POLSCI 630
Introducing Empirical Approaches to Political Science

Seminar: M/W 10:25 – 11:40 AM, Gross Hall 105
Lab: Friday TBD, Bunche Lab

Prof. Edmund Malesky, Ph.D
ejm@duke.edu
919-660-4340
Office hours: M/W 1-2 PM or by appointment Office location: Gross Hall 231

TA's: Anh Le (anh.le@duke.edu) and Jan Vogler (jan.vogler@duke.edu)

Course app for problem sets and announcements: http://ps630-f15.herokuapp.com/.

Course Description & Objectives

This course covers basic techniques in quantitative political analysis. It introduces students to widely-used procedures for regression analysis, and provides intuitive, applied, and formal foundations for regression and more advanced methods covered in later studies. This course will use rudimentary calculus and matrix algebra rather intensively. This course relies on R for statistical software.

This course strives to achieve four overarching goals. First, students will become literate in regression analysis. Even though basic ordinary least squares (OLS) is not still commonly used in many contemporary political science analyses, the regression framework—and related interpretations of marginal effects, hypothesis testing, causal identification, forecasting, and bias-efficiency tradeoffs—readily generalizes to more state-of-the-art applications. Second, students will establish a foundation in statistical theory and applied econometrics that will help students move forward with their methods training in the stats, economics and/or political science sequences. Third, students will develop experience working with data on topics related to political science, in the context of in-class examples, lab practicums, take-home problem sets and a final research paper. Fourth, students will practice applying the quantitative methods to analyses of their own research questions. Students will collect and analyze data as part of a final project.

Requirements

Grades in the course will be based on the following items:

- **60%** — Problem Sets. At the end of each lab session (Fridays), students will receive a take-home problem set to complete. The problem sets will ask students to demonstrate mastery of statistical theory, as well as in analyzing data to draw inferences. The problem sets are due by the start of class the following Tuesday. Problem sets should be submitted electronically in PDF form, ideally composed using Latex. In addition, students should submit a copy of their R code, with annotations to allow graders to understand their steps. Students will then cross-grade the anonymous submissions using the key provided by the instructors by the start of the following lab session. The peer graders are only responsible for identifying areas where corrections are needed—

the instructors are ultimately responsible for assigning the grades to be recorded for each assignment. During each lab session, the instructors will cover any questions about the previous problem set before moving on to the current week's material.

- **40%** — Methods paper, 15-20 pages. This paper will demonstrate students' technical mastery of the practical aspects of OLS regression in the context of a specific research problem of their own formulation. Students will choose a topic, develop a hypothesis, and test it quantitatively. The primary restriction here is that the dependent variable should be continuous rather than discrete, so that it is suitable for the kinds of techniques we will be learning in class. The format should be similar to a "research note" in *APSR* or *JOP*, but with greater emphasis on the technical details. The final paper is due on **December 13, by 2pm**. Final papers should be submitted electronically in PDF form, again, ideally composed using Latex. Students should also submit a copy of their R code, with annotations to allow graders to understand their steps.

Course Policies

Late assignments will be penalized. Each day the assignment is late will result in a drop of a letter grade, e.g., A to B, etc. Problem sets and the final paper will be graded on a 16-point scale as follows:

[15-16] ---- A
[14-15) ---- A-
[13-14) ---- B+
[11-13) ---- B
[10-11) ---- B-
[9-10) ---- C+
[7-9) ---- C
[6-7) ---- C-
[2-6) ---- D
[0-2) ---- F

The Duke community standard is in effect throughout the semester. By taking this course, you affirm that it is a violation of the code to cheat on assignments, to plagiarize, to deviate from the teacher's instructions about collaboration on work that is submitted for grades, to give false information to a faculty member, and to undertake any other form of academic misconduct. You also affirm that if you witness others violating the code you have a duty to report them.

Given the nature of this course, some amount of student collaboration is expected and permitted. Students may work on developing their R syntax together by sharing helpful tips, and students are welcome to compare outputs with one another, with the following stipulations: 1) the sharing of ideas must not be one directional, where one student is doing the work and the other is free riding; and 2) the actual write-up of the work that is handed in must be the work of each individual, with absolutely no copying and pasting from one student's work to another's.

Texts

We will rely on the following texts in this course:

- Angrist, Joshua D. & Jörn-Steffen Pischke. 2015. *Mastering 'Metrics: The Path from Cause to Effect.* Princeton University Press.
    - If you feel comfortable with your math skills, you may substitute the more advanced *Mostly Harmless Econometrics* by the same authors.
- Moore, William H. & David A. Siegel. 2013. *A Mathematics Course for Political & Social Research*. Princeton University Press.
- Wooldridge, Jeffrey. 2013. *Introductory Econometrics: A Modern Approach*, 5th Edition. Cengage Learning.

Schedule

1. **August 29- 31**: Probability theory and distributions (Moore & Siegel chs. 9 & 10)
    a. Probabilities
    b. Conditional Probabilities
    c. Random variables

2. **September 5-7**: Properties of random variables (Fox chs. 1-4, Moore & Siegel ch. 11)
    a. Expected value and variance
    b. Theoretical distributions
    c. Basic matrix notation

3. **September 12-14**: Comparisons and inference (Mastering 'Metrics ch. 1)
    a. Counterfactual comparisons, covariance, correlation and cross-tabs
    b. Hypothesis testing
        i. Experimental ideal
        ii. Difference of means
        iii. Correlation coefficient

4. **September 19-21**: Regression model estimation (Wooldridge ch. 1 & 2)
    a. Least squares estimators
    b. Predictions (expectations and errors)
    c. Model fit
        i. PRE
        ii. R^2
        iii. Root-MSE
        iv. F-test

5. **September 26-28**: Regression model interpretation (Wooldridge ch 3 & 4, Mastering 'Metrics ch. 2)
    a. Marginal effects and intercepts
    b. Hypothesis testing
    c. Multiple regression
    d. Graphical representations
        i. Coefficient plots and distributions
        ii. Substantive effects
        iii. Plotting residuals

iv. Simulation approaches

6. **October 3 to 12** (two weeks): Dummy variables and interactions (Wooldridge ch. 7)
    a. Additive and interactive effects
    b. Dummy variables and fixed effects
    c. ANOVA
    d. Interpretations of interaction models
    e. Graphical representation of conditional marginal effects

7. **October 17 to 19**: Statistical Theory of OLS (Wooldridge ch. 5)
    a. Matrix representation of multiple regression
    b. Gauss Markov Theorem
    c. Bias and efficiency

8. **October 24 to 26**: Omitted Variable Bias and Endogeneity (Mastering 'Metrics ch. 3, Wooldridge ch 15)
    a. Matching
    b. Control variables
    c. IV/2SLS regression
    d. Sample selection bias

9. **Oct 31 – Nov 9** (two weeks): Model specification and diagnostics (Wooldridge ch. 6 & 8)
    a. Outliers and leverage
    b. Missing data and imputation
    c. Nonlinear transformations and functional forms
    d. Heteroskedasticitity
    e. Collinearity
    f. Bias/efficiency tradeoff

10. **November 14–16**: Autocorrelation across time and space (Wooldridge 10-12)
    a. Serial correlation in the errors
    b. Spatial correlation
    c. Primer on mixed effects models

11. **November 21–23**: Other primers in causal identification (Mastering 'Metrics chs. 4–6)
    a. Natural experiments
    b. Difference in differences
    c. Regression discontinuity designs

12. **November 28–30**: Application
    a. Applying Statistics to Political Science Research
    b. Research Design
    c. Presenting your work