Advanced Regression Modeling

Andrew Gelman

25 May 2018

1. Overview

- We will go through two forthcoming books:

  - *Regression and Other Stories*, by Andrew Gelman, Jennifer Hill, and Aki Vehtari
  - *Multilevel Models and Advanced Regression*, by Andrew Gelman, Jennifer Hill, Ben Goodrich, Jonah Gabry, and Aki Vehtari

  These two book represent a revision and expansion of the book, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, by Andrew Gelman and Jennifer Hill.

- Topics covered in the course include data collection, modeling and inference, using the following tools: simulation, Bayesian inference, linear regression, logistic regression, generalized linear models, nonlinear and nonparametric regression, and multilevel models.

- You will program and fit models in R and Stan.

- 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, item response and ideal point modeling, economic and social behavior, effects of interventions, and policy analysis.

2. Student responsibilities

- **Three or four times a week**, you will write an entry in your **statistics diary**; see here: http://andrewgelman.com/2015/01/07/2015-statistics-diary/). Just set up a text or Word file and add to it every other day. The diary entries can be anything. They can be short slice-of-life observations ("Looking at faces on the subway this morning. Is it really true that people are less happy on Monday? How to measure this in a survey?"), quick questions ("Attitudes toward recreational drugs seem more permissive than in the past? Is this a real trend? If so, is it recent or has it been gradually happening for decades?"), research notes ("I'm comparing attitudes about military intervention in several European countries. Do I have to be concerned about question-wording effects in different languages?"), or things you're working on, difficult problems that you might be stuck on, or have an insight about. You can write as little or as much as you want each time. The only requirement is that you write something new in it, every other day. You're *not* allowed to go back a week later and fill in 3 entries at once. That would be cheating. Do it three or four times a week. Just type it in to the file.

- **Each week**, you will have two **homework assignments**. Each homework assignment needs to be uploaded to Courseworks.

- **Before every class**, you will have **readings** from the assigned textbooks.

- **Before every class**, you will have a **jitt** (just-in-time teaching assignment). Each of your jitts will be a set of three quick online items, separate from the main homework assignments, that are a mix of questions on the required reading, short exercises to get you ready for the upcoming class discussion, and feedback.

- **Each class** will involve your **active participation** in class discussion. Also, **bring your laptop computer** to class as we will be doing activities together in R.

- **At the end of the semester**, you will have a final exam.

3. Structure of course

We will cover the following topics:

1. Principles of data collection, measurement, and visualization

2. Review of basic mathematics and statistics

3. Programming and simulation

4. Linear regression

5. Prediction and Bayesian inference

6. Logistic regression and generalized linear models

7. Building, using, evaluating, and comparing models

8. Causal inference

9. Models for missing, incomplete, and structured data

10. Regularization for models with many predictors

11. Nonlinear models, nonparametric regression and machine learning

12. Multilevel/hierarchical models

13. The role of statistics in social science