Nutrition Data Analysis (NUTR311)
Updated 9/2/2016

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

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

Teaching Assistant(s)

Reem AlSukait, email: Reem.AlSukait@tufts.edu
Office Hours: By appointment

Corey O’Hara, email: Corey.O_Hara@tufts.edu
Office Hours: By appointment

TA duties and responsibilities

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

Corey is responsible for tutoring students and helping students plan, analyze and write course project research papers.

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, descriptive statistics, univariate analyses, regression analysis of continuous, binary and categorical outcomes, ANOVA & ANCOVA, analysis of clustered data including cluster randomized trials, panel data analysis & introduction to multilevel modeling, factor analysis; and the construction of scales and factor scores. Students pose a research question, identify appropriate statistical techniques for answering the research question, perform the analyses and report on the results in an article suitable for publication in an academic journal. Advanced Stata programming will be taught in weekly hands on lab sessions.

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
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, 9/23, due on Sunday, 10/2

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 10/9).

2nd graded homework assigned on Friday, 10/7, due on Sunday, 10/23
Table shells and analysis plan due on Sunday, 10/30
IRB paperwork must be submitted to the IRB office by Wednesday, 11/10
3rd graded homework assigned on Friday, 11/11, due on Sunday 11/20
Results and discussion section due on Sunday, 11/27
Complete research article including references and abstract due on Sunday, 12/11

Assignments are due before midnight on due dates. Homework assignments should be turned in via Trunk 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 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 7 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>
</tbody>
</table>
Average of 8 required online quizzes (For lessons 1, 2, 3, 4, 5, 6, 11 & 12) 18%

LOI including description of dataset, research question(s)/hypotheses and brief literature review 10%

Table shells and analysis plan due on Sunday 10/26 5%

First draft Results and discussion section of paper 10%

Complete research article 30%

**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 assignments (papers, drafts and homeworks).

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>
</tr>
</thead>
<tbody>
<tr>
<td>Juul</td>
<td>An Introduction to Stata for</td>
<td>3rd</td>
<td>978-1597180771</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td>---------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Kachigan</td>
<td>Multivariate Statistical Analysis: A Conceptual Introduction</td>
<td>2nd</td>
<td>978-0942154917</td>
</tr>
<tr>
<td>Treiman</td>
<td>Quantitative Data Analysis</td>
<td>2009</td>
<td>978-0470380031</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: http://library.tufts.edu/record=b2431675~S1

Software: STATA, ENA, and WHO anthropometry program for anthropometric calculation, Microsoft Access, Powerpoint. We will use STATA IC version 13 at the computer lab. If you purchase Stata this year, you should purchase Stata IC version 14.

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 14 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 14 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.

Miscellaneous Supplies: You should purchase a solar powered calculator that has basic statistical and logarithmic functions. You can purchase one for around 15 – 20 dollars at a Best Buy or Staples) or install a free scientific calculator app on your phone or tablet computer such as RealCalc for Android. Please purchase a USB drive (ideally you should look for one that has a capacity of 1 GB or higher). It is extremely important for you to make regular backups of files stored on your USB drive. You can copy the contents of your USB drive to a CDROM, DVD or
computer hard drive. You might want to purchase a rewritable CDROM or DVD disk. It is also extremely important that you do not store identifiable confidential data on your USB drive, CDROM disks or DVD disks. (in case they are lost or stolen). Yet another strategy for making backup copies is to email files to your own email address so long as they don’t contain identifying information. You can “zip” the files into an archive and email the archive file to yourself for safe keeping.

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). 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. http://uss.tufts.edu/studentAffairs/documents/TuftsStudentHandbook.pdf

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 enough to complete all of the lessons (both required and optional). Please make sure you use a Tufts supported browser when accessing Trunk. 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 9/11

Lesson 2: Review of Multiple Regression Assumptions
Online quiz 2 by 9/18

Lesson 3: Review of Assessing distributional normality & transforming variables
Online quiz 3 by 9/18

Lesson 4: Building statistical models, dummy variables & interaction terms
Online quiz 4 by 9/25

Lesson 5: Binary logistic regression analysis and polytomous logistic regression analysis
Online quiz 5 by 10/16

Lesson 6: Presenting statistical results in tables, figures and text
Online quiz 6 by 11/6

Lesson 11: Analysis of complex surveys
Online quiz 11 by 11/13

Lesson 12: Study design and selection of appropriate statistical methods
Online quiz 12 by 12/4

Optional Lessons & optional quizzes

Optional Lesson 7: One way 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 9/9</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</td>
<td>Robert Houser</td>
</tr>
<tr>
<td>Date</td>
<td>Time</td>
<td>Location</td>
<td>Activity</td>
<td>Due Date Details</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>-------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>9/16</td>
<td>2</td>
<td>Sackler room 514</td>
<td>Review of basic Stata commands &amp; anatomy of a basic Stata program</td>
<td>In-class Stata review exercise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9/16</td>
<td></td>
<td>Online lesson 1 this week Online Quiz 1 by 9/11</td>
</tr>
<tr>
<td>9/23</td>
<td>3</td>
<td>Sackler room 514</td>
<td>Getting started coming up with ideas for a research project, analyzing a</td>
<td>HW1 assigned this week and due on 9/18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9/23</td>
<td>dataset, and creating tables for a report or research article</td>
<td>Online lesson 4 Online quiz 4 by 9/25</td>
</tr>
<tr>
<td>9/30</td>
<td>4</td>
<td>Sackler room 514</td>
<td>Preparing data for analysis and navigating IRB permissions for conducting</td>
<td>HW1 due on 10/2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>9/30</td>
<td>research</td>
<td></td>
</tr>
<tr>
<td>10/7</td>
<td>5</td>
<td>Sackler room 514</td>
<td>Creating composite and synthetic variables &amp; how to create a data analysis</td>
<td>LOI including information about the data set and research questions/hypotheses,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10/7</td>
<td>plan and how to write a research proposal</td>
<td>methods and brief literature review for the research project you plan to conduct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(due by 10/9)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Homework 2 assigned on 10/7 (due on 10/23)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Group meetings</td>
<td>--</td>
<td>No class on Friday 10/14 Group meetings this week with Dr. Houser Homework 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>this week (10/10-</td>
<td></td>
<td>due on 10/23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10/14) to discuss</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Sackler room 514</td>
<td>Regression modeling with Stata &amp; Introduction to Report Writing</td>
<td>Online lesson 5 this week Online quiz 5 by 10/23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10/21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/28</td>
<td>8</td>
<td>Sackler room 514</td>
<td>Polytomous logistic regression modeling with Stata &amp; Report Writing (Part II)</td>
<td>Table shells and analysis plan due on Sunday 10/30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10/28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/4</td>
<td>9</td>
<td>Sackler room 514</td>
<td>Complex survey data analysis &amp; Analyzing data from a cluster randomized</td>
<td>Online lesson 6 this week Online quiz 6 by 11/6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11/4</td>
<td>trial</td>
<td></td>
</tr>
</tbody>
</table>

Robert Houser
<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Topic</th>
<th>Assignments</th>
<th>Instructor</th>
</tr>
</thead>
<tbody>
<tr>
<td>11/11</td>
<td>10</td>
<td>ANOVA &amp; ANCOVA models</td>
<td>Draft of literature review &amp; methods section due 11/13</td>
<td>Robert Houser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Homework 3 assigned on 11/11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>IRB paperwork submission to IRB office due 11/10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Online lesson 11 this week Online quiz 11 by 11/13</td>
<td></td>
</tr>
<tr>
<td>11/18</td>
<td>11</td>
<td>Introduction to scale construction</td>
<td>Homework 3 due on 11/22</td>
<td>Robert Houser</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Chronbach alpha &amp; Principle component analysis)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12/2</td>
<td>12</td>
<td>Data Reduction Techniques for multivariate data</td>
<td>Draft of results &amp; discussion due by 11/27</td>
<td>Robert Houser</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Thanksgiving break</td>
<td>Online lesson 12 this week Online quiz 12 by 12/4</td>
<td></td>
</tr>
<tr>
<td>12/9</td>
<td>13</td>
<td>Panel data analysis &amp; Introduction to Multilevel modeling</td>
<td>Research Paper Due on 12/11</td>
<td>Robert Houser</td>
</tr>
</tbody>
</table>

This schedule is subject to modifications at the discretion of the instructor.
# Academic Calendar Fall 2016

<table>
<thead>
<tr>
<th>AUGUST 2016</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>30-31</td>
<td>Tues.-Wed.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SEPTEMBER 2016</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Monday</td>
</tr>
<tr>
<td>6</td>
<td>Tuesday</td>
</tr>
<tr>
<td>20</td>
<td>Tuesday</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OCTOBER 2016</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Monday</td>
</tr>
<tr>
<td>11</td>
<td>Tuesday</td>
</tr>
<tr>
<td>31</td>
<td>Monday</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NOVEMBER 2016</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>Tuesday</td>
</tr>
<tr>
<td>11</td>
<td>Friday</td>
</tr>
<tr>
<td>23-25</td>
<td>Wed.-Fri.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DECEMBER 2016</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Monday</td>
</tr>
<tr>
<td>12</td>
<td>Monday</td>
</tr>
<tr>
<td>13</td>
<td>Tuesday</td>
</tr>
<tr>
<td>14</td>
<td>Wednesday</td>
</tr>
<tr>
<td>15</td>
<td>Thursday</td>
</tr>
<tr>
<td>22</td>
<td>Thursday</td>
</tr>
</tbody>
</table>
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 (9/9) 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.) 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.) Go over the assigned homework exercise and homework assignment Discuss Stata commands ttesti, tabi, and cci. Discuss odds ratio calculation, operational definitions, number of decimal places to report and why it matters, etc. Calculate bmi with a small number of cases to familiarize students with the construction of a “do file” Use Delaware BRFSS 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 (9/16) 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 watching the online Stata “movies” found at ...

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

***

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 (9/23) 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 (9/30) 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 (10/7) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

Learning Objectives for week 5: Upon completion of this week, students will

Become familiar with how to create a data analysis plan and how to write a research proposal.

Continue to become familiar with Stata data processing procedures. (continued from week 4).

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.

Required Readings for week 5:
Merging files
http://www.ats.ucla.edu/stat/Stata/faq/multmerge.htm

Anthropometry
Link for installing ANTHRO PLUS ...
http://www.who.int/growthref/tools/en/

Link for installing ENA 2011 ...
http://www.nutrisurvey.net/ena2011

Link for installing Anthro ... (WHO ANTHRO download) ...
http://www.who.int/entity/childgrowth/software/WHO_Anthro_setup.exe
http://www.who.int/childgrowth/software/en/
http://www.who.int/entity/childgrowth/software/anthro_pc_manual.pdf
http://www.cdc.gov/EpiInfo/
http://www.nutrisurvey.de/ena_beta/index.htm

Reading:
Also, please look over the section of the syllabus that shows an outline for a research
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 (10/21) 1:30pm-4:30pm, Sackler room 514

Instructor: Robert Houser

Learning Objectives for week 6: 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 6:

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:** Regression Modeling with Stata & Introduction to Report Writing

Friday (10/21) 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


***

**Week 8:** Polytomous logistic regression modeling with Stata & Report Writing (Part II)

Friday (10/28) 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](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 (11/4) 1:30pm-4:30pm, Sackler room 514  
Instructor: Robert Houser

<table>
<thead>
<tr>
<th>Learning Objectives for week 9:</th>
<th>Upon completion of this week, students will</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Understand how to perform Complex Survey Data Analysis with Stata including the svyset and svy: commands.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Required Readings for week 9:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ch. 9 in QDA</td>
</tr>
<tr>
<td><a href="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</a></td>
</tr>
<tr>
<td>Movies at ...</td>
</tr>
<tr>
<td><a href="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</a></td>
</tr>
<tr>
<td>Example with DHS data ...</td>
</tr>
<tr>
<td><a href="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</a></td>
</tr>
</tbody>
</table>

***

**Week 10:** Analysis of Variance (ANOVA) and Covariance (ANCOVA)  
Friday (11/8) 1:30pm-4:30pm, Sackler room 514  
Instructor: Robert Houser

<table>
<thead>
<tr>
<th>Learning Objectives for week 13:</th>
<th>Upon completion of this week, students will</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Be able to perform and interpret ANOVA and ANCOVA models with Stata</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Required Readings for week 13:</th>
</tr>
</thead>
</table>
Week 11: Introduction to Scale construction

Friday (11/18) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

Learning Objectives for week 11: Upon completion of this week, students will

Become familiar with Introductory Scale construction concepts and procedures including Chronbach alpha and Kuder-Richardson 20.

Be able to perform Chronbach alpha and Kuder-Richardson 20 analyses with Stata.

Required Readings for week 11:

Ch. 11 in QDA

Scale Construction (Very brief overview) ...

http://www.nd.edu/~rwilliam/stats2/l23.pdf

Scales and Measures (optional Kindle edition)

http://www.statisticalassociates.com/scalesandmeasures.htm

Validating scales and indexes by Bland et al.

BMJ 2002; 324 doi: 10.1136/bmj.324.7337.606 (Published 9 March 2002)

Cite this as: BMJ 2002;324:606.

http://www.bmj.com.ezproxy.library.tufts.edu/content/324/7337/606.1
**Week 12:** Data Reduction Techniques for multivariate data – part I
Friday (12/2) 1:30pm-4:30pm, Sackler room 514
Instructor: Robert Houser

**Learning Objectives for week 12:** Upon completion of this week, students will
Become familiar with Principle Component Analysis procedures and Stata.

**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
SEEMA VYAS AND LILANI KUMARANAYAKE
HIVTools Research Group, Health Policy Unit, Department of Public
http://heapol.oxfordjournals.org/content/21/6/459.short

**Learning Objectives for week 13:** Upon completion of this week, students will
Become familiar with how to perform and interpret k means cluster analysis with Stata.

**Required Readings for week 13:**
Ch. 8 in MSA

***

**Week 13:** Panel Data Analysis
Friday (12/11) 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_BnI