Syllabus DRAFT
UEP 294-25: Intermediate 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 (Room TBD)
There is no additional lab requirement for this course.

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

Course Description
This half-credit course 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 math or 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. The first half of the semester will cover topics related to linear regression, paying particular attention to realistic data scenarios where the assumptions of standard linear regression models do not apply. We will explore diagnostic tools for describing and graphing data, and for assessing the suitability of various modeling approaches. We will also build upon UEP 254 by discussing how to interpret regression coefficients, and to understand the effect of different functional forms. After reviewing the fundamentals of linear regression, we will adapt the second half of the semester to explore advanced methods related to time series, panel data, and discrete dependent variable analysis. These specific methods were chosen because they deal with data issues and research questions most common to policy analysis, such as trends over time, data clustering across states/municipalities, and survey questions with discrete outcomes such as yes/no or categorical responses. Other potential topics will vary based on student interests, and may include GIS, study design, factor analysis, nonparametric methods, environmental statistics, and simultaneous equation models.

Prerequisites
The course material 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 does not explore the underlying math theory in detail. As such, we will only skim the more “mathy” topics from the readings, and all essential math concepts will be covered as needed in class.
Course Materials

*Introductory Econometrics, 5th Edition 2014 (International edition!!),* by Jeffrey Wooldridge, is the required textbook for the course and you can purchase a copy online. We will use the international edition, which is available for purchase on Amazon for about $25.

**STATA software.** This class will use **STATA** 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 review sheets of the commands will be provided. Although STATA is available in the Eaton and GIS labs, I recommend purchasing the software and basic user manuals, which are available at a reduced rate through the university.

**Grading Policy**

Grades will be based on weekly problem sets, a project, and class participation:

- **Problem Sets – 36% (6% each)**
- **Class Participation – 20%**
- **Project – 44% (10% presentation, 11% guided questions, 23% final paper)**

**Problem Sets**

Regular bi-weekly problem sets will provide students the opportunity to practice the concepts covered in class on real-world problems and datasets. Problem sets will be graded on a simplified zero to six scale, with each worth 6% of your final grade. The homework is explicitly designed as a problem-solving exercise, and it is more important that you try to creatively solve quantitative problems than to come up with cookie cutter yes/no answers. I strongly encourage you to work on problem sets together.

**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 will represent approximately 10% of your final grade. The remaining 10% of your participation score will be determined by your contribution to the class discussions.

**Project**

Students will be required to pursue a mini research project during the semester (hopefully to coincide with a potential or ongoing thesis topic) and to submit a final project. Student projects will represent a culmination of work throughout the semester and the topic will gradually develop through weekly guided questions as outlined below. Each response to the guided project question will be worth 1% of your final grade. As part of the project, each student will also be required to present their work to the class in the form of a Power Point presentation 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.
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
## Course Outline

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Readings*</th>
<th>Assignments Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 21</td>
<td>Reviewing the Basics–Statistics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>January 28</td>
<td>Linear Regression</td>
<td></td>
<td>Project Q #1</td>
</tr>
<tr>
<td>February 4</td>
<td>Linear Regression contd.</td>
<td></td>
<td>Project Q #2</td>
</tr>
<tr>
<td>February 11</td>
<td>Understanding Variables</td>
<td></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>Regression Diagnostics</td>
<td></td>
<td>HW #2; Project Q #4</td>
</tr>
<tr>
<td>March 3</td>
<td>Regression Diagnostics contd.</td>
<td></td>
<td>Project Q #5</td>
</tr>
<tr>
<td>March 10</td>
<td><em>Advanced Topics TBD</em></td>
<td></td>
<td>HW #3; Project Q #6</td>
</tr>
<tr>
<td>March 17</td>
<td><em>Advanced Topics TBD</em></td>
<td></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><em>Advanced Topics TBD</em></td>
<td></td>
<td>HW #4; Project Q #8</td>
</tr>
<tr>
<td>April 7</td>
<td><em>Advanced Topics TBD</em></td>
<td></td>
<td>Project Q #9</td>
</tr>
<tr>
<td>April 14</td>
<td><em>Advanced Topics TBD</em></td>
<td></td>
<td>HW #5; Project Q #10</td>
</tr>
<tr>
<td>April 21</td>
<td><em>Advanced Topics TBD</em></td>
<td></td>
<td>Project Q #11</td>
</tr>
<tr>
<td>April 28</td>
<td><em>Advanced Topics TBD</em></td>
<td></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>

*NOTE: All readings will be outlined by the start of the semester. Weekly readings will typically include a relevant textbook chapter and an example paper from the empirical literature.

**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 (SPSS will also be accommodated).

**Reviewing the Basics – Statistics**
- Statistics terminology: hypothesis testing, p-values, confidence intervals, R², etc.
- Statistical distributions, i.e. normal, chi-square, etc.
- Understanding and working with various types of data – continuous vs categorical, cross-sectional vs time series, etc.
- Introduction to STATA software

**Linear Regression**
- Reviewing the basics of standard linear regression models
- Understanding the assumptions: what does it mean to be BLUE?
- Specifying a statistical model
- Interpreting regression results
Understanding Variables
- Organizing, summarizing, and displaying data
- Finding data ‘issues’ – missing variables, improperly coded, etc.
- Understanding and interpreting transformations – log, square, elasticity, etc.
- Using and interpreting interaction terms and dummy variables

Regression Diagnostics
- Using statistical tests and graphics to diagnose whether the BLUE assumptions are appropriate
- Identifying and dealing with common statistical issues
  - Multicollinearity
  - Heteroscedasticity
  - Omitted variable bias
  - Confounding
  - Outliers
  - Etc.

Potential Advanced Topics (final list will be determined by student data interests)
Discrete Dependent Variables: What to do when your dependent variable isn’t continuous?
- Understanding dummy and categorical variables common to survey work
- Building, understanding, and interpreting common modeling approaches – probit and logit models, linear probability models, etc.

Panel Data: What to do when your data are systematically grouped or clustered, with multiple observations within states, cities, classrooms, etc.?
- 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

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