Nutrition Data Analysis (NUTR311)
Updated 9/10/2018

Fridays 1:30pm-4:30pm, Sackler room 514

Robert F. Houser, Ph.D.
robert.houser@tufts.edu, Phone 617.636.3682 | Skype address: robyhouser
Office Hours: By appointment

Teaching Assistant: Robel Alemu
Robel.Alemu@tufts.edu
https://www.linkedin.com/in/robelalemu

TA duties and responsibilities

Robel is responsible for grading homework assignments and interacting with students about assignments, grading and grading feedback.

Tufts Graduate Credit: 1.0

Prerequisites for taking this course:

1. Two statistics courses are prerequisite (NUTR207 and NUTR307 or equivalent courses)
2. Students must have completed a research ethics on-line tutorial. If you have not done so, please sign up and complete the on-line tutorial found at ...
   http://www.tufts.edu/central/research/IRB/citi.htm
   More details are available at ...
   http://www.tufts.edu/central/research/IRB/HumanEducation.htm
3. It is highly recommended that students take a research methodology course before taking NUTR311 or take it concurrently. If you have not taken a research methodology course before and do not plan to take one concurrently, you are strongly encouraged to read ...

Course Description:

This course will cover knowledge of advanced Stata statistical computing, data base construction, error detection and correction, creation of composite variables,
Course Objectives:

Course provides students with important applied skills needed for careers in Nutrition and Public Health including ...`

1) how to apply statistical techniques to real datasets
2) proficiency in database design, data management and statistical programming
3) how to translate a research question into an analysis plan, research proposal and analytical strategy
4) how to determine what statistical tests and methods are appropriate for a given dataset and research question
5) how to write advanced level Stata programming code
6) how to interpret research findings and present them in tables, figures and text
7) how to analyze a dataset and produce a research article suitable for submission to an academic journal

Throughout the semester, students will learn how to apply data management and statistical analysis techniques to pre-selected data sets from nutrition and public health research projects. Emphasis will be on analytic fluency, on developing proficiency and confidence in thinking about data analysis problems, selecting and applying statistical methods and procedures, and formulating theories based on statistical findings. Computer lab exercises will provide hands-on experience, first with hands-on tutorials and progressively with less assistance. By the end of the course, students will be able to tackle routine analysis problems independently and to request and make use of specialized statistical expertise.

Many important skills are taught in the course that may be helpful to students when they apply for employment after graduation. Students may be able to list several important applied skills and experiences on their CV’s including database design, STATA statistical programming skills, student research conference presentations and submission of an article to an academic journal. Analyzing a dataset and writing a research article substitutes for the Master’s thesis required in some graduate programs and may also be used as a writing sample when graduates apply for employment.

NUTR311 is not designed to teach students how to obtain funding or plan research studies. It is designed to compliment courses such as Survey Research Design. NUTR311 assumes...
students know how to design studies and collect research data. NUTR311 is designed to prepare MS Nutrition students and MPH students to be able to work with health and nutrition data as research analysts and report writers. The course also helps prepare Ph.D. students for conducting dissertation research. Many of our graduates have worked as data analysts and researchers after graduating from our program.

No other course that Friedman school devotes so much time to hands on data analysis and statistical programming instruction. All classes are held at a computer lab with students analyzing data during almost every class session. Many hours of computer training are provided (far more than in any other Friedman course). The primary goal of the course is to provide students with useful data analysis skills that can help them obtain good jobs after graduation.

Students are encouraged to present their research findings at the annual Friedman Research Conference or submit their research article to an academic journal for consideration of publication.

**Description of assignments, tests, and other required activities:**

You MUST do your homework assignments by YOURSELF. You may only receive assistance from the course TA and instructor.

You should work with 1 or 2 other students on your research project analyses and papers and each of you must collaborate on all aspects of the work (i.e. You should not divide up the tasks and complete each part separately. You must each have a full understanding and involvement with each part.)

1st graded homework assigned on Friday, 9/16, due on Sunday, 9/30
Email to instructor your letter of intent (LOI). The LOI should contain information about the data set and research questions/hypotheses, methods and brief literature review for the research project you plan to conduct (due by Sunday 10/7).

2nd graded homework assigned on Friday, 10/6, due on Sunday, 10/21
Table shells and analysis plan due on Sunday, 10/28
IRB paperwork must be submitted to the IRB office by Friday, 11/9
3rd graded homework assigned on Friday, 11/9, due on Sunday 11/18
Results and discussion section due on Sunday, 12/1
Complete research article including references and abstract due on Sunday, 12/19
Assignments are due before midnight on due dates. Homework assignments should be turned in via Canvas and the research article should be turned in to the instructor via email.

The LOI and analysis plan must be submitted and accepted by the instructor before students begin analyzing data for the research article. Students MUST obtain permission or an exemption from the Tufts ethics IRB (Institutional Review Board) before working on the analyses for the research article and provide evidence to the instructor that approval or exemption has been approved.

Early in the semester, students will select a dataset to analyze. The instructor will provide students with a few choices. Students can use publicly available data sets such as NHANES, BRFSS, IFPRI datasets. They can also use data from faculty projects. In addition to IRB rules about secondary data, students should obtain written permission from faculty to use their data and complete an authorship agreement.

Authorship agreements for students using data from funded projects

There is always a very strict requirement from US funders to have specific wording used in acknowledging the grant support. Any papers emerging from analysis of such data need to include the funder and grant numbers, etc. When data are put together to support a publishable paper, it’s important to have key PIs and co-Is involved and listed as authors, especially when it would represent a first output (published or not) from a research work stream.

Please see the Friedman student policies and procedures manual (Appendix 1) for additional information about authorship guidelines.

Students working together must collaborate on all aspects of the work and will receive the same grade. Students must receive permission to begin work on the project by getting the approval of the instructor. Students should prepare an LOI (1-4 pages) of their research plans that describes the ...

a) proposed title of the paper
b) intended journal name
c) a description of the dataset
d) research questions and/or hypotheses that will be examined
e) brief description of the purpose and procedure
f) data analysis plan
g) a brief statement regarding the possible policy implications and/or possible usefulness of the research findings

Students who have not picked a data set to analyze by October 6 will be assigned data and research questions by the instructor.
Summary of Assignments and Grading

<table>
<thead>
<tr>
<th>Assignment(s)</th>
<th>Grading Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Assignment #1</td>
<td>9%</td>
</tr>
<tr>
<td>Homework Assignment #2</td>
<td>9%</td>
</tr>
<tr>
<td>Homework Assignment #3</td>
<td>9%</td>
</tr>
<tr>
<td>Average of 8 required online quizzes (For lessons 1,2,3,4,5,6, 11 &amp; 12)</td>
<td>18%</td>
</tr>
<tr>
<td>LOI including description of dataset, research question(s)/hypotheses and</td>
<td>10%</td>
</tr>
<tr>
<td>brief literature review</td>
<td></td>
</tr>
<tr>
<td>Table shells and analysis plan due on Sunday 10/26</td>
<td>5%</td>
</tr>
<tr>
<td>First draft Results and discussion section of paper</td>
<td>10%</td>
</tr>
<tr>
<td>Complete research article</td>
<td>30%</td>
</tr>
</tbody>
</table>

Penalties for late or incomplete assignments:

Class attendance policy: Please do not miss any classes and speak with the instructor if it is impossible for you to attend every class so you can make alternative arrangements such as audio recording the class session and obtaining class handouts and notes.

Material covered each week builds on material from previous weeks. So, it is EXTREMELY important that you do not miss any classes. If you are unable to attend a class, please contact one of the teaching assistants to request the class session be recorded and then follow up with the TA or instructor to make sure that you don't fall behind. If you miss more than 4 weeks of classes you will be given an incomplete and might be expected to retake the course next fall.

Late assignments: At the discretion of the instructor, a 5% penalty may be applied to each late assignment (papers, drafts and homework assignments).

People who are not enrolled in 311 might not be covered by the confidentiality permissions obtained for the datasets used in the course. Thus, no one should audit or sit in without first obtaining permission of the instructor (advanced notice is required so the instructor can seek permission well in advance for confidentiality permission to include such individuals).

Course texts and Materials (for the course as a whole):

Texts and manuals:
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Edition</th>
<th>ISBN-13:</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Juul</td>
<td>An Introduction to Stata for Health Researchers</td>
<td>3rd</td>
<td>978-1597180771</td>
<td>Recommended but not required</td>
</tr>
<tr>
<td>Kachigan</td>
<td>Multivariate Statistical Analysis: A Conceptual Introduction</td>
<td>2nd</td>
<td>978-0942154917</td>
<td>Recommended but not required</td>
</tr>
<tr>
<td>Treiman</td>
<td>Quantitative Data Analysis</td>
<td>2009</td>
<td>978-0470380031</td>
<td>Recommended but not required</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Title</td>
<td>Edition</td>
<td>ISBN</td>
<td>Online Access</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>---------</td>
<td>------</td>
<td>---------------</td>
</tr>
</tbody>
</table>

This text is relatively inexpensive (about $73.20 for the hardcover book, or $0.00 for the E-book). Because Tufts has a site license with Springer, the publisher, students may read the text online for free. The link is:

Software: STATA, ENA, and WHO anthropometry program for anthropometric calculation, Microsoft Access, Powerpoint. We will use STATA IC version 14 at the computer lab. If you purchase Stata this year, you should purchase Stata IC version 15.

You can use STATA on the Jaharis lounge computers or at Sackler library. You can obtain the software at ...
http://www.stata.com/order/new/edu/gradplans/gp-campus.html

The perpetual license for the “Stata/IC 15 and Getting Started manual” is your best option. There are leases available but the option that is the best long term value is the purchase of the Stata/IC 15 perpetual version for less than 200 dollars. Please wait until we meet in class to discuss software considerations before purchasing or leasing software.

Here is a summary of how the different “flavors” of Stata differ ...
http://www.stata.com/products/whichstata.html

Stata IC cannot directly read in files that have more than 2000 variables. However, Stata SE is usually not needed because it is possible to read variables from a very large file with the following “Stata magic trick” ...

Reading big Stata SE files with Stata IC ...

In the Stata command box you can get a list of the variable names in the big file with the “describe using” command (see below). Next, you can open a portion of the big file by specifying a list or range of variables in the big file (e.g. district-ndvi10_3).

describe using "I:\district yr mo collapsed (from Yoga).dta"

use district-ndvi10_3 using "I:\district yr mo collapsed (from Yoga).dta"
After accessing the variable you need you can save them in a new smaller file that Stata IC can open directly.

**Drawing software:** Draw.io  (Helpful for creating study design figures, conceptual framework figures, etc.  [https://www.draw.io/](https://www.draw.io/)

### Academic Conduct

Academic integrity, including avoiding plagiarism, is critically important. Each student is responsible for being familiar with the standards and policies outlined in the Friedman School’s *Policies and Procedures* manual ([http://nutrition.tufts.edu/student/documents](http://nutrition.tufts.edu/student/documents)). It is the responsibility of the student to be aware of, and comply with, these policies and standards. In accordance with Tufts University’s policy on academic misconduct, violations of standards of academic conduct will be sanctioned by penalties ranging from grade reduction or failure on an assignment; grade reduction or failure of a course; up to dismissal from the school, depending on the nature and context of any infraction.  


There are several required and optional online lessons and quizzes. Quizzes are “open book” comprehension quizzes. You can review the lessons and readings while taking the quizzes but you must do them by yourself. Although students can progress through the online lessons at their own pace, the recommended pace is shown below with “due dates” listed for each of the required lessons. Ph.D. students are encouraged to complete all of the lessons (both required and optional).  **Please save a screen shot of your quiz showing the answers BEFORE clicking on submit.**  Here is a list of the required and optional lessons ...

### Required Lessons

**Lesson 1** - Review of Introduction to multiple regression analysis  
Online Quiz 1 by 9/9

**Lesson 2**: Review of Multiple Regression Assumptions  
Online quiz 2 by 9/16

**Lesson 3**: Review of Assessing distributional normality & transforming variables  
Online quiz 3 by 9/16

**Lesson 4**: Building statistical models, dummy variables & interaction terms  
Online quiz 4 by 9/23

**Lesson 5**: Binary logistic regression analysis and polytomous logistic regression analysis  
Online quiz 5 by 10/14
Lesson 6: Presenting statistical results in tables, figures and text  
Online quiz 6 by 11/4

Lesson 11: Analysis of complex surveys  
Online quiz 11 by 11/11

Lesson 12: Study design and selection of appropriate statistical methods  
Online quiz 12 by 12/2

Optional Lessons & optional quizzes

Optional Lesson 7: Oneway analysis of variance and post hoc tests of statistical significance  
Optional quiz 7– no due date

Optional Lesson 8: Multi-factor Analysis of variance (ANOVA)  
Optional quiz 8– no due date

Optional Lesson 9: Repeated measures analysis of variance and mixed-design ANOVA  
Optional quiz 9– no due date

Optional Lesson 10: Analysis of covariance (ANCOVA)  
Optional quiz 10 – no due date

Optional Lesson 13: Cluster randomized trials  
Optional quiz 13 – no due date

Course & Assignment Schedule [to be adapted as necessary]:

<table>
<thead>
<tr>
<th>CLASS DATES &amp; LOCATION</th>
<th>WEEK OR SESSION</th>
<th>TOPIC OR CLASS TITLE</th>
<th>ASSIGNMENTS &amp; ACTIVITIES</th>
<th>LECTURER(S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sackler room 514 9/7</td>
<td>1</td>
<td>Introduction and course overview &amp; Stata software overview</td>
<td>Go over the syllabus in class and in class discussion of “working with Stata” and “accessing information on how to use Stata” Introduction to writing Stata programs (i.e. “do files”)</td>
<td>Robert Houser</td>
</tr>
<tr>
<td>Room</td>
<td>Week</td>
<td>Topic</td>
<td>Online Lesson</td>
<td>Online Quiz</td>
</tr>
<tr>
<td>---------------------</td>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>---------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Sackler room 514</td>
<td>9/14</td>
<td>Review of basic Stata commands &amp; anatomy of a basic Stata program</td>
<td>Online lesson 1 this week</td>
<td>Online Quiz 1 by 9/9</td>
</tr>
<tr>
<td></td>
<td>9/16</td>
<td>Getting started coming up with ideas for a research project, analyzing a dataset, and creating tables for a report or research article</td>
<td>Online lesson 2 this week</td>
<td>Online quiz 2 by 9/16</td>
</tr>
<tr>
<td></td>
<td>9/21</td>
<td>Preparing data for analysis and navigating IRB permissions for conducting research</td>
<td>Online lesson 3 this week</td>
<td>Online quiz 3 by 9/16</td>
</tr>
<tr>
<td></td>
<td>9/28</td>
<td>Creating composite and synthetic variables &amp; how to create a data analysis plan and how to write a research proposal</td>
<td>Online lesson 4</td>
<td>Online quiz 4 by 9/23</td>
</tr>
<tr>
<td></td>
<td>10/5</td>
<td>Research Group meetings</td>
<td>No class on Friday 10/11 Group meetings this week with Dr. Houser Homework 2 due on 10/21</td>
<td></td>
</tr>
<tr>
<td>Sackler room 514</td>
<td>10/19</td>
<td>Regression modeling with Stata &amp; Introduction to Report Writing</td>
<td>Online lesson 5 this week</td>
<td>Online quiz 5 by 10/20</td>
</tr>
<tr>
<td></td>
<td>10/26</td>
<td>Polytomous logistic regression modeling with Stata &amp; Report Writing (Part II)</td>
<td>Table shells and analysis plan due on Sunday 10/28</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11/3</td>
<td>Complex survey data analysis &amp; Analyzing data from a cluster randomized trial</td>
<td>Online lesson 6 this week</td>
<td>Online quiz 6 by 11/3</td>
</tr>
<tr>
<td>Sackler room 514</td>
<td>11/8</td>
<td>Research Group meetings</td>
<td>No class on Friday 11/8</td>
<td></td>
</tr>
<tr>
<td>Date</td>
<td>Time</td>
<td>Event Description</td>
<td>Location</td>
<td>Instructor</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>------------------</td>
</tr>
<tr>
<td>11/9</td>
<td></td>
<td>Group meetings this week with Dr. Houser</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Draft of literature review &amp; methods section due 11/11</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Homework 3 assigned on 11/9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRB paperwork submission to IRB office due 11/8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online lesson 11 this week</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online quiz 11 by 11/10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11/16</td>
<td>Sackler room 514</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Introduction to scale construction (Chronbach alpha &amp; Principle component analysis)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Homework 3 due on 11/18</td>
<td></td>
<td>Robert Houser</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thanksgiving break</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online lesson 12 this week</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Online quiz 12 by 12/2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11/30</td>
<td>Sackler room 514</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data Reduction Techniques for multivariate data</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Draft of results &amp; discussion due by 12/1</td>
<td></td>
<td>Robert Houser</td>
</tr>
<tr>
<td></td>
<td>12/7</td>
<td>Sackler room 514</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Panel data analysis &amp; Introduction to Multilevel modeling</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Research Paper Due on 12/19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This schedule is subject to modifications at the discretion of the instructor.
Academic Calendar for NUTR311

• Tuesday, September 4, 2018        Classes begin
• Monday, October 8, 2018          Indigenous People’s Day – NO CLASSES
• Monday, November 12, 2018        Veterans Day Observed – NO CLASSES
• Thursday-Friday, November 22 – 23 Thanksgiving Recess – NO CLASSES
• Friday, December 7, 2018         Last day of class for NUTR311
• Monday, December 10, 2018        Last day of classes for Fall 2018 Semester
Course Schedule
(Weekly Readings, Learning Objectives and Assignments)

Abbreviations:

QAD (“Quantitative Data Analysis: Doing Social Research to Test Ideas” by Donald J. Treiman


The following week schedule will be modified somewhat if moving more slowly or quickly is advantageous. NUTR311 is a “hands on” interactive course that takes place in a computer lab. Most courses rely on the lecture method and weekly assigned readings. Courses taking place in a lab environment are experiential and thus don’t conform to the traditional lecture style of non-lab based courses. So, much of the course “readings” involve students independently looking up information they need in Stata manuals, online web resources and research articles. Few formal readings will be assigned. Experiential learning takes place in real time. Readings are considered supplementary and texts should mostly be consulted as needed.

Week 1: Introduction and course overview & Stata software overview
Friday (9/7) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

<table>
<thead>
<tr>
<th>Learning Objectives for week 1: Upon completion of this week, students will</th>
</tr>
</thead>
<tbody>
<tr>
<td>Become familiar with the goals of the course by discussing the course syllabus. Go over the syllabus (books, readings, grading, paper, homework, forming groups, Stata software, etc.)</td>
</tr>
<tr>
<td>Understanding Stata files and Stata functions and procedures including “do files, “ado files”, and how to access Stata information resources (getting information on Stata commands, etc.)</td>
</tr>
<tr>
<td>Go over the assigned homework exercise</td>
</tr>
<tr>
<td>Discuss Stata commands ttesti, tabi, and cci. Discuss odds ratio calculation, operational definitions, number of decimal places to report and why it matters, etc.</td>
</tr>
<tr>
<td>Calculate BMI with a small number of cases to familiarize students with the construction of a “do file” using postmeno data for BMI introductory analysis</td>
</tr>
</tbody>
</table>

Required Readings for week 1: Meeting Recommendations for Multiple Healthy Lifestyle Factors: Prevalence, Clustering, and Predictors Among Adolescent, Adult, and
Senior Health Plan Members, Nicolaas P. Pronk, PhD, Louise H. Anderson, MS, A. Lauren Crain, PhD, Brian C. Martinson, PhD, Patrick J. O’Connor, MD, MPH, Nancy E. Sherwood, PhD, Robin R. Whitebird, PhD

***

**Week 2:** Review of basic Stata commands & anatomy of a basic Stata program
Friday (9/14) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

**Learning Objectives for week 2:** Upon completion of this week, students will:

Be familiar with basic Stata commands and the anatomy of a basic Stata program.

**Required Readings for week 2:**

Readings: Look over the Stata Manual (Juul)
If you decide not to purchase the Stata manual you should start reading the Stata “modules” found at ...
https://stats.idre.ucla.edu/stata/modules/
Please take a look at the first two topics on “fundamentals” (parts 1 and 2).

***

**Week 3:** Getting started coming up with ideas for a research project, analyzing a dataset, and creating tables for a report or research article
Friday (9/21) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

**Learning Objectives for week 3:** Upon completion of this week, students will:

Understand how to begin the analysis of a dataset.
Become familiar with how to create tables for a report or research article.
Become familiar with how to report on cell counts and percentages and how to understand simple bi-variate relationships (crosstabulations of nominal variables).
Understand how to recode categories and recompute crosstabulations and statistical measures of association (Stata immediate “tabi” and “cci” commands).
Become familiar with how to come up with ideas for a secondary data analysis research project and how to gain access to datasets to use for your class research project.

**Required Readings for week 3:** Reading: Chapters 1-3 in QDA

For information on creating tables with Stata, please see ...


***

**Week 4:** Preparing data for analysis and navigating IRB permissions for conducting research
Friday (9/28) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

**Learning Objectives for week 4:** Upon completion of this week, students will:

Have discussed the first homework assignment.

Understand how to collapse data from one unit of analysis to another.

Be able to use the Stata collapse function to create aggregate measures such as total household food and non-food expenditure, housing cost burden, % of total household expenditure devoted to food purchases, etc.

Understand how to navigate the IRB permission process and gain written approval to conduct secondary data analysis at Tufts University.

**Required Readings for week 4:**

No reading assignment this week but please examine the color statistics tree diagram you will receive in class and compare it to the UCLA Stata “Choosing the correct statistical test” web page found at ...

http://www.ats.ucla.edu/stat/stata/whatstat/default.htm

http://www.ats.ucla.edu/stat/stata/modules/collapse.htm

Also, please examine the example of a completed IRB exemption from review form
completed successfully by previous students.

2. Students must have completed a research ethics on-line tutorial. If you have not done so, please sign up and complete the on-line tutorial found at ...

http://www.tufts.edu/central/research/IRB/citi.htm

More details are available at ...

http://www.tufts.edu/central/research/IRB/HumanEducation.htm
**Week 5:** Creating composite and synthetic variables & how to create a data analysis plan and write a research proposal  
Friday (10/5) 1:30pm-4:30pm, Sackler room 514  
Instructor: Robert Houser

<table>
<thead>
<tr>
<th>Learning Objectives for week 5: Upon completion of this week, students will</th>
</tr>
</thead>
<tbody>
<tr>
<td>Become familiar with how to create a data analysis plan and how to write a research proposal.</td>
</tr>
<tr>
<td>Continue to become familiar with Stata data processing procedures. (continued from week 4).</td>
</tr>
<tr>
<td>Become familiar with procedures used to create composite and synthetic variables and how to write a research proposal including a data analysis plan and table shells.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Required Readings for week 5:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Merging files</td>
</tr>
<tr>
<td><a href="http://www.ats.ucla.edu/stat/Stata/faq/multmerge.htm">http://www.ats.ucla.edu/stat/Stata/faq/multmerge.htm</a></td>
</tr>
<tr>
<td>Anthropometry</td>
</tr>
<tr>
<td>Link for installing ANTHRO PLUS ...</td>
</tr>
<tr>
<td>Link for installing ENA 2011 ...</td>
</tr>
<tr>
<td><a href="http://www.nutrisurvey.net/ena2011">http://www.nutrisurvey.net/ena2011</a></td>
</tr>
<tr>
<td>Link for installing Anthro ... (WHO ANTHRO download) ...</td>
</tr>
<tr>
<td><a href="http://www.who.int/entity/childgrowth/software/WHO_Anthro_setup.exe">http://www.who.int/entity/childgrowth/software/WHO_Anthro_setup.exe</a></td>
</tr>
<tr>
<td><a href="http://www.cdc.gov/EpiInfo/">http://www.cdc.gov/EpiInfo/</a></td>
</tr>
<tr>
<td><a href="http://www.nutrisurvey.de/ena_beta/index.htm">http://www.nutrisurvey.de/ena_beta/index.htm</a></td>
</tr>
<tr>
<td>Reading: Also, please look over the section of the syllabus that shows an outline for a research</td>
</tr>
</tbody>
</table>
proposal and look over the research proposal example (handed out in class).

Ch. 4 “On the manipulation of data by computer” in QDA

Analyzing Health Equity from Household Surveys, chapter 4

http://www.who.int/pmnch/topics/economics/r_economics20080506/en/index.html

or


***

**Week 6:** Group meetings to discuss group project research plans and letters of intent

**Week 7:** Regression modeling with Stata – Part One

Friday (10/19) 1:30pm-4:30pm, Sackler room 514

Instructor: Robert Houser

**Learning Objectives for week 7:** Upon completion of this week, students will

- Become familiar with OLS regression analysis procedures with Stata
- Become familiar with Robust regression analysis procedures with Stata
- Understand how to perform dummy coding of categorical variables procedures with the `ib[#]`. function in Stata
- Understand how to interpret regression models performed with dummy coding of categorical variables with Stata
- How to decide which level of a categorical variable should be the reference category.

**Required Readings for week 7:**

http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter1/statareg1.htm


http://www.ats.ucla.edu/stat/stata/faq/rregr2.htm
http://www.ats.ucla.edu/stat/stata/output/reg_output.htm


Ch. 4 in MSA

Ch. 5, 6 and 7 in QDA

“How can I create dummy variables in Stata?”

http://www.ats.ucla.edu/stat/stata/faq/dummy.htm
**Week 7 (continued):** Regression Modeling with Stata & Introduction to Report Writing
Friday (10/19) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

**Learning Objectives for week 7:** Upon completion of this week, students will
Understand how to perform and interpret Regression diagnostics procedures in Stata

**Required Readings for week 7:**
Ch. 10 in QDA
http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter2/statareg2.htm

***

**Week 8:** Polytomous logistic regression modeling with Stata & Report Writing (Part II)
Friday (10/26) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

**Learning Objectives for week 8:** Upon completion of this week, students will
Understand how to perform and interpret Logistic regression modeling with Stata
Understand how to write a research report/article suitable for submission to an academic journal.
Understand how to write the various sections of a research article including
1 Purpose and Overview
2 Review of the literature
3 Study design and execution (Methods)
4 Analysis and Interpretation
5 Summary and Conclusions
6 References

**Required Readings for week 8:**
Ch. 13 in QDA
http://www.ats.ucla.edu/stat/stata/seminars/stata_logistic/default.htm
Logistic regression “movie” ...

http://www.ats.ucla.edu/stat/stata/seminars/stata_logistic/Movies/Stata_Binary_Logistic.html

http://www.ajcn.org/site/misc/ifa_format.xhtml

Manuscript peer review: A helpful checklist for students and novice referees ...
Vol. 23, Number 1. Advances in Physiology Education, June 2000 ...


Please make sure you have submitted your IRB paperwork by this week.
**Week 9:** Complex survey data analysis & analyzing data from a cluster randomized trial  
Friday (11/2) 1:30pm-4:30pm, Sackler room 514  
Instructor: Robert Houser

**Learning Objectives for week 9:** Upon completion of this week, students will understand how to perform Complex Survey Data Analysis with Stata including the `svyset` and `svy:` commands.

**Required Readings for week 9:**

Ch. 9 in QDA  
[http://www.ats.ucla.edu/stat/stata/seminars/svy_stata_intro/default.htm](http://www.ats.ucla.edu/stat/stata/seminars/svy_stata_intro/default.htm)

Movies at ...
[http://www.ats.ucla.edu/stat/stata/seminars/svy_stata_8/default.htm](http://www.ats.ucla.edu/stat/stata/seminars/svy_stata_8/default.htm)

Example with DHS data ...
[http://www.cpc.unc.edu/research/tools/data_analysis/statatutorial/sample_surveys/svy_commands](http://www.cpc.unc.edu/research/tools/data_analysis/statatutorial/sample_surveys/svy_commands)

***

**Week 10:** Group meetings to discuss group project research plans and letters of intent

Analysis of Variance (ANOVA) and Covariance (ANCOVA) Online (optional)  
Instructor: Robert Houser

Upon completion of this week, students will be able to perform and interpret ANOVA and ANCOVA models with Stata

**Week 11:** Introduction to Scale construction

Friday (11/16) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

<table>
<thead>
<tr>
<th>Learning Objectives for week 11: Upon completion of this week, students will</th>
</tr>
</thead>
<tbody>
<tr>
<td>Become familiar with Introductory Scale construction concepts and procedures including Chronbach alpha and Kuder-Richardson 20.</td>
</tr>
<tr>
<td>Be able to perform Chronbach alpha and Kuder-Richardson 20 analyses with Stata.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Required Readings for week 11:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch. 11 in QDA</td>
</tr>
<tr>
<td>Scale Construction (Very brief overview) ...</td>
</tr>
<tr>
<td><em>Scales and Measures (optional Kindle edition)</em></td>
</tr>
<tr>
<td><a href="http://www.statisticalassociates.com/scalesandmeasures.htm">http://www.statisticalassociates.com/scalesandmeasures.htm</a></td>
</tr>
<tr>
<td>Validating scales and indexes by Bland et al.</td>
</tr>
<tr>
<td>BMJ 2002; 324 doi: 10.1136/bmj.324.7337.606 (Published 9 March 2002)</td>
</tr>
<tr>
<td>Cite this as: BMJ 2002;324:606.</td>
</tr>
<tr>
<td><a href="http://www.bmj.com.ezproxy.library.tufts.edu/content/324/7337/606.1">http://www.bmj.com.ezproxy.library.tufts.edu/content/324/7337/606.1</a></td>
</tr>
</tbody>
</table>
**Week 12:** Data Reduction Techniques for multivariate data – part I  
Friday (11/30) 1:30pm-4:30pm, Sackler room 514  
Instructor: Robert Houser

<table>
<thead>
<tr>
<th>Learning Objectives for week 12</th>
<th>Upon completion of this week, students will become familiar with Principle Component Analysis procedures and Stata.</th>
</tr>
</thead>
</table>

| Required Readings for week 12 | Ch. 7 in MSA  
http://dss.princeton.edu/training/Factor.pdf  
[Constructing socio-economic status indices: how to use principal components analysis](http://heapol.oxfordjournals.org/content/21/6/459.short)  
[SEEMA VYAS AND LILANI KUMARANAYAKE HIV Tools Research Group, Health Policy Unit, Department of Public and Policy, LSHTM ...](http://heapol.oxfordjournals.org/content/21/6/459.short) |

<table>
<thead>
<tr>
<th>Learning Objectives for week 13</th>
<th>Upon completion of this week, students will become familiar with how to perform and interpret k means cluster analysis with Stata.</th>
</tr>
</thead>
</table>

| Required Readings for week 13 | Ch. 8 in MSA |

***

**Week 13: Panel Data Analysis**

Friday (12/7) 1:30pm-4:30pm, Sackler room 514  
Instructor: Robert Houser

| Learning Objectives for week 13: Analysis of panel data | Be able to describe the advantages and disadvantages of panel data. Define panel data, also called longitudinal data or cross-sectional time series data, which are data where... |
multiple cases (people, firms, countries etc) were observed at two or more time periods. Be able to “reshape” data from wide form to long form in Stata. Be able to use the `xtset` command to tell Stata that your dataset is panel data. Define fixed effects regression. Perform a fixed effects regression with Stata with the `xtreg` command. Be able to choose between a fixed effects and random effects regression.

http://dss.princeton.edu/online_help/stats_packages/stata/panel.htm
https://www3.nd.edu/~rwilliam/stats3/index.html

Analysis of poverty over time


Multilevel modeling

Recognize a research problem requiring a multilevel modeling approach. Describe the technical and substantive advantages of multilevel models. Explain the basic principles of multilevel modeling using graphical, verbal, and statistical language for a range of multilevel models. Develop a variety of models that enable quantitative assessment of contextual effects. Apply multilevel models to a research problem according to a well articulated research strategy.

Optional: Multilevel modeling with Stata ...

http://www.esourceresearch.org/Portals/0/Uploads/Documents/Public/Subu_FullChapter.pdf
https://www.youtube.com/embed/SbwApki_Bnl
Appendix 1 from Tufts Friedman Policies and Procedures Manual

Authorship Guideline Relating to Manuscripts Involving Friedman School of Nutrition Science and Policy Students

These guidelines apply to all written submissions involving students of the Friedman School when using their school affiliation in the context of publication. Their intent is to protect students’ interests in the context of research-based publications.

The school encourages student participation in research and related authorship of publishable papers, be it alone or in collaboration with faculty or other students at the school. These guidelines identify minimum requirements for authorship as well as for acknowledgments. Some journals have additional requirements. Every author must comply with the authorship requirements of the journal to which a manuscript is submitted provided that he/she also meets the minimum requirements of the Friedman School.


***

All persons designated as authors should meet the criteria for authorship as follows:

1. Authorship credit should be based only on: a) substantial contributions to the conception and design, data acquisition, or data analysis and interpretation (participation in a doctoral thesis committee does not in itself guarantee co-authorship of resultant publications); b) contributing new reagents, new animal models (e.g. transgenic animal), or other novel tools for successful implementation of the study design, c) drafting or revising an article critically for important intellectual content; and d) final approval of the version to be published. Conditions a, and c and d must all be met; condition b is specific to only certain kinds of biological research.

2. When doctoral dissertation work is converted into publishable articles or other paper the doctoral student concerned is typically first author. The order of other authors on the by-line should be a joint decision among co-authors, guided by the amount of time invested in bringing the publication to fruition. Relative contributions to the research, writing and analysis process, and hence likely order of co-authorship, should as far as possible be agreed at that outset of paper preparation (or as early in the process as possible), rather than at the point of submission (see attached references for further guidance on the ethics of deciding authorship order).

Authors should be prepared to explain the order in which authors are listed. If issues cannot be resolved jointly among authors, potential co-authors concerned about perceived unethical practice can bring the issue to the attention of the Academic Dean. If unresolved at that level, a case may be brought to the Grievance Committee, and ultimately to the University’s Grievance Committee.

3. An author may list more than one contribution to a publication, and more than one author may have contributed to the same aspect of the work. Authors should be prepared to indicate their specific contributions to the published work. The corresponding author must obtain permission from all authors for any change in the order of authorship prior to publication.
Students should pay close attention to the order of authorship at the time of submission of drafts to a journal, and also when signing any copyright agreement.

4. Other members of the group who do not meet authorship criteria should be listed, with their permission, in the Acknowledgments or in an appendix. The acquisition of funding, data collection, supervision of the research group, or participation as a support service center member, do not justify authorship.

5. Authorship of multi-center trials should be attributed to a group. All individuals in a group who are named as authors should meet the criteria for authorship. Members who do not meet authorship criteria should be listed, with their permission, in the Acknowledgments or Appendix sections.

6. List all other contributors who do not meet the criteria for authorship, such as individuals who provided purely technical or writing assistance, a laboratory director who provided only general support, and/or material support in the Acknowledgments section.

7. Groups of people who have contributed to the paper but whose contributions do not justify authorship may be listed in the Acknowledgments section under a heading such as “participating investigators.” Their contributions can be described as “scientific advisors,” “critically reviewed the study proposal,” “collected data,” or “provided and cared for volunteers and/or animals.”

8. Because readers may infer endorsement of data by those acknowledged, all persons must give written permission to be acknowledged.

9. Conflict of interest exists if there is a direct or indirect financial relationship with industry through employment, consultancies, stock ownership and honoraria, either directly with the author, or
through the author’s family. However, conflicts can also occur for other reasons, such as antagonistic personal relationships among researchers or students, or academic competition. The perception of conflict of interest should be avoided. All authors should be prepared to defend any relationship that could be construed or perceived as a conflict of interest. For further details, see Tufts University’s Policy on Conflict of Interest as published in Tufts’ Business Conduct Policy (2000) booklet or seek advice and counsel for Tufts’ Office of Associate Provost for Research.

Recommended Reference Materials on Ethics of Authorship.