**Syllabus for Principles of Quantitative Political Research 2
(POLS 4712)**

Andrew Gelman
18 Jan 2022

We will go through the second half of the book, *Regression and Other Stories*, by Andrew Gelman, Jennifer Hill, and Aki Vehtari (Cambridge University Press). This is a follow-up to the course, Principles of Quantitative Political Research 1 (POLS 4710), which covers the first half of the book, including measurement, data visualization, modeling and inference, transformations, and linear regression.

Topics covered in the course include:
- Applied regression: logistic regression, generalized linear models, poststratification, and design of studies.
- Causal inference from experiments and observational studies using regression, matching, instrumental variables, discontinuity analysis, and other identification strategies.
- 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: Review of basic statistics and regression modeling

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:
- Chapters 1-12 and Appendix A (review)
- Appendix B: 10 quick tips to improve your regression modeling

Homework assignments:
- The *American Political Science Review* recently published an article, "Can Exposure to Celebrities Reduce Prejudice? The Effect of Mohamed Salah on Islamophobic Behaviors and Attitudes," https://doi.org/10.1017/S0003055421000423 and for convenience we posted the article at http://www.stat.columbia.edu/~gelman/regression_course/salah.pdf and its appendix at http://www.stat.columbia.edu/~gelman/regression_course/salah_supp.pdf
  The key claim in the article is that Mohamad Salah joining the Liverpool football club reduced hate crimes in that English city. Frame this question as a linear regression problem, stating clearly what are the data points, the pre-treatment predictors, the treatment variable, and the outcome variable. The actual analysis they do to lead to their Figure 2 is complicated, so just try to set up the basic regression analysis. You can also take a look at Figure A-2 and Section A-4 of the appendix, but it is important here to set up the model from scratch rather than trying to copy what they did. Once you have written the model, discuss challenges of estimating the model from the available data. Consider problems of validity and reliability of measurement as well as the statistical assumptions of the model. Your answer to this assignment should be no more than 1 page long.

Week 2: Modeling probabilities.

Reading:
- Chapter 13: Logistic regression

Homework assignments:
- Exercises 13.1, 13.2, 13.3

Week 3: Logistic regression pro tips.

Reading:

- Chapter 14: Working with logistic regression

Homework assignments:
- Exercises 13.4, 13.5
- Exercises 14.1, 14.6
- In pairs: Exercise 13.13

Week 4: Building models from the inside out.

Reading:
- Chapter 15: Other generalized linear models

Homework assignments:
- Exercise 14.3
- Exercise 15.1, 15.2

Week 5: To understand the past, you must first know the future.

Reading:
- Chapter 16: Design and sample size decisions

Homework assignments:
- Exercise 15.10
- Exercises 16.1, 16.2
- In pairs: Exercise 15.18

Week 6: Enough about your data. Tell me about the population.

Reading:
- Chapter 17: Poststratification and missing-data imputation

Homework assignments:
- Exercises 16.3, 16.4, 16.7
- Exercise 17.1
- In pairs: Exercise 16.14

Week 7: How can flipping a coin help you estimate causal effects?

Reading:
- Chapter 18: Causal inference and randomized experiments

Homework assignments:
- Exercise 17.4
- Exercises 18.1, 18.2
- In pairs: Exercise 17.11

Week 8: Using correlation and assumptions to infer causation.

Reading:
- Chapter 19: Causal inference using regression on the treatment variable

Homework assignments:

- Exercise 18.4, 18.5
- Exercises 19.3
- In pairs: Exercise 18.18

Week 9: Causal inference is just a kind of prediction.

Reading:
- Chapters 18 and 19 again

Homework assignments:
- Exercises 19.4, 19.5
- Exercises 19.6, 19.7

Week 10: Imbalance and lack of complete overlap.

Reading:
- Chapter 20: Observational studies with all confounders assumed to be measured, sections 20.1-20.6 (skip sections 20.7-20.9)

Homework assignments:
- Exercises 19.8, 19.9
- Exercises 20.3, 20.11
- In pairs: Exercise 19.12

Week 11: More assumptions, more problems.

Reading:
- Chapter 21: More advanced topics in causal inference

Homework assignments:
- Exercises 21.1, 21.15, 21.17
- In pairs: Exercise 20.12

Week 12: Who's got next?

Reading:
- Chapter 22: Advanced regression and multilevel models

Homework assignments:
- Exercises 22.1, 22.2
- In pairs: Exercise 21.18

Week 13: All the things you forgot to learn.

Reading:
- Review entire book

Homework assignments:
- In pairs: Summarize what you have learned about your example from all the analysis you have done during the semester.