UEP 294-25: Advanced Quantitative Reasoning
Spring 2016

Prof. Mary Davis
Urban and Environmental Policy and Planning
Office location: 72 Professor’s Row
mary.davis@tufts.edu; 617-627-4719

Class Meetings
Thursday 10:30-11:45am (Eaton Rm 208)
There is no additional lab requirement for this course.

Office Hours
Tuesdays 12-2pm (office, phone, email, or Skype: medavis129)

Course Description
This half-credit module will provide a broad overview of advanced data analysis tools relevant to policy and planning practitioners. Although completion of the introductory quantitative course (UEP 254 or equivalent) is a prerequisite for this class, no additional statistics background is required. The course is designed to be flexible to student interests, and adaptations to the syllabus may be required to cover the range of student data needs. Most of the first half of the semester will cover topics related to linear regression, paying particular attention to realistic data scenarios where the standard assumptions do not apply. We will explore diagnostic tools to describe and graph data, and assess the suitability of various modeling approaches and functional forms. After reviewing the fundamentals of linear regression, we will explore advanced methods related to limited dependent variables and panel data. Final advanced topics will be selected by student consensus, and may include spatial and temporal data analysis, study design, factor analysis, nonparametric methods, among many other possible topics.

Prerequisites
This course takes up where UEP 254 ends, so it is assumed that all students in the class have the basic knowledge covered in an introductory college-level statistics class. This class focuses on the applied nature of statistics, and students will not be tested on the underlying math theory.

Course Materials
Introductory Econometrics, 5th Edition 2014 by Jeffrey Wooldridge is the required textbook for the course (noted as ‘W’ in the required reading list). The international edition is available on Amazon for about $25, and the previous 4th edition is similar enough to the 5th that it should also work as a substitute. I also encourage you to review your introductory statistics textbook as needed to refresh your memory of the basics.

STATA software. This class will use STATA 14 as the primary statistical software. However, students interested in continuing to use and expand their knowledge of SPSS will be able to do so throughout the course and in the assignments. All STATA commands will be reviewed in class and supplemental handouts of the commands will be provided. Although STATA is available in the Eaton and GIS labs, I recommend purchasing the software for your own computer, which is available for $125 (annual license) or $200 (perpetual license).
Grading and Assessment Policy
Grades will be based on bi-weekly problem sets, class participation, and a data analysis project:

- Problem Sets – 42% (7% each)
- Class Participation – 10%
- Project – 54% (10% presentation, 22% guided questions, 22% final paper)

Problem Sets
Regular bi-weekly problem sets will provide students the opportunity to practice the concepts on real-world problems and datasets. The homework will be graded on a simplified zero to seven scale as noted below, with each assignment worth 7% of your final grade. They are explicitly designed as a problem-solving exercise, and it is more important that you try to creatively solve quantitative problems than come up with cookie cutter answers. I strongly encourage you to work on problem sets together.

<table>
<thead>
<tr>
<th>HW Score</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero</td>
<td>Completely missing response</td>
</tr>
<tr>
<td>One to two</td>
<td>Majority of questions (&gt;70%) not fully answered and/or incorrect</td>
</tr>
<tr>
<td>Three to four</td>
<td>Around half of questions fully answered and/or correct</td>
</tr>
<tr>
<td>Five to Six</td>
<td>Most of questions (&gt;70%) fully answered and/or correct</td>
</tr>
<tr>
<td>Seven</td>
<td>All questions fully answered and correct</td>
</tr>
</tbody>
</table>

Class participation
The teaching format will be more discussion-based than a typical statistics course and it is essential that you come to class prepared and willing/able to participate in the discussions. Please let me know in advance if you will not be able to attend class, as attendance and participation will represent approximately 10% of your final grade.

Project
Students will be required to complete a data analysis project during the semester that will hopefully coincide with a potential or ongoing thesis topic. Student projects will represent a culmination of work throughout the semester and the topic will gradually develop through weekly guided questions as outlined in the syllabus, each worth 2% of your final grade. As part of the project, students will also be required to present their work to the class at the end of the semester. The presentations are designed to provide students the opportunity to learn to convey and simplify quantitative concepts to a broad audience.

Students with Disabilities
Tufts University values the diversity of our students, staff, and faculty; recognizing the important contribution each student makes to our unique community. Students with disabilities are assured that the Student Accessibility Services (SAS) office will work with each student individually to create access to all aspects of student life. Tufts is committed to providing equal access and support to all qualified students through the provision of reasonable accommodations so that each student may fully participate in the Tufts experience. If you have a disability that requires reasonable accommodations, please contact the Student Accessibility Services office at accessibility@tufts.edu or 617-627-4539 to make an appointment with an SAS representative to determine appropriate accommodations. Please be aware that accommodations cannot be enacted retroactively, making timeliness a critical aspect for their provision.
## Course Outline

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Required Readings</th>
<th>Assignments Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 21</td>
<td>Introduction</td>
<td>W 1.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chingos 2015</td>
<td></td>
</tr>
<tr>
<td>January 28</td>
<td>Linear Regression</td>
<td>W 2.1-2.3; 3.1-3.2; 4.2; 4.5; 6.3</td>
<td>Project Q #1</td>
</tr>
<tr>
<td>February 4</td>
<td>Linear Regression</td>
<td>W 2.5-2.6; 3.3-3.5; 8.1-8.3</td>
<td>Project Q #2</td>
</tr>
<tr>
<td>February 11</td>
<td>Understanding Variables</td>
<td>W 2.4; 6.1-6.2 and summary; 7.1-7.4</td>
<td>HW #1; Project Q #3</td>
</tr>
<tr>
<td>February 18</td>
<td>NO CLASS (Monday schedule)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>February 25</td>
<td>Understanding Variables</td>
<td>Bell et al. 2008</td>
<td>Project Q #4</td>
</tr>
<tr>
<td>March 3</td>
<td>Regression Diagnostics</td>
<td>STATA handout</td>
<td>HW #2; Project Q #5</td>
</tr>
<tr>
<td>March 10</td>
<td>Discrete Dependent Variables</td>
<td>W 7.5; 7.7; 17.1 Logistic handout</td>
<td>HW #3; Project Q #6</td>
</tr>
<tr>
<td>March 17</td>
<td>Discrete Dependent Variables</td>
<td>Besser and Dannenberg 2005</td>
<td>Project Q #7</td>
</tr>
<tr>
<td>March 24</td>
<td>NO CLASS (Spring Break)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>March 31</td>
<td>Panel Data</td>
<td>W 13 (all)</td>
<td>HW #4; Project Q #8</td>
</tr>
<tr>
<td>April 7</td>
<td>Panel Data</td>
<td>W 14 (all) White et al. 2013</td>
<td>Project Q #9</td>
</tr>
<tr>
<td>April 14</td>
<td>Advanced Topics</td>
<td>TBD</td>
<td>HW #5; Project Q #10</td>
</tr>
<tr>
<td>April 21</td>
<td>Advanced Topics</td>
<td>TBD</td>
<td>Project Q #11</td>
</tr>
<tr>
<td>April 28</td>
<td>Advanced Topics</td>
<td>TBD</td>
<td></td>
</tr>
<tr>
<td>Final Exam</td>
<td>In-class project presentations</td>
<td></td>
<td>HW #6; Final project</td>
</tr>
<tr>
<td>Period TBD</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Project Progression (responses turned in weekly)
1. Identify a broad research area and list of potential data sources
   a. Assignment to turn in: write this information in a short paragraph and provide a list of potential data sources
2. Narrow your broad research area to a specific hypothesis and identify datasets that might help you address that question
   a. Assignment to turn in: write this information in a short paragraph and attach potential STATA/SPSS dataset(s)
3. Do a literature search on your research question, specifically looking for how other quantitative research has developed and tested similar hypotheses
   a. Assignment to turn in: produce an annotated bibliography with detailed information on the statistical approach and data sources
4. Finalize your research question and data source, and propose a statistical modeling approach to answer that question
   a. Assignment to turn in: provide a clear paragraph describing your question, the data you have available to answer that question, and how you plan to quantitatively answer that question
5. Explore your dataset using the various summary and graphic tools in STATA/SPSS, and identify potential data issues such as poorly specified or missing variables
   a. Assignment to turn in: Provide a paragraph description of your data, table of summary statistics, and relevant graphics; describe any potential data issues or problems you identified as a result of this exercise
6. Continue to explore data analysis tools appropriate to your question and develop a final analytical approach
   a. Assignment to turn in: Write a methods section composed of all the information you have compiled to date that includes introduction to topic and hypothesis, description of data, and proposed modeling approach
7. Refine your methods section based on feedback from instructor and class discussions
   a. Assignment to turn in: Revised methods section
8. Analyze your data using the previously identified modeling approach and summarize your results
   a. Assignment to turn in: Write a paragraph describing the results of your analysis, and attach any relevant tables and figures (note: a table of results is required); based on these results, describe how you might re-tool the analysis to better answer your research question
9. Develop preliminary conclusions regarding your hypotheses
   a. Assignment to turn in: Write a paragraph describing what these results mean in layperson terms, and how they are significant to your broader research area
10. Develop rough first draft of your final project
    a. Assignment to turn in: write-up that includes introduction to the topic and hypothesis, description of data, and proposed modeling approach, results of the analysis, implications/conclusions drawn from those results
11. Develop less rough second draft of your final project
    a. Assignment to turn in: write-up that includes introduction to topic and hypothesis, proposed modeling approach, description of data, results of the analysis, implications/conclusions drawn from those results
**Topic Descriptions**

The following descriptions provide a cursory overview of the general concepts that will be covered each week. Additional content will be updated as needed. All topics will include instruction for using STATA statistical software as needed.

**Introduction**
- Understanding the goals of empirical models, prediction, and analysis using statistics
- Identifying the importance of appropriate model specification
- Working with various types of data – continuous vs categorical, cross-sectional vs time series, etc.

Required Readings:
- Woolridge textbook section 1.3

**Linear Regression**
- Reviewing the basics of standard linear regression models
- Understanding the assumptions: what does it mean to be BLUE?
- Interpreting regression results

Required Readings:
- Woolridge textbook sections 2.1-2.3; 2.5-2.6; 3.1-3.5; 4.2; 4.5; 6.3; 8.1-8.3

**Understanding Variables**
- Organizing, summarizing, and displaying data
- Understanding and interpreting transformations – log, square, elasticity, etc.
- Using and interpreting interaction terms and dummy variables

Required Readings:
- Woolridge textbook sections 2.4; 6.1-6.2 and summary; 7.1-7.4

**Regression Diagnostics**
- Identifying and controlling for common statistical issues, including multicollinearity, heteroscedasticity, omitted variable bias, and outliers
- Finding data ‘issues’ – missing variables, improperly coded, etc.

Required Readings:
- STATA handout from User Manual
Discrete Dependent Variables

- What to do when your dependent variable isn’t continuous
- Predicting dummy and categorical variables common to survey work
- Building, understanding, and interpreting common modeling approaches – probit, logit, logistic, and linear probability models

Required Readings:
- Woolridge textbook sections 7.5; 7.7; 17.1
- Logistic Regression handout

Panel Data

- What to do when your data are grouped or clustered
- Understanding and interpreting panel data
- Understanding options for addressing data clustering – fixed vs random effects and multi-level modeling
- Specifying, interpreting, and trouble-shooting panel data models

Required Readings:
- Woolridge textbook chapters 13 and 14

Potential Advanced Topics (final list will be determined by student data interests)

Time Series: What to do when your data are sequenced and correlated over time?
- Understanding time series data
- Diagnosing correlation over time
- Understanding options for addressing temporal correlations
- Specifying, interpreting, and trouble-shooting time series models

Spatial Data: What to do when your data are correlated over geographic space?
- Understanding spatial data
- Diagnosing spatial correlation
- Understanding options for addressing spatial correlations and software needs
- Specifying, interpreting, and trouble-shooting spatial models

Study Design: How to design primary data collection efforts to answer the right questions?
- Understanding study design from a statistical perspective
- Designing meaningful (and analyzable) quantitative surveys and questions
- Collecting the right amount of data (sample size calculations)
- Exploring sources of bias in poorly designed studies and data collection efforts