Readings and assignments for applied regression course

Andrew Gelman, Merlin Heidemanns, Pablo Argote

21 January 2020

We will go through the recently completed book, *Regression and Other Stories*, by Andrew Gelman, Jennifer Hill, and Aki Vehtari. The book has not completed publication yet, so we will hand out photocopied chapters.

Topics covered in the course include:
- Applied regression: data collection, modeling and inference, linear regression, logistic regression, Bayesian inference, and poststratification.
- 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's:* Answering questions, going over assignments, computer help, there will be one section per week, attendance is optional
- *Questions and thoughts:* Before each class, enter one or two questions or thoughts related to the reading, research, or anything else. We will discuss these in class. This will be a shared document so you can give a new entry or respond to someone else.
- *Readings:* Readings from the textbook due every class.
- *Homework:* Homework assignment due every class.
- *Ongoing project:* You will work on this in pairs throughout the semester.
- *Final exam:* Intended to make sure you learned basic skills and concepts.

Organization:

The semester is 14 weeks long, so the classes are listed in order as 1a, 1b, 2a, 2b, etc. 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 TAs will not hold 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:

Class 1a: Design-based inference

No reading or homework due before the first class.

Class 1b: Prediction as a unifying theme in statistics and causal inference

Reading:
- Chapter 1: Overview
- Appendix B: Computing in R, sections B.1-B.3

Homework assignment:
- Exercises 1.1, 1.2, 1.3

Class 2a: Data collection and visualization are important.

Reading:
- Chapter 2: Data and measurement

Homework assignment:
- Exercises 1.7, 1.8, 1.9

Class 2b: Here's the math you actually need to know.

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

Homework assignment:
- Exercise 2.1. Or, if there is no example that you can think of, use the Polity IV Annual Time-Series at https://www.systemicpeace.org/inscrdata.html, which includes a composite index of democracy (democ) whose components are described in page 15 of the codebook (http://www.systemicpeace.org/inscr/p4manualv2018.pdf). Download the data in Excel format, save as .csv, read into R, extract the data from 2018, and then track down the individual components of the measure and use scatterplots to understand how the measure works, as was done for the example of the Human Development Index in the textbook.
- Exercise 2.7

Class 3a: Time to unlearn what you thought you knew about statistics.

Reading:
- Chapter 4: Generative models and statistical inference

Homework assignment:
- Exercises 3.1, 3.3, 3.5

Class 3b: You don't understand your model until you can simulate from it.

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

Homework assignment:
- Exercises 4.1, 4.2, 4.3, 4.4
- Step 1 on your project: Plot data and discuss measurement issues.

Class 4a: Let's think deeply about regression.

Reading:
- Chapter 6: Background on regression modeling

Homework assignment:
- Exercises 5.1, 5.2, 5.3

Class 4b: You can't just do regression, you have to understand regression.

Reading:
- Chapter 7: Linear regression with a single predictor

Homework assignment:
- Exercises 6.1, 6.4
- Step 2 on your project: Construct a probability model and simulate fake data.

Class 5a: Least squares and all that

Reading:
- Chapter 8: Fitting regression models

Homework assignment:
- Exercises 7.2, 7.3, 7.6

Class 5b: Let's be honest about our uncertainty.

Reading:
- Chapter 9: Prediction and Bayesian inference, sections 9.1-9.3

Homework assignment:
- Exercises 8.1, 8.5
- Step 3 on your project: Fit and display a linear regression with a single predictor.

Class 6a: When do priors matter?

Reading:
- Chapter 9: Prediction and Bayesian inference, sections 9.4-9.6

Homework assignment:
- Exercises 9.4, 9.6

Class 6b: You don't just fit models, you BUILD models.

Reading:
- Chapter 10: Linear regression with multiple predictors

Homework assignment:
- Exercises 9.1, 9.7
- Step 4 on your project: Perform predictive simulation from a fitted regression model.

Class 7a: We're assuming WHAT?

Reading:
- Chapter 11: Assumptions, diagnostics, and model evaluation, sections 11.1-11.5

Homework assignment:
- Exercises 10.1, 10.2

Class 7b: Can you convince ME to trust YOUR model?

Reading:
- Chapter 11: Assumptions, diagnostics, and model evaluation, sections 11.6-11.9

Homework assignment:
- Exercises 11.1, 11.2, 11.3
- Step 5 on your project: Fit and display a linear regression with multiple predictors.

Class 8a: Only fools work on the raw scale.

Reading:
- Chapter 12: Transformations and regression

Homework assignment:
- Exercises 11.4, 11.5

Class 8b: What makes a master modeler a master modeler?

Reading:
- Review chapters 1 through 12

Homework assignment:
- Exercises 10.4, 10.8

- Step 6 on your project: Check the fit of your model and discuss the assumptions needed to use it to make real-world inferences.

Class 9a: Thinking like a quantitative social scientist

Reading:
- One or more research articles in political science

Homework assignment:
- Evaluate and criticize an article based on what you have learned so far in this course.

Class 9b: Modeling probabilities

Reading:
- Chapter 13: Logistic regression

Homework assignment:
- Exercises 12.1, 12.2
- Step 7 on your project: Think like a social scientist.

Class 10a: Logistic regression pro tips

Reading:
- Chapter 14: Working with logistic regression
- We will skip chapter 15, but you can read it and ask questions about it in class if you'd like.

Homework assignment:
- Exercises 13.1, 13.2, 13.3

Class 10b: To understand the past, you must first know the future

Reading:
- Chapter 16: Design and sample size decisions

Homework assignment:
- Exercises 14.1, 14.2, 14.6
- Step 8 on your project: Fit a logistic regression.

Class 11a: Enough about your data. Tell me about the population.

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

Homework assignment:
- Exercises 16.1, 16.2, 16.4
- (no exercises on chapter 15)

Class 11b: How can adding noise by flipping a coin help you learn?

Reading:
- Chapter 18: Causal inference and randomized experiments

Homework assignment:
- Exercises 17.1, 17.4, 18.1
- Step 9 on your project: Perform a design analysis, addressing issues of measurement, precision, and sample size.

Class 12a: When does correlation actually imply causation?

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

Homework assignment:
- Exercises 19.4, 19.5

Class 12b: Causal inference is just a kind of prediction.

Reading:
- Chapters 18 and 19 again

Homework assignment:
- Exercises 19.6, 19.7
- Step 10 on your project: Frame your problem in terms of causal inference.

Class 13a: Imbalance and lack of complete overlap

Reading:
- Chapter 20: Observational studies with all confounders assumed to be measured, sections 20.1-20.6
- We will skip sections 20.7-20.10, but you can read them and ask questions about them in class if you'd like.

Homework assignment:
- Exercises 19.9, 20.11

Class 13b: More assumptions, more problems

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

Homework assignment:
- Exercises 21.8, 21.15, 21.16
- Step 11 on your project: Perform a causal regression.

Class 14a: What's next?

Reading:
- Chapter 22: Advanced regression and multilevel models

Homework assignment:
- Exercises 22.1, 22.2

Class 14b: All the things you forgot to learn

Reading:
- Review entire book

Homework assignment:
- Final step on your project: Summarize what you have learned.