

Quantitative Analysis of Political Data (Post-Irma Edition)

Class: Mondays & Wednesdays, 11:00-12:20, ACE 328

Professor: Jack Reilly

jreilly@ncf.edu

Office: SSC 205

Office Hours: Wednesdays 1-3 and by appointment.

Appointments can be booked at jackreilly.youcanbook.me

Syllabus Revision: September 19, 2017

Summary Description: This course is intended for all students who intend to conduct quantitative research in political science, and will also be useful for other social science students interested in quantitative analysis. It will introduce students to the scientific norms of the study of political "science", research design, and the fundamental problem of statistical inference. Major goals for the course include the introduction and use of major political and social science datasets, including the American National Election Studies, the General Social Survey, the Cooperative Congressional Election Study, and others, as well as the practical use of advanced statistical techniques to analyze these datasets.

Course Structure: This will be the most applied "stats" course you will ever take. We have two main components to the class: a theoretical track, introducing relevant statistical techniques and methods, and an applied track, in which we learn about writing code for statistical analysis software (Stata), conduct analyses, and replicate previous studies.

Prerequisite: Introduction to Statistics or equivalent. Students should already be familiar with the concept of hypothesis testing and bivariate regression to take the class. This course is recommended for students who intend to take Econometrics in the spring. Upper-division work in social science is highly recommended before taking the course.

Books and Software

Required

- Lewis-Beck and Lewis-Beck, 2015. Applied Regression: An Introduction, Second Edition. Sage "Little Green Book" #22.
- Tufte, 1974. Data Analysis for Politics and Policy.
 - ebook: <http://www.edwardtufte.com/tufte/ebooks>
- Acock, 2016. A Gentle Introduction to Stata (Fourth or Fifth Edition)

Recommended

- Long, 2009. Workflow of Data Analysis Using Stata. Stata Press.
- Berry, 1985. Multiple Regression in Practice. Sage "Little Green Book" #50.

Secondary and Optional Reference

- Long and Freese, 2014. Regression Models for Categorical Dependent Variables Using Stata, 3rd Edition. Stata Press.
- Lewis-Beck, 1995. "Data Analysis: An Introduction". Sage "Little Green Book" #103.
- Jaccard & Turrisi, 2003. Interaction Effects in Multiple Regression. Sage "Little Green Book" #72.
- Aldrich, 1984. Linear Probability, Logit, and Probit Models. Sage "Little Green Book" #45.
- Agresti and Finlay, 2008. Statistical Methods for the Social Sciences. Pearson.
- Kellstedt and Whitten, 2013. The Fundamentals of Political Science Research. Cambridge.
- Long, 1997. Regression Models for Categorical and Limited Dependent Variables. Sage.
- Shalizi, 2015. Advanced Data Analysis from an Elementary Point of View. Online.
- Wheelan, 2014. Naked Statistics.
- Gonick and Smith, 1993. The Cartoon Guide to Statistics.
- Huff and Gels, 1993. How to Lie with Statistics.
- Pampel, 2000. "Logistic Regression: A Primer" Sage "Little Green Book" #132.
- Fox, 1991. "Regression Diagnostics" Sage Green Book #70
- <http://students.brown.edu/seeing-theory/>

Software: A primary component of the class is learning how to effectively and practically use statistical software. The main software package we will use, Stata, is the standard package used by practicing political scientists, and is very common in sociology and economics as well. It is also frequently used by businesses, political think tanks, policy analysts, and statistical consultants. New College has licenses available for use in our classroom, ACE 328, as well as the computers in the Quantitative Social Science Lab (ACE 228) and the ARC.

If you want to use Stata on your personal computer, you can purchase Stata as either a temporary six-month license or a perpetual license. If you wish to do this, please talk to me before purchasing to make sure you buy the right version (in short: do not buy "small Stata").

Course Requirements

1. Daily Reading & Preparation
2. Assignments
 1. Technical
 1. Problem Sets
 2. Replications
 2. Interpretive
 1. Class Summaries (in ~200 words or less)
 2. Assignment Summaries
3. Exams
 1. In-class exams (2)
 2. Take-home exam
4. Final Project
 1. "Pre-Registration" Papers

2. Final Presentation
3. Peer Review & Discussant Duties
4. Final Research Paper

Assignments: There are two kinds of assignments in this class: generic “problem sets”, testing statistical know-how and abilities, and replications, which require you to come as close as you can to replicating an existing piece of analysis (to be assigned by the professor). For each kind of assignment, you will be evaluated not only on whether your answers are mathematically correct, but also on coding style and the clarity of your presentation of statistical results.

All assignments are due on the Wednesday of each class week, at the beginning of class, electronically. As we will go over assignments in class the day they are due, late assignments **will not count for credit**. Assignments are due to me via SubmitBox here:

<https://jacklreilly.submitbox.org/>

The course password is “ols”, without quotes.

Exams: There are three exams in this class: two in class and one take home. The course is cumulative, and each exam will be comprehensive. These exam dates will not change¹: please write them down in your calendar now to make sure conflicts don’t arise.

The **first exam** is on **October 9, 2015**.

The **second exam** is on **November 15, 2017**.

The **take-home exam** will be assigned **after** the second exam and due November 17, 2017.

Final Project: You will be required to conduct an original research project using an existing social science dataset and present it to the class. Presentations will take place during the last two weeks of class, and papers will be due at the end of exam week: December 11, 2015.

Course Expectations

Etiquette: Course participants must be courteous to the professor and fellow students. Attend class on time, listen to fellow students when they talk, disagree (or agree) with others’ arguments professionally. Cell phones should be silent and kept out of sight, computers should be used only for educational purposes: no text messaging, Internet browsing, facebook-ing, tweeting, etc.

Office Hours, etc: I encourage you to stop by our office hours at any point if you have questions about the course, the readings, school, etc. In addition to formal office hours, I have an “open-blind” policy: if the blinds to my office are open, you are welcome to come in. (Just

¹ Assuming, you know, that Irma or her friends don’t come back.

knock and open the door.) If you want to be sure you can speak with me, setting up an appointment beforehand is always a good idea. I maintain an open booking system online that allows you to sign up for a particular timeslot: jacklreilly.youcanbook.me

E-mail: Students can generally expect a response to all e-mails within 24 hours, excepting weekends. I'm happy to answer any questions over e-mail that require less than a paragraph in response. Questions that require more than a short paragraph in response should be addressed in person.

Academic Policies

Students with Disabilities: (NCF Policy) "Students in need of academic accommodations for a disability may consult with the office of Students Disability Services (SDS) to arrange appropriate accommodations. Students are required to give reasonable notice prior to requesting an accommodation. Students may request an appointment with SDS in-person (HCL3), via phone at 941-487-4496, or via email at disabilityservices@ncf.edu."

Religious Observance: (NCF Policy) No student shall be compelled to attend class or sit for an examination at a day or time when s/he would normally be engaged in religious observance or on a day or time prohibited by his or her religious belief. Students are expected to notify their instructors if they intend to be absent for a class or announced examination, in accordance with the policy, prior to the scheduled meeting.

Academic Integrity: (NCF Policy) Any suspected instance of plagiarism will be handled in accordance with the College's policy on academic dishonesty.

Course Overview

There are two main tracks to the course. The first track, statistics, will cover topics related to the linear regression model. This includes some or all of the central limit theorem, hypothesis testing, bivariate regression, multiple regression, regression with categorical independent variables, interactive effects, multicollinearity, nonspherical errors, and an introduction to regression with categorical dependent variables (the generalized linear model.)

The second track, the workflow of data analysis, focuses on the practical components of statistical analysis. Topics include replication, coding and writing style, debugging, annotation, automation, presentation, graphics, data cleaning, storage, and management. Generally speaking, we'll cover material from the first track on Mondays and material from the second on Wednesdays.

Course Schedule (BETA - Particular Topics, Readings and Assignments May Change)

Week	Statistics (M)	Data Workflow (W)	Reading*	Work Due
1	Measurement & Central Limit Theorem Review	DIAGNOSTIC QUIZ	<i>Kellstedt & Whitten, ch 7, Tufte, ch 1, Acock 1-3</i>	Diagnostic Quiz
2	LABOR DAY	Introduction to Stata and Coding Style	<i>Kellstedt & Whitten, ch 8, Tufte, ch 2, Acock ch 4</i>	Assignment 0
3	IRMA	IRMA	-	-
4	Univariate and Bivariate Inference	Replicability, Cleaning & Recoding Data	Tufte, ch 3, Lewis-Beck ² , ch 1 & ch 2 Acock ch 8	Assignment 1: Stata Basics
5	Multiple Regression	Surveys, Choosing Variables, Review	Tufte, ch 4 Lewis-Beck ² , ch 3 Acock ch 10	Assignment 2: Regression
6	Categorical Independent Variables	Replication	Acock, ch 5 Jacoby, ch 1-2	Assignment 3: Replication 1
7	EXAM I: BASICS	Graphics and Visualization I	Acock, ch 6, Lewis-Beck ² , ch 4	EXAM
B	FALL BREAK			
8	Nonlinearity & Interactions	Predicted Values & Marginal Effect Plots	<i>Jaccard & Turrisi, chs 1-2</i>	Assignment 4: Replication 2
9	Nonlinearity & Transformations	Graphics and Visualization II	<i>Berry & Feldman, ch 5; Long, ch 4; Jacoby, Ch 3</i>	Assignment 5: Interactions & Dummies
10	Outliers	Reproducibility, Preparing Data, Weights, Loops	<i>Berry & Feldman, chs 4</i>	Pre-Registration Assignment 1
11	Logistic Regression	EXAM II: EXTENSIONS	Pollock, <i>Logistic Regression</i> ; Acock, ch 11	TAKE-HOME EXAM
12	Ordinal and Multivariate Logit	Giving a Data Based Presentation, Review	Acock, ch 11	Pre-Registration Assignment 2
13	PRESENTATIONS			
14				PAPER

*Additional secondary readings will be announced in class and posted online. Italicized readings will be available as pdfs on the course Google Drive.