

POLS 641: INTRODUCTORY ANALYSIS OF POLITICAL DATA

Northern Illinois University

Fall 2023

Tuesdays and Thursdays 2-3:15

Instructor: Dr. Colin Kuehl

Email: ckuehl@niu.edu

Office Hours: Tuesday & Thursday 12-1

Office: Zulauf 410 and [Zoom](#)

R guru: Robert Durbin

Email: rdurbin1@niu.edu

Office Hours: Tuesday 12-2

Office: Dusable 476

Course Description: This course is the first in NIU Political Science Department's graduate data analysis sequence. It is designed to provide students with an introduction to quantitative methodology used in political research. We will cover basic probability and statistics through multivariate regression with a focus on the use of these methods in the field of political science. This course concentrates on the consumption and application of quantitative methods. Accordingly, it does not focus on the elaborate math driving much of the statistical analysis. Instead, we will focus on the proper use of statistics and how to critically analyze quantitative research. As such, much of the course is focused on learning by doing. Students will learn to program using the R programming language (more below) and will conduct original quantitative analysis using existing data.

Prerequisites: I assume only high school algebra and a tolerance for stubborn work. Regarding the latter, getting your head around the material and making your statistical software package do what you want it to do can sometimes be both frustrating and time consuming. I can only encourage you to keep trying. With persistence (and sometimes a little help), you will eventually figure it out. This is an essential part of the process. Trust me on this. I've been there, too. Regarding the former, calculus is helpful since some of the material we cover makes use of it. However, don't panic: it is not necessary. I will show you some math from time to time in lecture, but this will be solely to provide you with motivating, behind-the-scenes intuition. You are not expected to be able to reproduce it in problem sets.

A Note on Statistical Software: Knowledge of statistical software (or qualitative software) is an increasingly important component of any political scientist's toolbox. The choice of statistical software is one of continuous debate. However, most quantitative researchers today use either STATA or R (others such as SPSS and SAS have dedicated niche followings and Python is growing) to complete their analysis.

This course will be taught using R for a number of reasons. 1) It's free! This allows students to use R on their home computer, laptop, etc and ensures access to R long past your time here at NIU. 2) Cutting edge political methods work is increasingly done using R. This is because, 3) the things you can do using R (and associated packages) is far broader than in STATA or other packages. Finally, 4) basic computer coding and understanding coding structures is an increasingly important skill across a number of domains (its developers actually describe R as a computing environment instead of statistical software to reflect this.). This is not to say that R does not have disadvantages.

R has little to no point-and-click options and is not very intuitive. Accordingly, it has a relatively high start-up cost. However, I strongly believe it will be worth your investment.

For those interested, the course will also provide an introduction to the use of L^AT_EX document preparation software. L^AT_EX is also free and increasingly used across the discipline of political science. It also pairs well with work in R.

Course Meetings: We will meet twice per week. The first, on Tuesday, will be a combination of lecture and discussion on fundamental concepts, such as research design, probability theory, etc. The second meeting, on Thursdays, will focus on application and coding in R. Meetings will be primarily held in person, however depending on scheduling issues and student interest we may move some sessions online.

Readings We have one assigned text for the course:

- Michael A. Bailey, *real stats: Using Econometrics for Political Science and Public Policy*, Oxford. (second edition, first edition is OK)

I recommend purchasing this book. You will continue to refer to it throughout your time here and portions will also be used in 642

In addition to the required text, there are a number of recommended books.

“Pop” data science books:

- Carl Bergstrom & Jevin West *Calling Bullshit*
- Nate Silver *The Signal and the Noise*
- Charles Wheelan *Naked Statistics*
- Tim Harford *The Data Detective*

Basic introductions

- Neil Salkind *Statistics for People who (think they) Hate Statistics*
- Scott Cunningham *Causal Inference Mixtape*
- Gary Klass *Just Plain Data Analysis*

More advanced texts for students looking to go beyond what is taught in the course:

- Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*
- Damodar N. Gujarati *Essentials of Econometrics*
- Jeffrey Wooldridge *Introductory Econometrics*
- Elena Llaudet and Kosuke Imai *Data Analysis for Social Science*
- Jenine Harris *Statistics with R*
- Guido Imbens and Donald Rubin *Causal Inference for Statistics, Social, and Biomedical Sciences*

Additional assigned readings will be found on the course blackboard page. The blackboard page also contains links to helpful websites and online books, especially for R help.

Grading: Final Project (50%), Problem Sets (25%), Midterm (5%), Datacamp (5%), Participation (15%),

- **Final Project:** Your final project will be an independent quantitative analysis of a political science topic of your choosing. You will obtain a dataset, formulate a hypothesis, conduct basic exploratory data analysis and perform, as well as correctly interpret, a multivariate regression. You will present your results to the class at the end of the semester and write up what you found in the form of the empirical section of journal article. An assignment sheet will be provided and we will discuss details in the following weeks.

You will present your findings to the class at the end of the semester. Your final project will be due **December 14th** by midnight by email.

- **Problem Sets:** Throughout the semester you will be expected to complete a number of problem sets. These exercises will give you a chance to apply the ideas and concepts using the R coding we are learning. Problem sets will be distributed on every other Thursday starting in week 2 and will be due on Wednesday at midnight of the second week.

Problem Set Due Dates:

Problem Set #1	Sept. 20th
Problem Set #2	Oct. 4th
Problem Set #3	Oct. 18th
Problem Set #4	Nov. 8th
Problem Set #5	Nov. 22nd

I strongly encourage you to form small study groups to work through the problem sets. However, the write-ups *must be your own*. This means that you can talk through how to solve a problem with someone else, have them check your code for errors, etc. but you must then on your own (later in the day, in silence) put the solution down in writing. No late assignments will be accepted without prior permission. Extensions should be arranged at least twenty-four hours in advance.

- **Midterm:** You will have a midterm in class on **October 26th**. It will assess your understanding of fundamental aspects of research design and statistics as well as some basics of coding in R.
- **Participation:** Class time will be divided between discussion and lecture of the various statistical concepts and techniques (on Tuesdays) and time spent learning and practicing in the R programming language (on Thursdays). My expectation is that this will be a true discussion and students will ask questions when they are confused or having difficulties. Your active participation is especially important given the wide range of backgrounds and small class size. Speaking the language of methodology is an essential component of your graduate training.
- **Datacamp:** During weeks in which no problem set is due you will be expected to complete selected datacamp courses. Datacamp is an online platform for learning various coding and data analysis techniques. As a member of the class you will a complementary six-month membership. These courses are intended to give you some additional practice and exposure to differing techniques as a supplement to what is taught in the course.

Course Policies:

- **Attendance:** Attendance is mandatory without prior arrangement. If circumstances prevent you from attending please let me know by email prior to the beginning of class.. You will be responsible for covering the material you missed.
- **Academic Honesty:** Cheating will not be tolerated. All students will be held to the highest standards of NIUs student code of conduct. All cases will be referred to campus authorities. As noted above, helping each other will be key to your success in this class, however the work you turn in must be your own.
- **Disability Services:** If you need an accommodation for this class, please contact the Disability Resource Center as soon as possible. The DRC coordinates accommodations for students with disabilities. It is located on the 4th floor of the Health Services Building, and can be reached at 815-753-1303 or drc@niu.edu. Also, please contact me privately as soon as possible so we can discuss your accommodations.
- **Email:** I am available through email M-F 8 am to 5 pm. I will respond to all emails within 24 hours during these times. If I forget, please send me a reminder.
- **Office Hours:** I recommend you come to office hours early and often. This will give me a chance to know you, your projects, etc. and give us a chance to talk about things beyond this course
- **Mental Health:** The first semester of graduate school is challenging and anxiety inducing at the best of times. I understand this and have my own battles. We all will handle this uncertainty and stress in different ways. Please take the time for self-care and if you need extensions for mental health reasons do not hesitate to ask. If you or somebody you know is struggling with anxiety or other issues, do not hesitate to reach out. Resources available include the DRC, Student Counseling Services or call 815-306-2777.
- **Health Precautions:** Masks are not required for this class (subject to changes in protocols from the university). Please be respectful of others decisions. Above all, please be considerate of your fellow students. If you are feeling sick stay home. If you might be sick wear a mask. At all times, please do your part to keep everyone healthy.
- **Gender Identity:** As a faculty member, I am committed to using your proper name and pronouns. We will take time during our first class together to do introductions, at which point you can share with all members of the course what name and pronouns you use, as you are comfortable. Additionally, if these change at any point during the semester, please let me know and we can develop a plan to share this information with others in a way that is safe for you.
- **Land Acknowledgement Statement:** Northern Illinois University operates and is built on the traditional lands of the Oceti Sakowin (Sioux), Miami, Bodewadmiakiwen (Potawatomi), Sauk and Meskwaki, and Peoria. These lands are subject to Cession 50 and 148, though their terms have been consistently violated. We seek to acknowledge this land and these peoples in order to honor the legacies, struggles and current existence of Indigenous peoples; situate ourselves within settler-colonial projects; disrupt the erasure of Indigenous peoples; and begin/continue the work of collectively learning and fulfilling our obligations, for those of us who are uninvited guests on Indigenous peoples' lands

Tentative Course Outline:

The tentative schedule for the course is below. Some topics will take just a week, others more. Given our small size and wide variety of backgrounds, we will be flexible in how fast we move. If we move quickly, there are a number of optional topics we can cover at the end of the semester. Treat this as a preliminary overview, and not a complete roadmap. Reading assignments (beyond your book chapters) for the following week will be provided at the end of each class and posted on Blackboard.

Meeting Date (Tuesday)	<u>Discussion Topic</u>	<u>Lab Topic</u>
29-Aug	Introductions	Grad School Tech
5-Sep	Causality and Inference	R Basics & Data Management
12-Sep	Probability Theory	Probability and Loading Packages
19-Sep	Random Variables and Measurement	Getting to Know Your Data
26-Sep	Describing Data	Validity, Cleaning data
3-Oct	Hypothesis Testing	Looking at Relationships
10-Oct	Bivariate Regression	Regression!
17-Oct	Multivariate Regression	More Regression and Merging Data
24-Oct	Review	Midterm
31-Oct	Binary and Categorical Variables	Visualization
7-Nov	Model Specification, OLS Assumptions	Regression Diagnostics
14-Nov	Causal Inference and Experiments	Experiments, Power Analysis
21-Nov	Hackathon(Thanksgiving Week)	
28-Nov	Mixed Methods	Mixed Methods and Student Choice
5-Dec	Research Presentations	
12-Dec	Finals Week/Papers Due	

Optional Topics: Factor Analysis, ANOVA, Power Analysis, Mapping, L^AT_EX, Advanced Visualization

Course Readings and Due Dates:**Week 01, 08/29 :** Introductions

Readings:

- Mutz and Rao (2018) “The Real Reason Liberals Drink Lattes” *PS: Political Science & Politics*, 51(4)
- Achen “Advice for Students Taking a First Political Science Graduate Course in Statistical Methods” *The Political Methodologist*

Recommended:

- Wheelan 2012 “Why I hate calculus, but love Statistics” from *Naked Statistics*

Week 02, 09/05 : Causality and Inference

Readings:

- Bailey Ch. 1
- Bush and Clayton 2022 “Facing Change” *American Political Science Review*

Recommended:

- Angrist and Pischke 2009 “Questions about Questions” from *Mostly Harmless Econometrics*

Week 03, 09/12 : Probability Theory

Readings:

- Ch 3 Probability *OpenIntro Statistics*

Recommended:

-

Due: Datacamp - R Basics**Week 04, 09/19 :** Random Variables

Readings:

- Bailey Ch. 3 pgs 49-64
- Buckley et al. 2023 “Endogenous Popularity: How Perceptions of Support Affect the Popularity of Authoritarian Regimes.” *American Political Science Review*

Due: Problem Set #1

Week 05, 09/26 : Measurement

Readings:

- Adcock and Collier. 2002 “Measurement Validity: A Shared Standard for Qualitative and Quantitative Research” *American Political Science Review*
Read 1 and Skim 1
- Pollock Ch 1 and 2
- Imai. 2017. ”Measurement” from *Quantitative Social Science*

Recommended:

- Gerring “Concepts and Measures” from *Social Science Methodology*

Due: Datacamp - Introduction to the Tidyverse**Week 06, 10/03 :** Hypothesis Testing

Readings:

- Bailey Ch 4
- Resnick. 2019. “Statitistical Significance” *Vox*
- Hullman. “ How to Get Better at Embracing Unknowns”

Recommended:

- Explore “Seeing Theory”

Due: Problem Set #2**Week 07, 10/10 :** Bivariate Regression

Readings:

- Bailey Ch 3,
- OI Regression(skim)

Due: Datacamp - Exploratory Data Analysis**Week 08, 10/17 :** Multivariate Regression

Readings:

- Bailey Ch 5

Recommended:

- Sykes. 1993. “An Introduction to Regression Analysis” Coase-Sandor Institute for Law & Economics Working Paper

Due: Problem Set #3

Week 09, 10/24 : Review and Midterm**Week 10, 10/31 :** Binary and Categorical Variables

Readings:

- Bailey Ch 6

Due: Datacamp - Simple Linear Regression, Assessing Model Fit

Week 11, 11/07 : Model Specification

Readings:

- Bailey reskim 3-5, Ch 7

Recommended:

- Xu. 2021. "A Basic Checklist for Observational Studies in Political Science"

Due: Problem Set #4

Week 12, 11/14 : Causal Inference and Experiments

Readings:

- Bailey Ch 10
- Imbens and Rubin "Causality: The Basic Framework" from *Causal Inference*
- Healy Ch. 1 from *Data Visualization*

Recommended:

- List, Sadoff, and Wagner. 2010. "So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design" *Experimental Economics*
- Tufte "The Visual Display of Quantitative Information"

Due: Datacamp - Understanding Data Visualization (Course) and Cleaning data (chapter)

Week 13, 11/21 : Hackathon (Thanksgiving Week)

Due: Problem Set #5

Week 14, 11/28 : Mixed Methods

Readings:

- Bailey Ch. 12, Ch. 16
- Lieberman 2005 "Nested Analysis as a Mixed-Method Strategy for Comparative Research" *American Political Science Review*
- Weller and Barnes 2014 Ch 1 - 3 *Finding Pathways*
- Notes from the Editors. 2020 *American Political Science Review*

Recommended:

- Seawright *Multi-Method Social Science*
- Bowers and Voors 2016 “How to Improve Your Relationship with Your Future Self”

Due: Datacamp - Analyzing Election and Polling Data in R & 1 Chapter(at least) of your choice

Week 15, 12/05 : Research Presentations