GRADUATE SEMINAR IN DATA MANAGEMENT
Florida State University
CCJ6741 • Spring 2021 • EPS 0215 • T 10:00am-12:30pm

Course Instructor: Dr. Sonja Siennick
Contact Information: 405 Eppes Hall, 850-645-9265, ssiennick@fsu.edu
Student Hours: Zoom meetings by appointment

Course Description

Most researchers spend far more time examining, cleaning, preparing, and documenting their data than they do actually analyzing those data. This course introduces students to key issues in and strategies for the preparation of quantitative data for analysis, or the transformation of “raw” data into an analysis-ready dataset. We will give special attention to multilevel and panel survey data. Although there are no prerequisites for the course, it is best-suited for doctoral students who have completed our basic methods and statistics sequence (CCJ5705, 5706, and 5740) or who otherwise have gained familiarity with descriptive statistics, linear and logistic regression, and research design. Basic familiarity with Stata or another similar program (e.g., SPSS, SAS) is assumed.

We will use Stata and other software in this course, but this is a “skills” course as much as it is a “tools” course. Software changes so rapidly that any specific syntax we create soon could be outdated. Our emphasis thus will be on the principles and practice of data management and the development of your intuition and logical skills. We will use software and syntax in service of those emphases.

Upon successfully completing this course, you will know the following:
- The philosophy and goals of data management
- The types and sources of missing data
- Strategies for dealing with missing data
- Key issues in the protection of human subjects and data security
- Rationales for data sharing and the basics of data management plans

You will be able to do the following:
- Extract, merge, and restructure raw data files
- Write more efficient syntax
- Check raw data for errors and outliers
- Transform non-normally distributed variables
- Create publication-ready tables and graphs
**Required Materials**

**Readings.** You will read several articles, chapters, and guides. Unless otherwise noted, please complete the readings before the class meeting for which they are assigned. Required readings are starred (*) on the course schedule.

**Software.** You will need access to Stata and Microsoft Excel. These programs are available in the graduate student computer lab or through FSU’s myFSUVLab. If you need a Stata refresher, review Stata’s Data-Management Reference Manual (https://www.stata.com/manuals/d.pdf) or one of the many books and websites devoted to this topic. If you wish to use the Virtual Lab, please note that you may have to contact IT Support to install some user-written commands. If you wish to use your personal copy of Stata, please note that older versions may not have the commands you will need. You should use SE or Intercooled Stata because Small Stata puts major limits on the size of your datasets. You also will need internet access.

**Course Requirements**

**Assignments.** You will complete eight assignments. Together they will make up 80% of your grade. Most will require you to obtain, clean and impute data and to document your work. Some will require you to present results. One will require you to give an in-class presentation. Detailed instructions for each assignment will be given in class. Timely submission is a must! In some cases, your fellow students will need your work before they can proceed with theirs. Missing or late assignments will receive 0s.

**Participation.** The success of this course truly depends on your active participation. I expect you to attend each class meeting. I also expect you to respect your fellow students’ work and time by coming to class prepared and by completing and submitting all of your own assigned work on time. If you don’t do these things, you will receive few participation points. Participation is worth 10% of your grade.

**Judging.** Your final task in the course, worth 10% of your grade, will be to judge your favorite student-produced deliverables. Competition will be in two categories: syntax files and codebooks. The prizes are the esteem of your colleagues and a sense of self-satisfaction. Yes, this means that your colleagues will be reading your work. That is part and parcel of being a data manager!

**Grading scale**

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>94%-100%</td>
<td>A</td>
</tr>
<tr>
<td>90%-93%</td>
<td>A-</td>
</tr>
<tr>
<td>87%-89%</td>
<td>B+</td>
</tr>
<tr>
<td>84%-86%</td>
<td>B</td>
</tr>
<tr>
<td>80%-83%</td>
<td>B-</td>
</tr>
<tr>
<td>77%-79%</td>
<td>C+</td>
</tr>
<tr>
<td>74%-76%</td>
<td>C</td>
</tr>
<tr>
<td>70%-73%</td>
<td>C-</td>
</tr>
<tr>
<td>60%-69%</td>
<td>D</td>
</tr>
<tr>
<td>Below 60%</td>
<td>F</td>
</tr>
</tbody>
</table>

To protect your privacy, I cannot discuss grades over email.
Course Expectations

Working with raw data can be challenging and at times frustrating, but the best way to learn many of the things we will cover is to try them and to be persistent when they do not work the first time. I will teach and demonstrate important topics and techniques, but I am no substitute for hands-on experience. I feel your pain (believe me, I really do), but I am going to be firm about this: I am happy to help you with the technical aspects of the assignments AFTER you have tried to resolve any issues on your own at least THREE times. If and when you reach that point, don’t be shy about bringing me your questions, syntax, datasets, and anything else you would like me to review.

I strive to make my classroom a safe, respectful, and effective learning environment. I need you to help me do this. Please do not arrive late, sleep, eat, read outside material, hold side conversations, use your cell phone for anything, or use your laptop for purposes unrelated to the material. I also expect you to treat me and each other with respect.

Please check Canvas regularly for course announcements and changes.

University Policies

University attendance policy. Excused absences include documented illness, deaths in the family and other documented crises, call to active military duty or jury duty, religious holy days, and official University activities. These absences will be accommodated in a way that does not arbitrarily penalize students who have a valid excuse. Consideration will also be given to students whose dependent children experience serious illness.

Academic Honor Policy. The Florida State University Academic Honor Policy outlines the University’s expectations for the integrity of students’ academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process. Students are responsible for reading the Academic Honor Policy and for living up to their pledge to “...be honest and truthful and...[to] strive for personal and institutional integrity at Florida State University.” (Florida State University Academic Honor Policy, found at http://fda.fsu.edu/Academics/Academic-Honor-Policy)

Americans with Disabilities Act. Students with disabilities needing academic accommodation should:
1. register with and provide documentation to the Office of Accessibility Services; and
2. request a letter from the Office of Accessibility Services to be sent to the instructor indicating the need for accommodation and what type; and
3. meet (in person, via phone, email, skype, zoom, etc…) with each instructor to whom a letter of accommodation was sent to review approved accommodations.

This syllabus and other class materials are available in alternative format upon request.

For the latest version of this statement and more information about services available to FSU students with disabilities, contact the:
Human subjects research. The Graduate Student Handbook reminds students that “students, faculty or employees who conduct research involving human subjects when such data will be available for generalized knowledge, requires review and approval by the Institutional Human Subjects Committee.” If you wish to use the class data to pursue your own independent research, please be sure to obtain HSC (IRB) approval first.

Syllabus change policy. Except for changes that substantially affect implementation of the grading statement, this syllabus is a guide for the course and is subject to change with advance notice.
COURSE SCHEDULE

* REQUIRED READING
† RECOMMENDED READING

January 12  Course Introduction
Topics: What is data management?; a brief refresher on Stata.
Readings: Read online Stata tutorials if you need a refresher.

January 19  Raw Data Basics
Topics: Codebooks; extractions; skip patterns; the NLSY97.
Readings:
An overview of the steps and considerations in the programmer/analyst’s “workflow.” If you use Stata regularly I recommend you purchase the book. An overview of key issues in data compilation and archiving.
* From Add Health: sect23.pdf (posted to Canvas)
  fc_sect23.pdf (posted to Canvas)
  Example codebook, skip pattern, and question map.
For next time: Find a partner and send me your team’s ranked list of construct preferences.

January 26  Lab Session: NLSY97 Practice
Topics: Extracting NLSY97 data.
Readings: Visit the NLS Investigator and read the Investigator Guide:
https://www.nlsinfo.org/investigator/pages/search.jsp?s=NLSY97
https://www.nlsinfo.org/content/cohorts/nlsy97/topical-guide
Please give special attention to the portions of the Guide that deal with your construct.
For next time: Assignment 1: Use the NLS Investigator to extract variables needed for your assigned construct; create question maps for the resulting dataset.
February 2  Cleaning Data
Topics: Exploring and cleaning data.
Readings:
  Should we clean “messy” data at all? A survey of faculty in three disciplines.
  Tips on ensuring that your data and recodes are correct.
  Strategies for the pre-analysis exploration of variables and relationships.

February 9  Lab Session: Intermediate Programming in Stata
Topics: Annotation; automation and repetition; other intermediate programming topics.
Readings:
  Helpful introductions to Stata macros and the logic of loops.
For next time: Assignment 2: Submit your dataset and codebook.

February 16  Data Structures
Topics: Data structures; merging; multilevel and panel data.
Readings:
  The basics of the relevant Stata commands.
  For the last two: What do you think these datasets looked like? Try to sketch them out and start thinking about the necessary steps in syntax.
For next time: Assignment 3: Merge all class data files; restructure.
February 23  Missing Data: Theory
Topics: Missing data (sources and treatments); review of linear, logistic, ordinal regression.

Readings:
  *A criminology-specific discussion of the key issues.*
  *Another treatment with some additional ideas.*
  *A nontechnical explanation of the processes that generate missing data and methods to address them, including multiple imputation.*

March 2  Missing data: Practice
Topics: Implementing multiple imputation in Stata.

Readings:
  *The “how-to” manual for Stata’s new canned MI facility.*

For next time: Assignment 4: Comps question.

March 9  Lab Session: Multiple Imputation
Readings: None

For next time: Assignment 5: Regression analyses and comparison of results under different treatments of missing data.

March 16  Complex Survey Data
Topics: Complex sampling designs; weights; design effects.

Readings:
  *Tips on when to weight (and when not to weight).*
* “Sample Weights and Design Effects.” Available at [https://www.nlsinfo.org/content/cohorts/nlsy97/using-and-understanding-the-data/sample-weights-design-effects](https://www.nlsinfo.org/content/cohorts/nlsy97/using-and-understanding-the-data/sample-weights-design-effects). 
  *NLSY97-specific information.*
March 23  Presenting Results in Tables
Topics: Creating tables; interpretation of tables.
Readings: None
For next time: Assignment 6: Comps question.

March 30  Lab Session: Predicted Values and Graphing
Topics: Predicted values; graphing; creating interaction terms; graphing interactions.
Readings: None
For next time: Assignment 7: Graphing interactions.

April 6  Data Sharing and Security
Topics: Human subjects protections; data security; data sharing; data management plans.
Readings:

For next time: Assignment 8: Identify and prepare a presentation about a Stata tip or trick; make any desired changes to your syntax and codebook and submit them for judging.

April 13  More Cool Things You Can Do in Stata
Topics: Student-identified tips and tricks in Stata programming.
Readings: None
For finals week (due 4/20): Judging of colleagues’ deliverables.