POLSCI 630: Probability and Regression in Political Science

Seminar: MW 10:20 – 11:10, Gross Hall 105

Lab: F 10:20 – 11:10, Physics 227

Instructors:

Jon Green

294H Gross Hall

[jon.green@duke.edu](mailto:jon.green@duke.edu)

Office hours: email to set up a time

Chris Johnston

294J Gross Hall

[cdj19@duke.edu](mailto:cdj19@duke.edu)

Office hours: email to set up a time

TA: Stephanie Wright

Office hours: Thursdays 10-11am

**Course Description and Objectives**

This course covers basic techniques in quantitative political analysis with a focus on linear regression. It introduces students to widely used procedures for regression analysis, and provides intuitive, applied, and formal foundations for linear regression as well as some extensions.

This course will use calculus and matrix algebra. For statistical software, we will focus on R. If you wish to learn an alternative, such as Stata or Python, you may of course do so, but it must be in parallel. That is, you must complete all assignments in R. We may be able to provide some guidance with other software, but no guarantees.

This course strives to achieve four overarching goals. First, students will become literate in regression analysis, one of the most widely used modeling approaches in the social sciences. Second, students will establish a foundation in statistical theory and applied econometrics that will help them move forward with their methods training. Third, students will develop experience working with data on topics related to political science, in the context of in-class examples, lab practicums, take-home problem sets, and exams. Fourth, students will develop their programming skills in widely used statistical software.

**Evaluation**

Grades in the course will be based on the following items:

* 10% Attendance. The first part of learning is showing up.
  + Lab is class, and is included in our expectation that students will attend class.
* 25% Quizzes. After the week’s lecture (i.e. after class on Wednesday), students will receive a take-home quiz to be completed before lab on Friday. The quiz will be short and straightforward, designed to evaluate your understanding of core concepts from that week’s material.
  + The lowest two quiz scores will be dropped when calculating the final grade.
* 25%— Problem Sets. On Fridays, students will receive a problem set to be completed over the course of the upcoming week. The problem sets will ask students to demonstrate mastery of statistical theory, as well as in analyzing data to draw inferences.
  + We will use lab time to work on the first question or two from the week’s problem set together, and students will be responsible for completing the rest on their own.
  + Problem sets should be submitted electronically to the Assignment page on Canvas before the following week’s lab (i.e. Friday morning). Grades will be deducted by 1 point (out of 10) for every day they are late.
  + We ask that you submit two files: 1) a professional-looking PDF with your answers and 2) the code that produced those answers.
  + The lowest problem set score from the semester will be dropped when calculating the final grade.
* 20% each — Midterm and Final Exams. There will be two exams. For each, students will be given a dataset and prompt to work on in-person. We will use one class period for them midterm and the separate final exam period for the final.

**Text**

Please purchase the following textbook (if you want to use an earlier edition, make sure the chapters match up to the 7th edition):

* Wooldridge, J. 2019. *Introductory Econometrics: A Modern Approach*, 7th Edition, Boston: Cengage.

**Additional Readings**

We will also read 3 chapters from the following text, which are posted on Canvas:

* James, Gareth, et al. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

**Schedule (13 weeks of content)**

Week 1 (1/10, 1/17): Introduction, linear and non-linear functions

Wooldridge, Appendix A

Week 1 meets 1/10 and 1/17

Week 2 (1/22, 1/24): Probability

Wooldridge, Appendix B

Week 3 (1/29, 1/31): Statistical inference

Wooldridge, Appendix C

Week 4 (2/5, 2/7): The linear regression model, definition and estimation

Wooldridge, Chapter 2

Week 5 (2/12, 2/14): Assumptions and properties

Wooldridge, Chapter 3

Week 6 (2/19, 2/21): Inference and interpretation

Wooldridge, Chapters 4 and 5

Week 7 (2/26, 2/28): Functional form and interactions

Wooldridge, Chapter 6

Week 8 (3/4, 3/6): Qualitative information

Wooldridge, Chapter 7

Wednesday 3/6: MIDTERM

Week 9: Spring Break

Week 10 (3/18, 3/20): Heteroskedasticity

Wooldridge, Chapter 8

Week 11 (3/25, 3/27): Endogeneity and Instrumental Variables

Wooldridge, Chapters 9 and 15

Week 12 (4/1, 4/3): The bias-variance trade-off and generalizability

James et al., Chapters 2, 5, 6

Week 13 (4/8, 4/10): Time series and time series cross-sections

Woolridge, Chapters 10 and 13

Week 14 (4/15, 4/17): Models for categorical dependent variables

Wooldridge, Chapter 17

**Course Policies**

Grade Scale

A standard ten-point A through F grading scale will be employed (i.e. the A-range is 90-100; the B-range is 80-89; and so on).

Attendance

You are expected to attend class, including lab, prepared to engage with the material for that class. Excused absences may be requested in writing with reasonable advance notice (more urgent reasons require less advance notice). Duke policies outline personal emergencies, illnesses, varsity athletic competition, and religious observances as acceptable reasons for an excused absence, but we are willing to consider other reasons that do not neatly fall into one of these categories if given sufficient advance notice.

Deadlines and Late Work

This course moves on a regular, fast-paced schedule. Late work will be penalized (as outlined above) and extensions on assignments will be strongly discouraged in the absence of a compelling reason. This is to incentivize you to stay on schedule, as it will be difficult to catch up if you fall behind.

Professionalism, Academic Honesty, and Collaboration

The Duke community standard is in effect throughout the semester. By taking this course, you affirm that it is a violation of the code to cheat on assignments, to plagiarize, to deviate from the teacher’s instructions about collaboration on work that is submitted for grades, to give false information to a faculty member, and to undertake any other form of academic misconduct. You also affirm that if you witness others violating the code you have a duty to report them.

Beyond the Duke community standard, we expect you to adhere to and maintain norms of professionalism throughout the course. This includes providing us with reasonable advance notice if you need to miss class or move a deadline (better reasons require less advance notice), maintaining a collegial learning environment with your colleagues both inside and outside of the classroom, and taking pride in your work.

Given the nature of this course, some amount of student collaboration is expected and permitted. You are encouraged to form study groups, BUT you must complete quizzes and problem sets on your own. That is, students may work on developing their statistical syntax together by sharing helpful tips, and students are welcome to compare outputs with one another, with the following stipulations: 1) the sharing of ideas must not be one directional, where one students is doing the work and the other is free riding; and 2) the actual write-up of the work that is handed in must be the work of each individual, with absolutely no copying and pasting from one student’s work to another’s. The difference between students learning from each other and one student representing another’s work as their own is usually quite obvious.