Syllabus for Principles of Quantitative Political Research 1
(POLS 4710)

Andrew Gelman
7 Sep 2021

We will go through the first half of the book, *Regression and Other Stories*, by Andrew Gelman, Jennifer Hill, and Aki Vehtari (Cambridge University Press). There is a follow-up course, Principles of Quantitative Political Research 2 (POLS 4712), which covers the second half of the book, including logistic regression, generalized linear models, poststratification, design of studies, and causal inference.

Topics covered in the course include:

- Applied regression: measurement, data visualization, modeling and inference, transformations, and linear regression.
- Simulation, model fitting, and programming in R.
- Key statistical problems include adjusting for differences between sample and population, adjusting for differences between treatment and control groups, extrapolating from past to future, and using observed data to learn about latent constructs of interest.
- We focus on social science applications, including but not limited to: public opinion and voting, economic and social behavior, and policy analysis.

Components of the course:

- *Class meetings* twice a week. Lectures, motivation, going over assignments, discussion.
- *Section meetings with TA:* Answering questions, going over assignments, computer help, there will be one section per week, attendance is optional.
- *Questions and thoughts:* We will have a shared document for the course. Before each class, enter one or two questions or thoughts related to the reading, research, or anything else. You can give a new entry or respond to someone else. We will discuss these in class.
- *Readings:* Readings from the textbook due every week.
- *Homework:* Assignment due every class.
- *Ongoing project:* You will work on this in pairs throughout each semester. The assignments for this project are included in the homework assignments below.
- *Final exam:* Intended to make sure you learned basic skills and concepts.
- *Grading:* 20% class participation (including contributing to the shared document), 40% homework (including the project), and 40% final exam.
- We follow the university's guidelines on academic integrity (http://www.college.columbia.edu/academics/academicintegrity) and accommodation for disabilities (http://www.college.columbia.edu/rightsandresponsibilities).

Organization:

Class periods will mostly be devoted to discussions and statistics activities involving the reading and homework assignments, which are to be completed *before* each class listed below.

The TA will hold weekly meetings and no other regular office hours. One-on-one meetings are possible via prior email. Emails received on the weekend will not be answered on the weekend.

Schedule:

Week 1: Quantitative social science

No reading or homework due before the first class. But you are required to enter something in the shared document, saying something about your expectations for the course. The following is due before the second class:

Reading:
- Chapter 1: Overview

Homework assignments:
- Exercise 1.1

Stories:
- Wikipedia experiment
- Literary Digest poll of 1936

Activities:
- Design a social science experiment to measure some quantity of interest
- Design an experiment to distinguish between two social science hypotheses of interest

Computer demos:
- Simulating statistical data collection and analysis
- Regression predicting election outcome from the economy

Drills:
- Designing a study
- Generalizing

Discussion problems:
- Finding the hidden assumption and error in a naive empirical claim
- Finding the hidden assumption and error in a naive empirical claim

Week 2: Prediction as a unifying theme in statistics and causal inference

Reading:
- Appendix A: Computing in R, sections A.1-A.3

Homework assignments:
- Exercises 1.2, 1.3, 1.4
- Exercises 1.5, 1.6
- In pairs: Exercise 1.10

Stories:
- Economist and Fivethirtyeight election forecasts
- Girls and sports

Activities:
- Candy weighing

- Gathering, plotting, and discussing two measurements of the same underlying quantity from students

Computer demos:

- Graph of data and lines
- Graph a function of 4 variables using a grid

Drills:

- Simple coding: computing and graphing functions
- Simple coding: sampling, looping, and vectors

Discussion problems:

- Evaluating election forecasts
- Variation in social science patterns

Week 3: Data collection and visualization are important.

Reading:

- Chapter 2: Data and measurement

Homework assignments:

- Exercises 1.7, 1.8, 1.9
- Exercises 2.1, 2.3

Stories:

- The Harvard study claiming North Carolina is less democratic than North Korea
- Political leanings of sports fans

Activities:

- Measuring handedness
- Scatterplot charades

Computer demos:

- Graphing the Human Development Index
- Baby Name Voyager and plots of baby names

Drills:

- Criticizing graphs
- Criticizing graphs

Discussion problems:

- Creating a better "electoral integrity index"
- Telling stories with graphs

Week 4: Here's the math you actually need to know.

Reading:

- Chapter 3: Some basic methods in mathematics and probability
- Appendix A: Computing in R, sections A.4-A.5

Homework assignments:

- Exercises 2.6, 2.7
- Exercises 3.1, 3.2, 3.3
- In pairs: Exercise 2.10

Stories:

- Death rate in the pandemic

- Galton was a hero to most

Activities:
- Amoebas and population growth
- Gather data on students in the class: normal and Poisson distributions

Computer demos:
- Matrix manipulations
- Computing weighted averages using vectors and matrices

Drills:
- Straight lines
- Normal distribution

Discussion problems:
- College admissions and weighted averages
- Squares, cubes, and metabolic rates

Week 5: Time to unlearn what you thought you knew about statistics.

Reading:
- Chapter 4: Statistical inference

Homework assignments:
- Exercises 3.5, 3.6, 3.7
- Exercises 4.1, 4.2, 4.3, 4.4

Stories:
- Apparent null effects in a study of heart stents
- The consulting project where they got the wrong standard error

Activities:
- Design a bogus social science study, following the Rolf Zwaan model
- Discuss effects in the context of a social science example

Computer demos:
- Simulating fake data and computing a confidence interval, plus looping
- Bias and unmodeled uncertainty

Drills:
- Binomial distribution (example of basketball shots)
- Sample size and standard errors

Discussion problems:
- Confidence intervals and true parameter values
- Approximate standard error for average "feeling thermometer" ratings

Week 6: You don't understand your model until you can simulate from it.

Reading:
- Chapter 5: Simulation
- Appendix A: Computing in R, sections A.6-A.7

Homework assignments:
- Exercises 4.7, 4.10
- Exercises 5.1, 5.2

Stories:

- The proportion of identical twins in the population
- Voting and coalitions probability example

Activities:

- Real vs. fake coin flips
- Simulating a probability process

Computer demos:

- Simulating a mixed discrete/continuous distribution (income example)
- The mad sd by hand and using the existing R function

Drills:

- Programming in R
- Propagation of uncertainty

Discussion problems:

- Designing a simulation to study potential impacts of selection
- Discuss a real-world process that would be hard to simulate

Week 7: Let's think deeply about regression.

Reading:

- Chapter 6: Background on regression modeling

Homework assignments:

- Exercise 5.3 and another exercise, simulating a comparison with measurement error
- Exercises 6.1, 6.2
- In pairs: Exercise 5.13

Stories:

- What does non-uniform partisan swing look like?
- Clinton/Trump vote vs. polls and predictions

Activities:

- Fake-data simulation and fitting a regression
- Before-after memory tests demonstrating regression to the mean

Computer demos:

- Fake-data simulation and fitting a regression
- Challenges in setting up the regression-to-the-mean simulation

Drills:

- Regression to the mean
- Scatterplots, regression lines, and regression functions

Discussion problems:

- Examples of cases not exhibiting regression to the mean
- Understanding uniform partisan swing (considering regression to the mean)

Week 8: You can't just DO regression, you have to UNDERSTAND regression.

Reading:

- Chapter 7: Linear regression with a single predictor

Homework assignments:

- Exercise simulating regression with measurement error
- Exercises 7.1, 7.2

Stories:

- 5^2 + 12^2 = 13^2
- Interpreting the regression of earnings on height

Activities:

- African countries in the United Nations
- Socioeconomic status and political ideology, studied using a class-designed questionnaire

Computer demos:

- Estimating the mean is the same as regressing on a constant term
- Estimating a difference is the same as regressing on an indicator variable

Drills:

- Sketching a fitted regression model
- Predicting probabilities using regression

Discussion problems:

- Regression line and modeling for future elections given fewer electoral landslides
- Interpreting statistically significant results given huge sample sizes

Week 9: Least squares and all that.

Reading:

- Chapter 8: Fitting regression models

Homework assignments:

- Exercises 7.3, 7.6
- Exercise 8.1
- In pairs: Exercise 7.10

Stories:

- No, Ronald Reagan did not win "overwhelming support from evangelicals"
- When is there undue influence from a single data point?

Activities:

- In pairs, simulate and recover regression lines
- How much do you have to move a point to shift the fitted line by a specified amount?

Computer demos:

- Playing around with the least squares estimate
- Comparing lm and stan_glm

Drills:

- Standard errors and sample size
- Comparisons as regression models

Discussion problems:

- Using the results of a regression to plan a new study
- Doubling sample size and statistical significance

Week 10: Let's be clear about our uncertainty.

Reading:

- Chapter 9: Prediction and Bayesian inference, sections 9.1-9.3 (skip sections 9.4-9.5)

Homework assignments:

- Exercise 8.5

- Exercises 9.1, 9.2
- In pairs: Exercise 8.11

Stories:

- Studying the fairness of random exams
- Disentangling bias and variance in election polls

Activities:

- Coverage of prediction intervals
- Prior distributions for real-world quantities

Computer demos:

- Different forms of predictive uncertainty
- Playing with the Bayesian weighted average formula

Drills:

- Examples of Bayesian combination of information
- Elections: calculating Bayesian posterior mean, standard deviation and probability of success

Discussion problems:

- Forecasting the 2020 elections: state and national predictions
- Forecasting the 2022 and 2024 elections

Week 11: You don't just fit models, you BUILD models.

Reading:

- Chapter 10: Linear regression with multiple predictors

Homework assignments:

- Exercise 9.3
- Exercises 10.1, 10.2
- In pairs: Exercise 10.11

Stories:

- Venn diagrams
- Regression of earnings on height and sex

Activities:

- Fitting multiple regression to data gathered from students
- Designing a study with regression in mind

Computer demos:

- Adding predictors and interactions to a model; comparing regression with interactions to two separate regressions
- Indicator variables

Drills:

- Thinking through predictors and potential problems
- Thinking through interactions

Discussion problems:

- Coming up with an example where regression with interactions gives a much different answer than fitting two separate regressions
- How much is gained by including a pre-test?

Week 12: Can you convince ME to trust YOUR model?

Reading:

- Chapter 11: Assumptions, diagnostics, and model evaluation, sections 11.1-11.7 (skip section 11.8)

Homework assignments:

- Exercises 10.5, 10.6
- Exercises 11.1, 11.2
- In pairs: Exercise 11.11

Stories:

- Actual vs. guessed exam scores
- Bill James does model checking

Activities:

- Compare some survey question for younger and older students in the class, discuss the interpretation of results
- Construct data that violate the regression assumptions, fit the model anyway, see what happens

Computer demos:

- Forming a linear predictor and using it in an interaction model
- Predictive simulation

Drills:

- List the assumptions of regression and give real-world examples of how they can fail
- Explain how regression assumptions can be tested, using real-world examples

Discussion problems:

- Consider the implications of regression assumptions for a real-world study
- Selection bias from conditioning on diagnostics

Week 13: Only fools work on the raw scale.

Reading:

- Chapter 12: Transformations and regression

Homework assignments:

- Exercises 12.1, 12.2
- In pairs: Summarize what you have learned about your example from all the analysis you have done during the semester.

Stories:

- Price elasticity of demand
- Logarithm of world population

Activities:

- Correlation of multiple measurements on students in the class
- Combining predictors to create a total score

Computer demos:

- Centering and standardizing predictors
- Interpreting regression coefficients on the log scale

Drills:

- Examples of exponential and power-law growth and decline
- Estimating the effect of an intervention using the log transformation

Discussion problems:

- Coming up with general rules of when to use the log scale
- The log(1 + x) rule and, more generally, the log(a + x) rule when transforming data with zeroes