

POLS 641: INTRODUCTORY ANALYSIS OF POLITICAL DATA

Northern Illinois University

Fall 2017

Instructor:	Dr. Colin Kuehl	Time:	Thursday 3:00 – 5:40
Email:	ckuehl@niu.edu	Lab:	Dusable 170
Office Hours:	Tuesday 2-5	Office:	Zulauf 410

Course Description: This course is the first in NIU Political Science Department's graduate data analysis sequence(642 & 643 are the follow-up courses). It is designed to provide graduate students with an introduction to quantitative methodology used in social 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. For those interested, courses in the Economics Department and Division of Statistics provide a much more deeper dive. 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 dirty hands. 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 thisI've been there, too. Regarding the former, calculus is helpful since some of the material we cover makes use of it. However, dont 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 if one of continuous debate. However, most quantitative researchers today use either STATA or R(others such as SPSS and SAS have dedicated niche followings) to complete their analysis.

This course will be taught using R for a number of reasons. 1) Its 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 in is increasing 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.

Readings We have one assigned text for the course:

- Michael A. Bailey, *real stats: Using Econometrics for Political Science and Public Policy*, Oxford, 2016.

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

Basic introductions and "pop" data analysis books:

- Neil Salkind *Statistics for People who (think they) Hate Statistics*
- Gary Klass *Just Plain Data Analysis*
- Nate Silver *The Signal and the Noise*
- Charles Wheelan *Naked Statistics*

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*
- Wooldridge, Jeffrey. *Introductory Econometrics*

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

Grading: Problem Sets (35%), Final Project (50%), Participation (15%)

- **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 as using R coding we are learning. Problem sets will be distributed on every other Friday starting in week 2 and will be due on Wednesday at midnight of the second week(12 days later).

Problem Set Due Dates:

Problem Set #1	Sept. 20th
Problem Set #2	Oct. 4th
Problem Set #3	Oct. 18th
Problem Set #4	Nov. 1st
Problem Set #5	Nov. 15th
Problem Set #6	Nov. 29th

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, but you must then on your own (in another room, later in the day, in silence) put the solution down on paper. No late assignments will be accepted without documentation. Extensions should be arranged at least twenty-four hours in advance.

- **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 as well as performing a regression. The form will be that of the empirical section of a journal article. We will discuss details in the following weeks. Your final project will be due **December 12th** by midnight by email.
- **Participation:** Class time will be divided between discussion and lecture of the various statistical concepts and techniques (usually in the first half of class time) and time spent learning and practicing in the R programming language. My expectation is that this will be a true discussion and students will ask questions when they are unsure. 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 methods training.

Course Policies:

- **Attendance:** Attendance is mandatory. If circumstances prevent you from attending please let me know in advance. 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 NIU's 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.

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.

1. Introduction: Causality and Research Design
Bailey Chapters 1 and 2
2. Probability and Probability Distributions
3. Descriptive Statistics, Measures of Central Tendency, Data Visualization
4. Univariate Analysis and Data Visualization and Management
5. Sampling and Limit Theorems
6. Hypothesis Testing
Bailey Chapter 4

7. Bivariate Regression
Bailey Chapter 3
8. Multivariate Linear Regression
Bailey Chapters 5 and 6
9. Optional Topics: Introduction to STATA or SPSS, Factor Analysis, ANOVA, Power Analysis, Mapping, \LaTeX , advanced visualization