010. Data Management Using Stata: A Practical Handbook.
Stata Press.

February 13. 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

February 20. Creating and Changing Variables (continued).

February 27. Labeling Variables, Values, Datasets.
Ch. 4 from Mitchell, Michael N. 2010. Data Management Using Stata: A Practical Handbook.
Stata Press.

March 13. Planning the Data Management Workflow.
Ch. 3, 5 from Long, J. Scott. 2009. The Workflow of Data Analysis Using Stata. Stata Press.

March 20. Automating Your Work (Loops and Macros).
Ch. 9 from Mitchell, Michael N. 2010. Data Management Using Stata: A Practical Handbook.
Stata Press.
Assignment 2 (Recoding and Labeling) due 9PM

March 27. Combining and Reshaping Data Files.
Ch. 6, 8 from Mitchell, Michael N. 2010. Data Management Using Stata: A Practical Handbook.
Stata Press.

April 3. Processing Observations Across Subgroups.
Ch. 7 from Mitchell, Michael N. 2010. Data Management Using Stata: A Practical Handbook.
Stata Press.

April 10. 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. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1741-3737.2011.00861.x
Assignment 3 (Merging and Reshaping) due 9PM*

April 17. Missing Data (continued).
Wulff, Jesper N. and Linda Ejlskov. 2017. "Multiple Imputation by Chained Equations in Praxis:
Guidelines and Review." *The Electronic Journal of Business Research Methods* 15(1):41-56.
http://www.ejbrm.com/issue/download.html?idArticle=450

April 24. Stata Programming.
Ch. 4, 6-10 from Baum, Christopher F. 2016. An Introduction to Stata Programming, 2nd ed.
Stata Press.
Assignment 4 (Subgroups, Macros, Loops) due 9PM*

May 1. Stata Programming (continued).
Ch. 11-12 from Baum, Christopher F. 2016. An Introduction to Stata Programming, 2nd ed. Stata
Press.
Assignment 5 (Missing Data) due May 13 at 9PM*