Syllabus

PLSC 501, Spring 2025

Instructor

- 👤 Dave Clark
- 🏛 LNG 57; LN2430
- ✉ dclark@binghamton.edu
- ⊙ clavedark
- office hours: M 1:30-3:30pm

TA

- 👤 Oguzhan Irguren
- ✉ oirgure1@binghamton.edu

Course details

- 📅 Spring 2025
- 📆 Wednesday
- 🕐 9:40-12:40
- 🌐 LNG 332 (Bing SSEL)

Lab

- 📆 Thursday
- 🕐 1:15-2:15
- 🌐 CW 326

Seminar Description

This 4 credit hour seminar is about the principles of linear regression models, focusing on the ordinary least squares approach to regression. The course is data science oriented, emphasizing data management, wrangling, coding in **R**, and data visualization.

The class requires some background (provided in preview materials) in probability theory, matrix algebra, and using **R**. A major goal of the class is to provide the intuition behind how and why we do empirical tests of theories of politics. Anyone can estimate a statistical model, but not everyone can estimate and interpret an informed, thoughtful model. Understanding regression and linking regression with theories of political behavior are key steps toward saying insightful things about politics.

An important thing to realize is that it takes time and repetition to really understand this stuff intuitively. Exposure to regression in general (whether OLS, MLE, Bayes, etc.) by reading, hearing, and doing over and over will make it stick. So don't get down or freaked out if you

feel like you don't get it. Instead, ask questions - of me, of the TA, in class, in workshop, in office hours. Asking is essential.

The class meets one time per week for three hours. The required workshop meets one time per week for 1 hour. My office hours are designed to be homework help hours where I'll work in the grad lab with any of you who are working on the exercises. The most productive pathway for this class is for your to get in the habit of working together, and those office hours are a good time for this.

Course Purpose

This seminar is a requirement in the Political Science Ph.D. curriculum. It is the second (of three) required quantitative methods courses aimed at giving students a wide set of empirical tools to put to use in testing hypotheses about politics. This seminar focuses on the method of Ordinary Least Squares, and the linear model, emphasizing data science skills.

Learning Objectives

At the end of the semester, students will be able to describe the linear model, its assumptions, and interpretation. Students will be able to manage data, estimate linear models, diagnose and correct models, and generate quantities of interest. Students will be able to plot and analyze their estimates and quantities of interest. Students will be able to generate their own research designs, and test hypotheses using the linear model.

Course Requirements and Grades

The seminar requires the following:

- Problem sets - 50% total
- Replication/Extension project (including log, poster, etc.) - 50%

Please note that all written assignments must be submitted as PDFs either compiled in *LaTeX* or in **R** markdown (Quarto).

You'll complete a series of problem sets, mostly applied. How many will depend on how things move along during the term. Regarding the problem sets - the work you turn in for the problem sets should clearly be your own, but I urge you to work together - doing so is a great way to learn and to overcome problems.

A word about completeness - attempt everything. To receive a passing grade in the course, you must finish all elements of the course, so all problem sets, all exams, papers, etc. To complete an element, you must at least attempt all parts of the element - so if a problem set has 10

problems, you must attempt all 10 or the assignment is incomplete, you've not completed every element of the course, and you cannot pass. I realize there may be problems you have trouble with and even get wrong, but you must try - the bottom line is don't turn in incomplete work. Ever.

For technical and statistics training and help, Oguzhan will lead a **required workshop session every week, Thursday 1:15-2:15pm.**

The paper assignment asks you to replicate some published piece of research (details on selecting this we'll discuss in seminar), to comment critically on the theory and design as you replicate, and then to build on the paper in a substantive way, extending the research. That extension is the main part of the assignment - this is one of your first major opportunities to develop a novel contribution to the empirical literature. The replication/extension paper is due the Monday of exam week. You'll keep a research log as you work on the paper during the semester - this is due along with the paper, and is part of your total grade. In addition, we'll have a poster session to present your extensions at the end of the semester. Details to follow.

Grades will be assigned on the following scale:

Grade	Range	Grade	Range
A	94-100%	C+	77-79%
A-	90–93%	C	73-76%
B+	87–89%	C-	70-72%
B	83-86%	D	60-69%
B-	80-82%	F	<60%

Reading

The book for the class is:

- Wooldridge, J.M. 2013. *Introductory Econometrics.* 5th edition, South-Western/Cengage. ISBN 978-1-111-53104-1.

I'm assigning an older edition of Wooldridge (the 5th) so cheaper copies are easier to come by. If you choose an edition other than that, you'll need to reconcile chapters, etc. Also, be warned that the cheaper international edition (paperback) is alleged to differ in substantial ways.

In addition, I'll assign a small number of articles over the course of the term.

It will rarely be the case that we discuss the readings in seminar. Their purposes are two-fold. First, the technical readings (Wooldridge) are to frame the derivations of the models and techniques. These are important because to understand how to interpret a model, you

need a technical understanding of its origins. This is not to say you need to be able to derive these yourselves, or to perform complex mathematical computations. It is to say that if you carefully read about the models themselves, you will certainly find their interpretations easier. The mantra for this course is this - interpretation is everything. Without some understanding of how these models arise, you will find interpretation hard.

Second, the applied readings (i.e. articles) provide just that - application in a political, economic, or social context, and application in terms of interpretation of results. Some applications are nicely done, some less so. Not only are these useful examples of what to do and of what not to do, they will provide really useful models for your own efforts to apply these statistical models. Interpretation is everything.

So reading is up to you, and is crucial. As much as I hate to say this, if I get the sense at any point during the term you are not reading, I will absolutely begin weekly reading quizzes; or perhaps require weekly papers on the readings. Making them up will suck. Taking them will suck. Grading them will suck. There really isn't much reading, and there's just no excuse not to do it.

How to Read

Reading academic stuff is a bit different from most other reading. Especially since you're reading large amounts in graduate school (and the rest of your academic lives), it's important to think about what you're trying to accomplish. With journal articles, the goal has to be to understand the novel claim the paper makes, to understand why that novel claim is interesting or novel (at least according to the author), to understand how the author assesses the evidence, and to understand what that evidence tells us. Most importantly, you should have two answers to each of these - what the author says, and what you think. The author may tell you the argument is important for some reason, and you might disagree - you might think it's unimportant, or important for a different reason, or wrong, or whatever.

Reading technical stuff is yet another category, but equally important. The best way to read math models is to read them aloud (yes, out loud) for three reasons. First, it forces you to slow down as you read it - you talk slower than your eyes move. Second, it forces you to deal with every element of the model - in silence, your mind will just skim right over it to the next set of English words. Third, it engages two senses reading silently does not - vocalization and hearing.

Reading out loud is not going to make everything crystal clear - but it will open some pathways whereby math language isn't completely foreign. Most importantly, doing so bridges the gap between the math and the English accounts that normally follow, and make those English accounts easier to make sense of.

® Resources

There are lots of good, free ® resources online. Here are a few:

- Modern Statistics with R
- R for Data Science
- R Markdown: The Definitive Guide
- R Graphics Cookbook
- Advanced R
- R Markdown Cookbook
- Data Science:A First Introduction
- The Big Book of R

Class Meetings, Office Hours, Assignments

The course will meet this spring entirely in-person in the Social Science Experiment Lab on Wednesdays 9:40am-12:40pm.

Office hours are Mondays 1:30pm-3:30pm. I'll likely hold these in the grad work room to help with your assignments. For an appointment, email me and we'll sort out a time.

All assignments should be turned in on Brightspace - please submit ::

- PDFs generated from *LaTeX* or R Markdown (Quarto).
- annotated R scripts.
- where necessary, data.

Assignments should be instantly replicable - running the code file should produce all models, tables, plots, etc.

Course Policies

Attendance

Attendance is expected, and is essential both in the seminars and lab sessions if you're to succeed in this class.

Academic Integrity

Ideas are the currency in academic exchange, so acknowledging where ideas come from is important. Acknowledging the sources of ideas also helps us identify an idea's lineage which can be important for understanding how that line of thought has developed, and toward promoting future growth. As graduate students, you should have a good understanding of academic honesty and best practices. Here are details of Binghamton's honesty policy.

Artificial Intelligence

Any and all use of AI in work for this course must be documented and cited. Documentation minimally includes reporting prompts used to generate text, code, data, etc. Citation includes a reference to the AI tool used, model, date, etc.

I encourage you to use AI to troubleshoot; I discourage your using it otherwise for coding at this point in your graduate careers for two reasons. First, the more you can code on your own, the better you'll be able to use AI - so you have to learn to code first. Coding implies both the technical, syntactical aspects of writing in a language, and the logical, structural aspects of writing a program. Second, the more you can code independently, the better you'll be able to connect your data and data structures to statistical models - your intuitions will be deeper this way.

It will be tempting to use AI to finish assignments and to get things "right." For this semester, getting them is much more important than getting them right. I'd rather you turn in a wrong answer you've worked through than a right answer you've not.

Here, incidentally, is a statement on AI produced without prompt by AI. I disavow this statement, but think it's instructive:

[I'm not a fan of AI in the classroom. I don't allow laptops, tablets, or phones in class. I don't allow recording of the class. I don't allow the use of AI to generate text for assignments. I don't allow the use of AI to generate code for assignments. I don't allow the use of AI to generate data for assignments. I don't allow the use of AI to generate plots for assignments. I don't allow the use of AI to generate models. I don't allow the use of AI to generate anything for this class. If you're caught using AI in any form for any assignment, you'll receive a failing grade for the assignment, and you'll be reported to the university for academic dishonesty. However, I do allow the use of AI to generate ideas for assignments. For example, you can use AI to generate ideas for your research paper, but you can't use AI to generate the paper itself. Alternatively, you can use AI to generate ideas for your problem sets, but you can't use AI to generate the solutions. If you're unsure whether something is allowed, ask me before you do it.]

Course Schedule

Referring to Wooldridge, 5th edition, 2013

Week 1. 22 Jan **Introduction, Regression Discussion**

- matrix notes, preview materials

Week 2, 29 Jan – **Matrix/Bivariate Regression**

- Wooldridge, chapter 2, Appendix D, Appendix E

Week 3, 5 Feb – **Implementing the Model**

- Wooldridge, chapter 3, Appendix E

Week 4, 12 Feb – **Multivariate Regression, statistical control**

- Wooldridge, chapter 3, Appendix E

Week 5, 19 Feb – **Inference in Regression**

- Wooldridge, chapter 4

Week 6, 26 Feb – **Specifying the Model**

- Wooldridge, chapters 6 and 9
- King (1986)

Week 7, 5 March – **Dummy Variables & Multiplicative Interactions**

- Wooldridge, chapter 7
- Brambor, Clark, and Golder (2006)

Week 8, 12 March – **Spring Break**

Week 9, 19 March – **Interactions (continued)**

- Wooldridge, chapter 7
- Brambor, Clark, and Golder (2006)

Week 10, 26 March – **Limited Dependent Variables**

- Wooldridge, chapter 7.5, and 17.1

Week 11, 2 April – **Panels, Fixed & Random Effects**

- Wooldridge, chapters 13, 14

Week 12, 9 April– **Time Series**

- Wooldridge, chapters 10, 12

Week 13, 16 April – **no classes; Monday classes meet**

Week 14, 23 April – **Non-constant Variance**

- Wooldridge, chapter 8

Week 15, 30 April – **IV models**

- Wooldridge, chapter 15

Week 16, 7 May – **Causal Inference**

- TBA

References

Brambor, T., W. R. Clark, and M. Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14 (1): 63–82.

King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30 (3): 666–87.