Quantitative Political Analysis II

COURSE INFORMATION

Term: Fall 2020
Level: Intermediate
Meet: T/F 1-2:20 (Class) R 7-8 (Lab)
Room: N/A; remote class. Course may be completed asynchronously
Syllabus Revision: August 18, 2020

Instructor: Jack Reilly
Office: Social Sciences 205
E-mail: jreilly@ncf.edu
Office Hours: Tuesdays and Fridays 4-5, scheduled via zoom or drop-in via slack
Appointments: jacklreilly.com/appointments

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.

REMOTE INSTRUCTION

Formally: This class is a remote course that meets twice per week. The first meeting is asynchronous via recorded lecture and follow-along code examples, and may be completed at any time.[1] The second meeting is primarily synchronous, but may also include asynchronous prerecorded material. Synchronous sessions are held over zoom. Attendance at synchronous sessions is not required to complete the course, but is highly recommended.

Informally: None of this whole pandemic thing is ideal, espeically for such a hands-on data class, and we're all just trying to make this work as best as we can, y'all. I think this plan is the best I can make it, but we'll switch it up if something's clearly not working.[2]

PREREQUISITE

An introductory class in statistics (Quantitative Political Analysis I, Introduction to Statistics, Dealing with Data I, Introduction to Biostatistics, etc). 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 a social science is highly recommended before taking the course.

Materials

BOOKS

Required

- Lewis-Beck and Lewis-Beck, 2015. Applied Regression: An Introduction, Second Edition. Sage Green Book #22.
- Tufte, 1974. Data Analysis for Politics and Policy. (ebook: `http://www.edwardtufte.com/tufte/ebooks`)

[1] The professor will be available over slack during the formal first session of the week for consultations.
[2] ie, if the first sessions of the week need to become fully synchronous for question and answer periods, they will.

- Acock, 2016. A Gentle Introduction to Stata (any edition fourth or newer should do)

Recommended

- Long, 2009. Workflow of Data Analysis Using Stata. Stata Press.
- Berry, 1985. Multiple Regression in Practice. Sage Green Book #50.
- Gelman, Hill, and Vehtari, 2020. Regression and Other Stories. Cambridge.

The primary books for the class are Lewis-Beck[2], Tufte, and Acock, which are readable and cover the core course material in a relatively accessible fashion. Lewis-Beck[2] and Tufte primarily cover the conceptual material; Acock covers Stata, the statistical software we will be using. What they give in accessibility, however, they give up slightly in comprehensiveness and depth. Berry & Feldman provides slightly more depth for our later material, but significantly more depth comes from the recently released Gelman, Hill, and Vehtari, which doubles as one of the best modern treatments of applied regression and provides a comprehensive treatment of our course material, as well as material for a successive class, in one volume.[3] Its depth comes at a cost, however, as the authors are somewhat eager to introduce relatively complex additional concepts quickly and early. There are reading assignments in the class for following along with either the combination of smaller books and with Gelman, Hill, and Vehtari - you may choose which track to read (or do both).

Finally, Long provides an excellent overview of workflow and pragmatic statistical practice considerations regardless of the software you choose; it just so happens that the language he focuses on is Stata, as well.

REMOTE COMPUTING & SOFTWARE

This is a remote course. As such, to successfully complete the class, you will need internet access and a device capable of running or accessing the following software: Canvas, Zoom, Google Drive, and Slack. You will also need the ability to play mp3 audio files and mp4 video files. You may find details, links, and locations for all electronic course materials on the course Canvas page.

STATISTICAL COMPUTING & 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 political think tanks, policy analysts, financial analysts, businesses, and statistical consultants. New College has licenses available for use in NCF computer classrooms as well as the computers in the Quantitative Social Science Lab (ACE 228), the ARC, HNS 108, and the Bon House Lab.

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, make sure you buy the right version (in short: buy "Stata/IC"). I strongly recommend this for the class if you are completing it remotely or planning on using your own computer exclusively on campus.[4]

New College has also made Stata available for student use through the virtual desktop client (`vdi.ncf.edu`), which you can access from your own computer anywhere with an internet connection. A guide to using the VDI for Stata, created by Professor Fidalgo, may be found on the course canvas page.

[3] It does so with examples in R, rather than Stata, as well.

[4] Think of the $45 price to rent the software as textbook expense. Books for this class are otherwise quite inexpensive; all can be purchased for under $20 combined on Amazon as of this writing.

If you prefer to complete the course using the free and open-source R statistical language, you may do so (or may do so in addition to completing it in Stata for an additional mod tutorial credit). However, while solutions in R will be provided, we will not have time to go over them in detail in class, and formal instruction will be more limited, so it is recommended that you have prior experience in R, as well as some significant stick-to-itiveness, before choosing this path.

Course Requirements

OVERVIEW

Satisfactory completion of the course requires completion of the following:

1. Daily Reading & Preparation
2. Assignments
   (a) Technical
      i. Problem Sets
      ii. Replications
   (b) Interpretive
      i. Class Summaries (in 200 words or less)
      ii. Assignment Summaries
3. Exams
   (a) Conceptual exams (2)
   (b) Practical exams (2)
4. Final Project
   (a) Pre-Registration Papers
   (b) Final Presentation
   (c) Peer Review & Discussant Duties
   (d) 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 Friday 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 canvas.

You may miss two regular assignments without any kind of penalty or comment in your narrative evaluation (aka, you get two "freebies").

EXAMS

There are two exams in this class. The course is cumulative, and each exam will be comprehensive. As befits a remote class, all exams will be take-home.

The **first exam** is due **September 29, 2020**, with its practical portion due three days later.
The **second exam** is due **October 27, 2020**, with its practical portion due three days later.

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 weeks of class, and papers will be due at the end of exam week: December 4, 2020.

Course Expectations

OFFICE HOURS — I encourage you to stop by my 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-door policy: if the door is open, you are welcome to come in. If you want to be sure you can speak with me, setting up an appointment beforehand at jacklreilly.youcanbook.me is always a good idea.

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.

A NOTE ON WRITING — Clear writing and argumentation is a critical element to success in this class. I strongly recommend exploring the options for writing (and revising!) assistance at the Writing Resource Center. You can schedule an appointment through the writing center here: https://ncf.mywconline.com

NEW COLLEGE POLICY STATEMENTS — **Students with Disabilities:** Any student who, because of a documented disability, may require special arrangements and/or accommodations, should contact the instructor as soon as possible to make necessary arrangements. Students must present appropriate verification from Student Disability Services during the instructor's office hours. Please note that instructors are NOT permitted to provide classroom accommodations to a student until appropriate verification from Student Disability Services has been provided. For additional information, please visit Student Disability Services in HCL 3 and/or email disabilityservices@ncf.edu.

Religious Observance: 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: Any suspected instance of plagiarism will be handled in accordance with the College's policy on academic dishonesty.

Class Schedule

OVERVIEW — There are two main tracks to the course. The first track, the conceptual track, 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 will cover material from the first track on the first day of the week and material from the second on the second day of the week.

TOPICS OUTLINE (Subject to change)

W	Conceptual	Workflow	Work
1	DIAGNOSTIC QUIZ	Stata/R Crash Course: Coding Style	Diagnostic Quiz
2	Regression and Inference	Replicability, Cleaning, Recoding	A1: Stata Basics
3	Multiple Regression	Surveys, Choosing Variables	A2: Regression
4	Categorical IVs	Graphics I	A3: More Regression
5	Categorical Interactions	Replication	A4: Replication 1
6	EXAM I	Catchup & Review	Exam
7	Continuous Interactions	Predicted Values & Marginal Effects	A5: Replication 2
B	BABY FALL BREAK		
8	Transformations	Marginal Effects Plots	A6: Interactions
9	Outliers & Error Terms	Graphics II	A7: Replication 3
10	EXAM II	Catchup & Review	Exam
11	Logistic Regression	Advanced Stata	A8: Replication 4
12	Ordinal & Multi Logit	Presenting Your Work, Catchup	Pre-Registration 1
13	TBD	Presentations	Pre-Registration 2
14	Presentations	Turkey Time	Presentations
F	FINALS		Papers

Additional Resources

- Introductory Statistics
  - Kellstedt and Whitten, 2013. The Fundamentals of Political Science Research. Cambridge.
  - Wheelan, 2014. Naked Statistics.
  - Gonick and Smith, 1993. The Cartoon Guide to Statistics.
  - Lewis-Beck, 1995. Data Analysis: An Introduction. Sage Green Book #103.
  - Jaccard and Turrisi, 2003. Interaction Effects in Multiple Regression. Sage Green Book #72.
  - Aldrich, 1984. Linear Probability, Logit, and Probit Models. Sage Green Book #45.
  - Agresti and Finlay, 2008. Statistical Methods for the Social Sciences. Pearson.
  - Huff and Gels, 1993. How to Lie with Statistics.
  - Pampel, 2000. Logistic Regression: A Primer. Sage Green Book #132.
  - Fox, 1991. Regression Diagnostics Sage Green Book #70
  - http://students.brown.edu/seeing-theory/
  - http://www.reed.edu/psychology/stata/index.html
- More Advanced Statistics
  - Fox, 2015. Applied Regression Analysis and Generalized Linear Models. (also the R companion)
  - Long, 1997. Regression Models for Categorical and Limited Dependent Variables. Sage.
  - Long and Freese, 2014. Regression Models for Categorical Dependent Variables Using Stata, 3rd Edition. Stata Press.
  - Shalizi, 2015. Advanced Data Analysis from an Elementary Point of View. Online.
  - McElreath, 2015. Statistical Rethinking.
  - Gelman and Hill, 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models.

  - Monogan, 20xx. Political Analysis Using R.
  - James et al. 2017. An Introduction to Statistical Learning.

- Graphics
  - Healy and Moody, 2014. "Data Visualization in Sociology". *Annual Review of Sociology.*
  - Healy, 2018. Data Visualization: An Introduction. https://socviz.co
  - Tufte, 2001. The Visual Display of Quantitative Information, 2nd ed.

- Workflow & Data Management
  - Bowers, 2011. "Six Steps to a Better Relationship With Your Future Self" *The Political Methodologist.*
  - Healy, 2018. The Plain Person's Guide to Plain Text Social Science. http://plaintext.co/