Sociology 511: Data Management
Fall 2016 (3 credits), Tues: 2:50-5:40, Wilson-Short 5 (and Wilson-Short 231)

Professor Thomas Rotolo
Office: Wilson Hall 252
Office Hours: TBA in class
Office Phone: 335-4414
E-mail: rotolo@wsu.edu (write **Soc 511** in subject line)

Introduction: This course will introduce you to core concepts and procedures that are used regularly by sociologists and researchers in other fields when conducting quantitative analysis of data. The techniques covered are those used regularly by sociologists and other researchers in universities and other settings. The skills are also necessary to succeed in subsequent statistics and methods courses offered by the Department of Sociology. We will cover basic data management and basic statistical procedures in Stata, a powerful statistical software package used by many social scientists, including most of the WSU Sociology faculty. As we learn a variety of data management tasks, we will review several core statistical topics that are required in more advanced statistics courses. You may need to use an introductory statistics textbook or internet sources in order to ensure that you understand the basic statistical concepts.

Most of the class will be dedicated to Stata applications, however we will also consider how another very popular (and free!) software package, R, can be used to address similar data management tasks. I will provide an overview of the distinction between Stata and R on the first day of class.

Goals of the course: The Sociology Graduate Program at Washington State University has three stated Graduate Program Objectives. This graduate course addresses the first objective: "To enable students to conduct original, empirical sociological research". The Graduate Program has a series of learning outcomes for this objective (see last page of the syllabus, Table 1).

Prerequisites: There are no formal course prerequisites for this class, although a familiarity with algebra and PC operating systems is helpful. In order to complete this class successfully, you need to have access to your own computer with a version of Stata installed or you will need to use the computer lab in Wilson-Short Hall (Room 231). All computers in this lab have Stata installed. You do not know anything about this statistical software package, or any other software, however some experience with other statistical software, along with knowledge of a spreadsheet software package such as Microsoft Excel will ease the learning process. It will also help if you have completed successfully at least one introductory statistics course.

We will use two books, available through the WSU Bookie or from other retailers:

1) **Mitchell, Michael N. 2010. Data Management Using Stata: A Practical Handbook. College Station, TX: Stata Press**

We will progress sequentially through the Mitchell text.

2) **Gaubatz, Kurt Taylor. 2015. A Survivor's Guide to R. Los Angeles, CA: Sage Publications.**

There are also *many* helpful R resources on-line.

For those interested in mastering Stata, I also recommend finding a version of:

Hamilton, Lawrence. ***Statistics with Stata.***

Acock, Alan. ***A Gentle Introduction to Stata***

Grading

Your final grade will be based on your performance on two large, take-home projects, occasional weekly assignments, and participation in class. The large projects may include statistics-related questions, as well as data management tasks. Some parts of the projects will require you to obtain your own data sets so these tasks may be assigned well in advance of the due date. The exact format of the projects will be discussed in class.

The weekly assignments will be related directly to the procedures covered in the textbook and in class. Many of the weekly assignments will be informal exercises that require you to get the software to "do something". Some weekly assignments will require simple replication, others might require you to replicate commands using your own data, and near the end of the semester assignments might require you to "put everything together". Most weekly assignments will be graded as "pass" (completed) or "fail" (not completed). If weekly assignments are graded, I will use a 3-point scale: excellent (100%), adequate (85%), incomplete (0%), unless announced otherwise.

You may be required to submit your work to the course web space on learn.wsu.edu. I will demonstrate the submission process in class. Most of the time, I will ask you to discuss the assignments in class with the intent to encourage class discussion. If it becomes evident that students have not completed the assignment before class, I will require you to submit a formal assignment to the web space within 24 hours. Submission details will be provided in class should this unfortunate situation occur.

Large project due dates and grading percentage breakdown

Thursday, October 13	(Project #1):	20%
Thursday, December 1	(Project #2):	20%

If you need to miss an exam, I would appreciate a notice of at least 24 hours, as circumstances allow. Make-up exams will be arranged on a case-by-case basis.

Weekly Assignments:	45%

Attendance / Participation	15%

Regular class attendance and active participation in class is required

This course will adhere to the following grading scheme, although I reserve the right to alter the percentages in your favor at the end of the semester.

A:	**94-100%**	**C+:**	**77-79%**	**F:**	**59% and below**
A-:	**90-93%**	**C:**	**74-76%**		
B+:	**87-89%**	**C-:**	**68-73%**		
B:	**84-86%**	**D+:**	**67-69%**		
B-:	**80-83%**	**D:**	**60-66%**		

Some Random Thoughts: At certain points throughout the semester, you might find yourself in a dark alley with the computer. Here are some general tips: First, can you get back to where you started? If so, try to redo the analysis. I know this sounds like a lot of work (and in many cases this suggestion is impractical), but for most of the exercises we will be working on in this class, starting over can often be the best plan of action.

Second, ask yourself if this error has happened before? Have you seen a similar error message? If so, look back through your notes to see what you did then. Third, when all else fails read the documentation. Stata help documentation is quite good or you may even pick up something that will help later, even if it didn't help you this time. And, there is one last hope: Ask me (or a friend). I'm competent with Stata and most of its unusual behaviors; I've probably, at one time or another, faced the same problem you are facing. If you having trouble with R, the internet is the best place to search for help.

Schedule

The table below provides an **approximate** schedule for the semester, based on the Mitchell textbook. The schedule is approximate because we will develop a pace that suits the needs and experience of the class. Our pace also depends on how much R we decide to incorporate into the course. Some of the topics might be covered in ten minutes while other topics might take an entire class period to cover. Some topics might extend over several class periods. The material is cumulative so we will often revisit topics learned early in the semester.

Week 1 (August 23)

Introductions / Introduction to course
Introduction to Stata (and R): Strengths and weaknesses

Introductory Statistics Review

Week 2 (August 30)

Mitchell Chapter 1: Using archival data; using your own data
Mitchell Appendix A: Sections A1-A3

Summarizing data (tables, graphs); summarizing data (central tendency, dispersion)

Basic Stata syntax: List, Describe, Edit

Week 3 (September 6)

Mitchell Chapter 2: Reading and writing datasets

The shape of a distribution (skew); the normal distribution

Stata operators
Reading raw data into Stata
Types of datasets
Saving data files
Graphing in Stata
The shape of a distribution

Week 4 (September 13)

Mitchell, Chapter 3: Data cleaning

Using the normal distribution

Checking your data
Correcting errors in data
Identifying duplicates

Week 5 (September 20)

Mitchell, Chapter 4: Labeling

Describing data
Labeling: Variables, values
Labeling utilities
Comments in datasets
Standardized variables

Week 6 (September 27)

Mitchell, Chapter 4 continued
Statistical inference, sampling distribution

Formatting display of variables
Changing order of variables in dataset
Appendix A.5: Data types

Week 7 (October 4)

Mitchell, Chapter 5: Creating variables
Rowntree, Chapter 5 (Pp. 94-101): Estimation and inference

Creating and recoding variables
Numeric expressions and functions
Appendix A.6: Logical expressions
Missing value coding
Dummy variable creation

Week 8 (October 11): Project #1 due Thursday, October 13. Details in class.

Week 9 (October 18)

Mitchell Chapter 5: Creating variables (continued)
Hypothesis testing, statistical significance

Date variables
Date and time variables
Computations across variables
Computations across observations
The "egen" command
Converting string variables to numeric variables
Renaming variables

Week 10 (October 25)

Mitchell, Chapter 6: Combining datasets
One- versus two-tailed tests

Appending
Merging
Joining
Crossing
One-sided versus two-sided tests

Week 11 (November 1)

Mitchell, Chapter 7: Subgroup processing
z-tests versus t-tests; t-distribution

Separate results for subgroups; computing values by subgroups
Computing values within subgroups: Subscripting observations
Computing values within subgroups: Computations across observations
Computing values within subgroups: Running sums
Appendix A.8-A.9: Subsetting observations with "if and in" and with "keep and drop"

Week 12 (November 8)

Mitchell, Chapter 8: Changing the shape of data
Correlation

Wide and long datasets
Reshaping long to wide / Reshaping wide to long

Week 13 (November 15)

Mitchell, Chapter 8

Multilevel datasets; collapsing datasets

November 21-25: Thanksgiving break

Week 14 (November 29)

Project #2 due Thursday, December 1. Details in class

Week 15 (December 6)

Analyzing relationships: Prediction and regression

Attendance: I will note attendance in this class each week. I will not provide any lecture materials outside of class, so if you miss class, you will need to get the notes from another student. If you don't know anyone in the class, please make a friend early in the semester so you can rely on this person if you need any notes later in the semester.

Academic Integrity: There is no excuse for using someone else's work or ideas and turning them in as your own. Academic integrity will be strongly enforced in this course. Any student caught cheating on any assignment or exam will be given an F grade for the course and will be reported to the Office Student Standards and Accountability. Cheating is defined in the Standards for Student Conduct WAC 504-26-010 (3). You should read and understand the definitions in the conduct standards document.

Students with Disabilities: Reasonable accommodations are available for students with a documented disability. If you have a disability and need accommodations to fully participate in this class, please either visit or call the Access Center (Washington Building 217; 509-335-3417) to schedule an appointment with an Access Advisor. All accommodations MUST be approved through the Access Center.

Safety information: Washington State University is committed to enhancing the safety of the students, faculty, staff, and visitors. It is highly recommended that you review the Campus Safety Plan (http://safetyplan.wsu.edu/) and visit the Office of Emergency Management web site (http://oem.wsu.edu/) for a comprehensive listing of university policies, procedures, statistics, and information related to campus safety, emergency management, and the health and welfare of the campus community.

Table 1: Goals of the course

Sociology Graduate Learning Outcome	**Course topics and dates that advance these learning goals**	**The objective will be evaluated primarily by**
1. Formulate an important, viable sociological research question within field of expertise, including situating that research question within the relevant literature.	The development of a relevant sociological research question precedes data collection and determines data management tasks. Specific dates of note: Week 1: Introduction to sociological research Week 2: Identifying sociological data; Describing variables Weeks 4-5: Reading and writing your own datasets; archival data used regularly by sociologists Week 8: Sociological literature search Weeks 13-14: Statistical association in sociology Week 8 and Week 14: Projects relate to this outcome	Assignments Projects #1 and #2
Sociology Graduate Learning Outcome	**Course topics and dates that advance these learning goals**	**The objective will be evaluated primarily by:**
2. Collect or identify data from which to draw conclusions about the research question.	Data management is an essential part of the collection and identification of data. The course will always stress the importance of drawing proper conclusions from data analysis. Specific dates of note: Weeks 3-5: Using archival data; data cleaning Week 9-11: Hypothesis testing and statistical significance; One- and two-sided tests; z-tests / t-tests Week 8 and Week 14: Projects relate to this outcome	Assignments Projects #1 and #2
3. Appropriately analyze research data and present results in a clear, concise manner.	Each week of the course will address important issues related to the collection and identification of data, by covering specific data management skills in Stata. Specific dates of note: Weeks 4-5: The normal distribution; standardized variables Week 11: Scale / index construction Week 8 and Week 14: Projects relate to this outcome	Assignments Projects #1 and #2
4. Draw implications about sociological knowledge from research findings.	Research findings will be discussed in the context of evaluations of statistical association between sociological variables. Specific dates of note: Weeks 13, 15: Statistical association Week 8 and Week 14: Projects relate to this outcome	Assignments Projects #1 and #2