Tufts University

Introduction to Data Management in Stata (CH 99-20)

Syllabus  Spring 2017

Updated 01/27/17

INSTRUCTOR & CONTACT INFORMATION
Andrea Acevedo, PhD
Community Health Department  Tufts University
574 Boston Avenue, Second Floor 208
Medford, MA  02155

Email: andrea.acevedo@tufts.edu

Phone: (617)627-2151

Grader: Christian Chamars  Email: c.chamars@gmail.com

OFFICE HOURS:
Mondays 11am-12pm
Additional office hours available by appointment. You can also email both Christian and I about
questions related to the assignment. Please always include both of us, and one of us will get back
to you within 24 hours if it’s a weekday or within 48 hours if it is a weekend.

CLASS SCHEDULE
The course meets on Mondays and Wednesdays from 3:00pm-4:15pm in the Mark Lab in the
Tisch Library.

COURSE DESCRIPTION
Stata is a statistical software package that is widely used for data management and data analysis.
The purpose of this course is to introduce students to basic programming in Stata and to provide
guidance on data management strategies for health related research. The course will focus on
command-based programming for modifying and managing data, and performing simple
statistical analysis in Stata. By the end of the course students will be able to comfortably
navigate the Stata environment, create simple datasets, access existing datasets, create variables,
use graphing functions, and run commands to calculate summary statistics as well as basic
inferential statistics. This course is intended for students who have no or minimal experience
using this statistical software program.
COURSE PRE-REQUISITES:

Students should be a Community Health major, and should have taken a statistics course prior to taking the course – CH31 or one of its equivalents.

COURSE OBJECTIVES

By the end of this course, students will be able to:

- Comfortably navigate the Stata environment
- Describe a dataset
- Create variables
- Use basic graphic functions
- Create new datasets
- Use strategies for data cleaning
- Conduct basic descriptive and inferential statistics
- Identify potential data sets for addressing research questions
- Apply techniques learned to answer research questions

REQUIRED MATERIALS & TEXTS

There are no textbooks for this course. Some readings will be made available at the course website on Trunk (trunk.tufts.edu) or distributed in class prior to specific sessions. You are responsible for reading all assigned material before the class date for which the readings are assigned.

These books are not required, but if you’re like me and like to have a reference source handy, these are some useful, beginner’s types books:


GRADING & EVALUATION

Final grades for this course will be determined by the following assessments:

1. Class Participation 10%
2. In-Class assignments 10%
3. Quiz 10%
4. Homework Assignments 35%
5. Final project 30%
6. Presentation 5%
**Class Participation (10%)**
Class participation includes attendance and coming to class on time, being prepared to discuss assignments, asking relevant questions, listening to other students and the instructor, and being engaged with the material. Students who miss class (unless with note from medical staff or their alpha Dean), who use computers for things other than class activities, or use cell phones, will have lower participation grades.

**In-Class Assignments (10%)**
In-class assignments will allow you to practice some of the new techniques you learn in class. In-class assignments cannot be made up if you miss class.

**Homework Assignments (35%)**
Assignments will include a mixture of practice exercises as well as parts of your final project.

**Final Project (30%)**
Your final project will involve identifying a dataset, developing research questions, and using the skills you learned in this class to answer your research questions. It will include a brief introduction, a methods section, a section on results, graphic representations of the sample and/or results, and a brief discussion. Parts of the project will be due throughout the course. Detailed instructions will be provided for each section. Maximum numbers of Pages: 10 (double-spaced).

**Presentation (5%)**
You will be required to make a 5-7 minute presentation of your project.

**COURSE POLICIES**

**Submitting Written Assignments:** Assignments should be formatted based on the homework sample (uploaded in Trunk). All assignments are to be submitted through Trunk by the due date and time.

**Late Assignments:** Because this is a skills based course which requires practice and topics tend to build on what was learned before, it is critical that assignments be completed on time. Late assignments will be accepted only with prior instructor approval and will receive a 10% reduction in grade for each 24 hour period for which it is late. Late begins at 1 minute past the deadline. If your assignment is late, you will not be able to submit the assignment via Trunk. Therefore, submit the assignment to Christian Chamars (and cc me) directly via email.
Cellphones:
No cell phone use during class.

ACADEMIC INTEGRITY
Tufts holds its students strictly accountable for adherence to academic integrity. The consequences for violations can be severe. It is critical that you understand the requirements of ethical behavior and academic work as described in Tufts’ Academic Integrity handbook. If you ever have a question about the expectations concerning a particular assignment or project in this course, be sure to ask for clarification. The Faculty of the School of Arts and Sciences and the School of Engineering are required to report suspected cases of academic integrity violations to the Dean of Student Affairs Office. If cheating or plagiarism is suspected, this must be reported to the Dean. More information is available at: http://uss.tufts.edu/studentAffairs/documents/HandbookAcademicIntegrity.pdf

STUDENT SERVICES
Tufts is committed to providing support and equal access for all students so that they may fully realize their academic potential. If you need academic accommodations because of a disability, please contact the course director early in the semester so that your learning needs may be appropriately met. All discussions will remain confidential. More information can be found at the Tufts University Student Accessibility Services website: https://students.tufts.edu/student-accessibility-services

Accommodations: Students needing academic adjustments or accommodations because of a documented disability must present documentation from the Academic Resource Center. All discussions remain confidential.

Additional Resources:
• Tisch Library: For research assistance- http://www.library.tufts.edu/tisch/subjectLib.html
• Academic Resource Center (ARC): For writing and other academic assistance: http://uss.tufts.edu/arc/
• The Tufts University Counseling and Mental Health Service (CMHS). For personal and academic concerns- Sawyer House at 120 Curtis Street. 617.627.3360. http://ase.tufts.edu/counseling/
# Course Calendar

Please note that the order and schedule of topics may shift based on student needs.

<table>
<thead>
<tr>
<th>Class</th>
<th>Day</th>
<th>Date</th>
<th>Class Topic</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Monday</td>
<td>January 23</td>
<td>Introductions</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Wednesday</td>
<td>January 25</td>
<td>do-files and log files and variable types</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exploring a dataset</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Monday</td>
<td>January 30</td>
<td>Exploring a dataset (cont’d)</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Wednesday</td>
<td>Feb. 1</td>
<td>Accessing help within Stata and other resources</td>
<td><strong>Assignment #1 Due Sunday 2/5/17 by 11:55pm</strong></td>
</tr>
<tr>
<td>5</td>
<td>Monday</td>
<td>Feb. 6</td>
<td>Variable labels &amp; values</td>
<td><strong>Assignment #2 Due Sunday 2/12/17 by 11:55pm</strong></td>
</tr>
<tr>
<td>6</td>
<td>Wednesday</td>
<td>Feb. 8</td>
<td>Creating a dataset / Creating new variables, Part 1: using arithmetic and relational expressions; converting string to numeric and vice-versa</td>
<td><strong>Assignment #3 Due Sunday 2/19/17 by 11:55pm</strong></td>
</tr>
<tr>
<td>7</td>
<td>Monday</td>
<td>Feb. 13</td>
<td>Creating new variables, Part 2: Recoding</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Wednesday</td>
<td>Feb. 15th</td>
<td>Discussion of final assignment / Introduction to ICPSR and other ways of finding datasets</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monday</td>
<td>Feb. 20th</td>
<td><strong>NO CLASS</strong> ** President’s Day / University Holiday**</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Wednesday</td>
<td>Feb. 22</td>
<td>Working with dates and time</td>
<td><strong>Assignment #3 Due Sunday 2/26/17 by 11:55pm</strong></td>
</tr>
<tr>
<td>10</td>
<td>Thursday</td>
<td>Feb. 23</td>
<td>Dropping/keeping variables; Selecting observations</td>
<td><strong>Quiz</strong> (due to another class scheduled in the lab, this class will end at 4pm)</td>
</tr>
<tr>
<td>11</td>
<td>Monday</td>
<td>Feb. 27</td>
<td>Importing data in other formats / Combining datasets to add observations</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Wednesday</td>
<td>March 1</td>
<td>Merging datasets</td>
<td><strong>Assignment #4 Due Sunday 3/5/17 by 11:55pm</strong></td>
</tr>
<tr>
<td>13</td>
<td>Monday</td>
<td>March 6</td>
<td>Graphing</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Wednesday</td>
<td>March 8</td>
<td>Analyzing data: Epidemiology</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Monday</td>
<td>March 13</td>
<td>Analyzing data: ttests and ANOVA’s</td>
<td><strong>Assignment #5 Due Sunday 3/19/17 by 11:55pm</strong></td>
</tr>
<tr>
<td>16</td>
<td>Wednesday</td>
<td>March 15</td>
<td>Analyzing data: Regressions</td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>Day</td>
<td>Date</td>
<td>Class Topic</td>
<td>Assignments</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Monday</td>
<td>March 20</td>
<td><strong>NO CLASS</strong> (Spring Recess)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wednesday</td>
<td>March 22</td>
<td><strong>NO CLASS</strong> (Spring Recess)</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Monday</td>
<td>March 27</td>
<td>Analyzing data: Logistic regressions</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Wednesday</td>
<td>March 29</td>
<td>Analyzing Data: Time-to-event</td>
<td>Assignment #6 Due Sunday 4/2/17 by 11:55pm</td>
</tr>
<tr>
<td>19</td>
<td>Monday</td>
<td>April 3</td>
<td>Long vs. wide datasets &amp; reshaping</td>
<td>DRAFT Project Due by Wednesday 4/5/17 11:55pm</td>
</tr>
<tr>
<td>20</td>
<td>Wednesday</td>
<td>April 5</td>
<td>Topic to be decided based on student interest</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Monday</td>
<td>April 10</td>
<td>TBD</td>
<td>Assignment #7 Due Tuesday 4/11/17 by 11:55pm</td>
</tr>
<tr>
<td>22</td>
<td>Wednesday</td>
<td>April 12</td>
<td>Peer Discussion of Draft Projects</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Monday</td>
<td>April 17</td>
<td><strong>No class</strong> Patriot’s Day / University Holiday</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Wednesday</td>
<td>April 19</td>
<td>Topic to be decided based on student interest</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Monday</td>
<td>April 24</td>
<td>Student presentations (Part 1)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Wednesday</td>
<td>April 26</td>
<td>Student presentations (Part 2) &amp; Wrap Up</td>
<td></td>
</tr>
</tbody>
</table>

Final Project Due:  Wednesday May 3rd, 11:55pm