Multilevel Regression Modeling for Public Policy

PPHA 41420

Logistics

- Prerequisites: Linear regression and R (e.g. PPHA 31102)
- Lecture: TuTh 2:00-3:20pm in Keller 0007
- Instructor: Eric Potash (epotash@uchicago.edu)
- Office Hours: TuW 3:30-5:30pm in Keller 3105 or by appointment
- Grader: David McMillon (mcmillon@uchicago.edu)

Summary

Grouped data, such as students within schools or workers within firms, are ubiquitous in public policy. Both to satisfy the assumptions of regression and to build realistic models that provide realistic inferences we should include group-level intercepts and slopes in our models. Traditionally this was accomplished using fixed effects and their interactions with covariates. However, as we commonly have few observations per group, this approach can yield noisy or degenerate estimates.

We will introduce a Bayesian perspective on regression modeling and use it to develop multilevel regression models (also known as hierarchical or mixed-effects models). Under certain assumptions, these models allow us to partially pool information across groups in order to efficiently model the group structure even when the number of observations within each group is small. Recent advances in computing have made the estimation of multilevel models much more practical.

Drawing on examples from the fields of epidemiology, education, and political science, we will study applications of multilevel models to heterogeneous treatment effects, small area estimation, longitudinal data, and prediction. Familiarity with R and linear regression are assumed.

Books and Resources

We will follow the text *Data Analysis Using Regression and Multilevel/Hierarchical Models* by Andrew Gelman and Jennifer Hill. Some other texts that I may borrow from and may be of interest to you include *Statistical Rethinking* by Richard McElreath, *Geospatial Health Data* by Paula Moraga, and *Doing Bayesian Data Analysis* by John Kruschke.

Grading

Assignments (50%)

Assignments will be shared at the end of class on Thursday covering that week's material and will be due the following Thursday before class (2:00pm) via Canvas. Late assignments will be penalized by 10% per day, with 1 day starting at 2:01pm. Solutions will be posted on Mondays so no credit will be given for assignments submitted after that. There will be 6 assignments total. Your lowest assignment score will be dropped. Assignments should be completed using RMarkdown and rendered to PDF with code blocks visible.

Exams (50%)

There will be two exams: a midterm (15%) and a final (35%).

Academic Dishonesty

Writing code is substantially different from writing essays: it is standard practice to google things that don't work, and copy a line or two from a manual or Stack Overflow. I encourage you to discuss general strategies for solving problems with your classmates and friends. Questions and answers on the discussion board will naturally include code. However, you should never ask to see a classmate's solutions nor copy code from your classmates. No one but you should type your code.

All students suspected of academic dishonesty will be reported to the Harris Dean of Students for investigation and adjudication. The disciplinary process can result in sanctions up to and including suspension or expulsion from the University. In addition to disciplinary sanctions, I will impose a grade penalty of 0 for students who have committed academic dishonesty. See the Harris and University of Chicago academic dishonesty policies.

ADA Accommodations

Any student who believes they may need assistance should inform the Harris Dean of Students office by the end of the first week of class. The Dean of Students office will coordinate any student accommodations with Harris instructors.

Software

In this course we will be using the R language. You will need to install two pieces of software. First install the R programming language itself. Next install RStudio. This is what's called an *integrated development environment*. It makes using R programming easier and more productive. Additionally you will need the rstanarm package which can be installed using the following command: `install.packages('rstanarm')`. Assignments should be submitted using RMarkdown. In RStudio create a new RMarkdown PDF document (`File > New File > R Markdown` select output format `PDF`). Save the file and then click the "Knit" button (or `Ctrl+Shift+K`) to render it to PDF, installing the necessary software.

Syllabus

This syllabus is tentative and will be updated throughout the quarter. Lecture slides, readings, and assignments will be posted each week on Canvas.

1. Introduction to multilevel modeling

Why multilevel modeling? Examples and challenges

2. Linear regression and inference

- Assignment 1 posted

3. Prediction

4. Regression to the mean; Interpretation

- Assignment 1 due, Assignment 2 posted

5. Multiple predictors; Assumptions, Diagnostics, and Evaluation

6. Multilevel Model Basics

- Assignment 2 due, Assignment 3 posted

7. Varying Intercept Models

8. More on Varying Intercepts

- Assignment 3 due, Assignment 4 posted

9. Cancelled

10. Simulation; Varying Slopes

- Assignment 4 due

11. Review

12. Midterm

13. Bias and variance of multilevel estimates; Leave-one-out cross validation

14. More multilevel structures

- Assignment 5 posted

15. Cross Validation; Meta-analysis

- Reading for next lecture: *Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments* by Rachael Meager

16. Meta-analysis; Correlated predictors and group-level effects

- Assignment 5 due, Assignment 6 posted

17. Multi-level logistic regression

18. Post-stratification; Repeated measures

- Assignment 6 due

19. Sample size and power calculations

Final exam is take-home, assigned 12pm Tuesday March 17 and due 12pm Wednesday March 18. I will be available during the scheduled final exam time (Tuesday March 17 3:30 to 6:20pm in Keller 0007) to discuss.