Syllabus: Quantitative Political Analysis II

COURSE INFORMATION	*Term:* Fall 2022	*Instructor:* Jack Reilly
	Level: Intermediate	*Office:* Social Sciences 205
	Meet: T/F 1-2:20 (Class) R 7-8 (Lab)	*E-mail:* jreilly@ncf.edu
	Room: ACE 329	*Office Hours:* Tues 10-11, Weds 1-3, by appt
	Syllabus Revision: September 1, 2022	*Appointments:* jacklreilly.com/appointments

DESCRIPTION

This course is designed for all students who intend to conduct quantitative research in political and social science. It will also be useful to other students interested in quantitative analysis. We will take a pragmatic approach to data analysis, focusing both on formal statistical analysis and on the actual practice of utilizing and managing social science data. Statistically, we will cover a number of topics, including multiple regression, regression with categorical independent variables, interaction terms, regression diagnostics, and regression with categorical dependent variables (logistic regression.) For data, we will use a variety of major political and social science datasets, including the American National Election Studies, the General Social Survey, the Cross National Election Project, the Comparative Study of Electoral Systems, and others.

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

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`)
- 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.[1] 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.

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 commonplace in sociology and economics as well. It is also frequently used by political think tanks, policy analysts, financial analysts, businesses, and statistical consultants.[2] 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.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.

If you want to use Stata on your personal computer, you can purchase Stata as a temporary license (six month or year) or as a perpetual license. If you wish to do this, make sure you buy the right version (in short, you probably don't need more than the basic edition, Stata/BE). I strongly recommend this for the class if you plan on using your own computer exclusively on campus. A six-month Stata license costs $48, and you can request a week long student trial version here: `https://www.stata.com/customer-service/short-term-license/`. [3]

Course Requirements

OVERVIEW

Satisfactory completion of the course requires completion of the following:

1. Daily Reading & Preparation
2. Assignments
   (a) Problem Sets (4)
   (b) Replications (4)
3. Exams (2)
4. Final Project

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

[2] 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, we will not have time to go over R solutions in detail in class, and formal instruction will be limited, so it is recommended that you have prior experience in R, as well as some significant stick-to-itiveness, before choosing this path.

[3] Think of the $48 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.

(a) Pre-Registration Paper

(b) Final Presentation

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". Don't use them too early!). Freebies cannot be used for tests or the final presentations. Double weight assignments cost two freebies to miss.

EXAMS

There are two exams in this class. The course is cumulative; each exam is comprehensive. The **first exam** is **Friday, October 8, 2021**, and the **second exam** is **Friday, November 12, 2021**. Exams may have in-class and take-home portions.

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 week of class.

Course Expectations & Guidelines

ETIQUETTE & DECORUM

A college course, especially a small one, is fundamentally a community. Be courteous to fellow students and the professor. Don't let yourself be distracted by your cell phone in class. If you disagree with something someone says, do so respectfully. Engage with each other and remember: your shared goal is to learn from each other as well as the professor.

OFFICE & CONSULTATION HOURS, APPOINTMENTS

I encourage you to chat with me at any point if you have questions about the course, the readings, college, political science, data science, etc. You have a variety of options available to you to consult with me: in person, over zoom, or on the phone. (If you'd like to meet in person, but prefer outdoors to indoors, please let me know and I will do my best to accommodate you.) For any of those, you can go to my website here: http://jacklreilly.com/appointments and sign up for an appointment at your convenience. At minimum, I am always available Tuesdays, from 10-11, for appointments (and typically many other times, as well.)

Second, I maintain "drop-in" hours every Wednesday from 1-3 in my office, Social Sciences 205 (Social Sciences is the small pink building on Dort Promenade before you cross College Drive to get to College Hall) - for these, there is no need to schedule an appointment, just come by. And don't be bashful! Come say hi! I'd like to get the chance to get to know you.

E-MAIL

Students can generally expect a response to all e-mails within 24 hours (and typically sooner), excepting weekends. If your email requires a long response (more than two or three sentences), expect me to encourage you to schedule an appointment with me so that we can more effectively discuss the matter.

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 Crash Course: Coding Style	Diagnostic Quiz
2	Crash Course: Regression Inference	Cleaning & Recoding Data	A1: Stata Basics
3	Multiple Regression	Large Across Time Surveys	A2: Regression
4	Categorical IVs	Weights	A3: More Regression
5	Categorical Interactions	Replication	A4: Replication I (Bartels I)*
6	Continuous Interactions	Predicted Values, Marginal Effects	A5: Interactions
7	Transformations	EXAM	Exam I
	FALL BREAK		
8	Outliers & Error Terms	Graphics I	*A6: Replication II (Valentino) (Opt)*
9	Logistic Regression	*Catchup/Exam Coverage*	A7: Replication III (Mond & Sand)
10	Ordinal & Multinomial Logits	Graphics II	A8: Replication IV (Bartels II)
11	Interactions & Logits	AMEs, MEMs, AEMs, oh my!	A9: Replication V (Reilly)
12	Advanced Stata & Data	EXAM (IN CLASS)	Exam II (Part 1)
	PROJECT SECTION		
13	*Catchup/Exam Coverage*	T(OF)URKEY TIME! & PUMPKIN PIE	Exam II (Part 2)
14	Presenting Quantitative Projects	READING DAYS	Pre-registration
F	Presentations		Presentations

$*$ = *double weight assignment*

ADDITIONAL RESOURCES

Not required for the course, but useful for further study

- 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/

Class Agenda

A NOTE ON READINGS AND REFERENCES

Readings in this class are different than many other classes. You shouldn't imagine just doing the readings beginning to end; but rather, you should look through them after the lecture video for the week to fill in and enhance your understanding.

There is a "main" track for reading in the class, generally focusing on Tufte, Lewis-Beck, and Acock. However, if you prefer a more in-depth treatment of the material, readings are given from Gelman, Hill, and Vehtari's excellent *Regression and Other Stories* as well. Please see the main syllabus for more about these books.

Textbook abbreviations used in this document:

- LB^2=Lewis-Beck and Lewis-Beck, *Applied Regression, An Introduction. 2nd Edition*
- Tufte = Tufte, *Data Analysis for Politics and Policy*
- Acock = Acock, *A Gentle Introduction to Stata*
- Long = Long, *The Workflow of Data Analysis Using Stata*
- GHV = Gelman, Hill, and Vehtari, *Regression and Other Stories*
- FR = Freeman and Ross, *Programming Skills for Data Science*

Assignments Policy

DUE DATES

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.

"FREEBIES"

You may miss two regular assignments without any kind of penalty or comment in your narrative evaluation (aka, you get two "freebies". Don't use them too early!). Freebies cannot be used for tests or the final presentations. Double weight assignments cost two freebies to miss.

References

WEEK 1

Introductions & Beginning Stata Crash Course

1. *Review*: Kellstedt & Whitten, Fundamentals of Political Science Research, ch 7-8
2. Complete: Course interest form (`https://forms.gle/EhWuf1iwQSmJ2i9c6`)
3. Confirm: that you have access to the course slack, google drive, canvas, and that you can use zoom. Install local applications for all of the above as desired.

WEEK 2

Regression & Inference

1. Main Reference:
   - Tufte, ch 1-3
   - LB^2, ch 1-2

- Acock, chs 1-4, 8

2. Alternate References:
    - Alternate Stata Intro: Getting Started With Stata, ch 1 (Mac) (Windows)
    - Alternate Stats Track: GHV, chs 6-7
    - Advanced R Track: FR, chs 9-10

WEEK 3 **Multiple Regression**

1. Main Reference:
    - Tufte, ch 4
    - LB^2, ch 3
    - Miller, Interpreting the substantive significance of multivariable regression coefficients [drive]
    - Nagler, coding style [drive]
    - Acock, section 10.1-10.4
    - Long, ch 1 [drive]
2. Alternate References:
    - GHV, ch 10

WEEK 5 **Categorical Interactions**

1. Main Reference:
    - LB^2, 64-71
    - Stata Track: Acock, section 10.8-10.11
    - Walkthrough: Fidalgo, Interaction Effects [drive]
2. Alternate References:
    - GHV, section 10.3-4

WEEK 6 **Continuous Interactions**

1. Main Reference:
    - Jaccard & Turrisi, chs 1-2
    - LB^2, 69-71
    - Acock, 10.11
    - Walkthrough: Using the Margins Command in Stata for Continuous Interactions
2. Alternate References:
    - GHV, section 10.3

WEEK 7 **Transformations**

1. Main Reference:
    - Acock, 10.12
2. Alternate References:
    - GHV, ch 12

WEEK 8 **Outliers, Error Terms, & Specification Issues**

1. Main Reference:
    - LB^2, ch 4
    - Acock, section 10.5-10.7
2. Alternate References:
    - GHV, ch 11

WEEK 9 **Logistic Regression**

1. Main Reference:
    - Pollock, Logistic Regression [drive]
    - Acock, ch 11.1-11.6
2. Alternate References:
    - GHV, ch 13.1-13.3

WEEK 10 **Interactions and Logits**

1. Main Reference:
    - Acock, ch 11.7
2. Alternate References:
    - GHV, ch 13.7; 14.1-14.2

WEEK 11 **Ordinal & Multinomial Logits**

1. Main Reference:
    - (Optional) Long and Freese (Scan)
2. Alternate References:
    - GHV, 15.5

WEEK 12 **Advanced Stata**

1. Main Reference:
    - Long, all
2. Alternate References:
    - Advanced R: FR, ch 5-8

WEEK 13-14 **Presentations & Projects**

EXTRA I **Graphics I**

1. Main Reference:
    - Acock, ch 5
2. Alternate Reference:
    - FR, ch 15

EXTRA II **Graphics II**

1. Main Reference:
    - Acock, ch 6
2. Alternate Reference:
    - FR, ch 16, 17

Campus Academic Resources

YOUR ACADEMIC ADVISOR Your academic advisor is your first resource at the college for navigating your courses and academic work at the college. Many first year students find themselves uncertain about when it is "ok" to go their advisor with questions or for advice—but the real answer is "anytime"! We all have open office hours that you can simply drop in at, even if you don't have an appointment, and are eager to help you.

WRITING SKILLS Clear writing and argumentation is a critical element to success in college (not to mention, life generally). That said, writing is hard, and students come to college with very different levels of preparation for college level and professional writing. Regardless of your skill and comfort with writing, 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

QUANTITATIVE SKILLS Like writing skills, quantitative literacy is an integral element to success in college (not to mention, life generally). That said, math is (also) hard, and students come to college with very different levels of preparation for college level and professional data literacy. While this course will cover many aspects of data literacy, should you desire additional support beyond what I and/or the course TA can provide, I recommend exploring the options for assistance at the Quantitative Resource Center. The QRC is located in the Academic Resource Center (ARC), located on the first floor of the Jane Bancroft Cook Library.

Student Success Center

Having trouble figuring out how to manage it all? In addition to your faculty mentor and professors, New College has peer to peer coaching and group workshops available at the Student Success Center. The SSC helps you develop the skills necessary for success in college. We offer one-on-one appointments with trained peer coaches, group study sessions, recurring appointments, workshops, printable resources, regular newsletters, and referrals to other campus services. You can find more information here: https://www.ncf.edu/academics/student-success-center/

Additional Resources

Please see the Campus Support Resources document—located in the course canvas page and google drive—for all the academic resources available to you at New College.

New College & Community Academic Policies

Critical Thinking

Students are encouraged to employ critical thinking and to rely on data and verifiable sources to interrogate all assigned readings and subject matter in this course as a way of determining whether they agree with their classmates and/or their instructor. No lesson is intended to espouse, promote, advance, inculcate, or compel a particular feeling, perception, viewpoint or belief.

Student Accessibility

New College of Florida is committed to creating a learning environment that meets the needs of its diverse student body. If you are a student with a disability, or think you may have a disability, you are encouraged to initiate a conversation with the AALC (Advocacy and Accessible Learning Center). The AALC works with students with disabilities to identify reasonable accommodations and plans ways to implement these with your faculty members. Please visit their website for additional information: `https://www.ncf.edu/departments/advocacy-accessibility/`. You may also contact the AALC in-person (HCL3), via phone at 941-487-4844, or via email at `aalc@ncf.edu`. Students are welcome to discuss privately any concerns related to barriers to both fully participating and learning in this course. Students with accommodations are highly encouraged to meet with their primary or partner instructor as soon as possible.

Title IX

New College of Florida is committed to equal access to education pursuant to Title IX of the Educational Amendments of 1972. The law protects all individuals on our campus from gender-based discrimination or exclusion or instances of sexual misconduct. All full-time faculty, full-time staff, and resident advisors are Responsible Employees required to report any known instances of sexual misconduct or gender discrimination to the Title IX Coordinator. Please contact our Title IX coordinator (`titleix@ncf.edu`) or see the website (`https://www.ncf.edu/campus-life/title-ix/`) for more information.

Equity, Diversity, and Equal Opportunity

New College's commitment to excellence can only be realized in a learning environment that is inclusive, characterized by openness to diverse perspectives, and marked by mutual respect. Anything short of this aspiration is inconsistent with our commitment. Equal access, and the opportunity to participate fully in all of our programs and facilities, without regard to race, color, creed, religion, political ideology, national origin, age, marital status, disability, public assistance status, veteran status, gender identity, gender expression, or sexual orientation, is essential to that commitment and will be the standard to which we expect all members of our learning community to adhere.

Academic Integrity

Academic integrity is essential to maintaining a vibrant, healthy, and engaging learning environment for which we all must take responsibility. The New College faculty considers academic dishonesty to be a serious violation of community standards. Students are expected to

refrain from acts of academic dishonesty, which may include:

1. cheating and/or plagiarism (such as: presenting the intellectual work of others as one's own; failing to cite sources; improper paraphrasing via failing to use own words even if a citation is given; partial, incomplete, or inaccurate citation of work of others);
2. unauthorized multiple submissions (submission of the same work for different academic activities, without the approval of the instructor);
3. false citation (false citation of a source or knowingly attributing work to a source from which the referenced material was not obtained);
4. falsifying data (fabricating or altering data to deliberately mislead; for example, changing data to get better experiment results is academically fraudulent);
5. falsifying information, signatures, or initials on official and academic forms.

If you are in doubt about what practices are permissible in an examination, you should consult the professor prior to sitting for the exam. If you lack understanding of how, in a paper or other presentation, to distinguish your thoughts from those of others, the faculty can refer you to standard guidelines and discuss specific questions.

CLASSROOM RECORDING

Florida State Law allows students to record classroom lectures without a requirement for prior notice and without the faculty member's consent, with specific limitations on where and how those recordings can be used.

"A recorded lecture may not be published without the consent of the lecturer," and it may be employed only

a. "for a student's own personal educational use"
b. "in connection with a complaint to the public institution of higher education where the recording was made"
c. "as evidence in, or in preparation for, a criminal or civil proceeding"

Recorded lectures used for other purposes may allow the faculty member to seek damages plus court costs and reasonable attorney fees, with the total recovery not to exceed $200,000.

A class lecture is defined as an educational presentation delivered by faculty or guest lecturer, as part of a New College course, intended to inform or teach enrolled students about a particular subject. A class lecture does not include lab sessions, student presentations, clinical presentations such as patient history, academic exercises involving student participation, assessments (quizzes, tests, exams), field trips, private conversations between students in the class or between a student and the faculty or lecturer during a class session.