Topics in Political Science:
Multilevel Models
PLAD 8500, University of Virginia
Spring 2017

Instructor: Constanza F. Schibber

Time and Location: Thursday 4-6:30 PM, Gibson Hall 287

Contact: constanza@virginia.edu

Office Hours: Thursday 2:30-3:30, Gibson Hall 266

Overview

This course covers statistical modeling with explicitly defined hierarchies. Social scientists encounter multilevel data all the time: voters clustered in electoral districts, students nested within classrooms, legislators clustered in congressional periods, countries nested within regions, and so forth. Classic time-series cross-sectional (TSCS) data can also be thought as multilevel data, with observations clustered by unit and time period. Even in survey research, multilevel models are used to estimate public opinion across geographic units from individual-level survey data (commonly known as MRP). The course will begin with a review of linear regression, logistic regression, and generalized linear models. Then it will proceed to multilevel nested models and follow with non-nested models for linear and generalized linear models. Hierarchical modeling can incorporate individual-level predictors, group-level predictors, and individual-by-group (also known as cross-level) interactions. The second half of the course will feature a Bayesian perspective on inference and computation of hierarchical models.

Required Text

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Additional Readings

Gill, Jeff. 2014. *Bayesian Methods : A Social and Behavioral Sciences Approach*. Chapman & Hall/CRC Statistics in the Social and Behavioral Sciences, Third Edition. (Available to read online at the UVa Library.)

Published articles will be available in the Collab website. See the schedule for a reading list.

Grading

Your grade will be structured as follows:

- Participation & Attendance: 5%
- In-Class Lab Sessions: 10%
- In-Class Discussion: 15%
- Assignments: 35%
- Research Paper: 20%
- Presentation of the Research Paper: 15%

Late assignments will not be accepted and no incomplete will be given for assignments or the course. Exceptions will be granted only under truly extraordinary circumstances.

The procedure to have any grade revised is as follows. Please write up a short description of your argument as to why your grade should be changed and hand it in, along with your initial assignment, within one week of receiving your grade. The instructor will respond in writing. The instructor's decisions regarding grades are final.

No adjustments will be made to final grades under any circumstances. Students will have the opportunity to earn extra credit over the course of the semester to provide an extra cushion in case of any unforeseeable problem.

Evaluation

Participation & Attendance: I expect students to attend all lectures and to arrive to class on time. Laptops are only allowed during in-class exercises. Forms of participation may include asking questions, answering questions from the instructor or from other classmates, actively participating in in-class group activities and class discussion, among others. Using the course email list to ask and answer questions is strongly encouraged and it will contribute towards your participation evaluation.

In-Class Lab Sessions: During the last 30 minutes of most classes there will be a "hands-on" lab in which each student will work on running and understanding the `R` and/or `Bugs/Jags` code presented during the lecture. The instructor will be available to help students complete the task. Students are allowed (and encouraged) to collaborate. At the end of the lab session, each student will fill in a short Quizz through Collab assessing their own work.

In-Class Discussion: On some assigned dates, we will have in-class discussions. Students have to carefully read the assigned papers and come prepared to class. Everyone has to participate in the discussion. Each student will prepare one question or topic for discussion based on the readings.

Assignments: Most weeks there will be an assignment which will consist of a combination of analytical problems and data analyses. Assignments should be written in a professional fashion and include the `R` code used to address specific problems. I recommend preparing the assignments using `RStudio` and the `R` library `knitr` (instructions will be provided separately), because it will be more efficient for you. Unless otherwise noted, all assignments should be completed by the time of class, 4 PM, and uploaded to the Collab website.

You are encouraged to work together with your fellow students and use the course email list to ask and answer questions. *But do not copy answers from another student, or allow your answers to be copied, or look for and copy solutions to the assignments on the internet.* Copying is cheating and will be referred to the Honor Committee.

Research Paper: The final assignment is writing a paper that applies a multilevel model to data in your field of study. The end product should look like the statistical and empirical sections of a paper published in a journal (10-12 pages, double-spaced). Along with your research paper, you will submit replication material for your statistical model and results in `R`.

There could be four types of research papers: (1) The paper is the start of a research manuscript that will eventually turn into a student's thesis, dissertation, or published work; (2) A student already has a manuscript that could be improved by reanalyzing the data with a hierarchical model; (3) The paper reanalyzes the empirics of a paper published in a leading journal that could (or should) have used a multilevel model but did not; (4) The paper reanalyzes the empirics of a paper published in a leading journal that uses a frequentist multilevel model; the student replicates the model and develops a Bayesian specification. *First-year graduate students are strongly encouraged to pursue (3) or (4).*

Students will submit a research paper proposal by March 1, 2017.[1] Students are strongly encouraged to ask the instructor for feedback on their proposal before submission. Even though the proposal is not graded, it will be approved, approved with revisions, or rejected. If the proposal is approved with revisions, the student will submit a written response on how to address the revisions by March 15th, 2017. If a proposal is rejected, the student will submit a new proposal by March 15th, 2017.

All students should meet with the instructor *at least once* to discuss their project before it is due on Friday, May 12, 2017 at noon.

Presentation of the Research Paper: Each student will create a poster summarizing their research paper. A poster session will be held jointly with *PLAD 7500 Time Series* on Friday 28, 2017, between 11:30 AM and 1 PM (date & time could be subject to change). Faculty and students will be invited and there will be food and beverages. Presenters should stay near their posters to take questions and comments and explain their findings to attendees.

A draft of the poster should be presented to the instructor on (or before) April 20, 2017. Take into consideration that posters should go into print at least 2 or 3 days before the poster session. The poster can be landscape or portrait, but no larger than 36 x 48 inches. The following provides helpful advice about structuring and organizing a good poster:

[1] Instructions will be uploaded to the Collab website.

Designing Effective Posters, Jeff Radel at the University of Kansas

There are a variety of software packages that can be used to design posters including Microsoft Power Point, LaTeX, and Adobe Illustrator.

Installing R

All students will need to download and install the latest R software. R is a free statistical programming language that we will use to fit models, simulation, computing probabilities, creating graphics, *etc.*. It may be obtained at the CRAN website. Go to http://lib.stat.cmu.edu/R/CRAN and click your choice of platform (Linux, MacOS X or Windows) for the precompiled binary distribution. Note the FAQs link to the left for additional information.

You will use R to complete assignments.

The University of Virginia Library has a number of workshops on R such as "Introduction to R", "Character Manipulation in R", "Introduction to R Markdown", "Advanced Visualization in R", among others.

Installing JAGS

For Bayesian modeling we will be using *Just Another Gibbs Sampler* (JAGS). JAGS is not part of R, it's a stand-alone application, but we will command it from R. Windows and OS X users download the installers at: http://goo.gl/tbw7Lt. Then, install a few extra R packages by typing (in R): `install.packages(c("R2jags", "rjags", "coda"))`.

Schedule

Thursday	
Jan 19th 1. **Introduction to the Course. Why Multilevel Models?**	**1**
26th 2. **Linear and Generalized Linear Models Review** Reading: Gelman and Hill, Chapters 1, 2, 3, & 4	**2**

Thursday	
Feb 2nd 3. **Simulation of probability models and statistical inferences** Reading: (1) Gelman and Hill, Chapters 7 & 8; (2) Gill, Chapter 9 4. **Hypothesis Testing** Discussion: - Gill, Jeff. 1999. The Insignificance of Null Hypothesis Significance Testing, *Political Research Quarterly* 52(3):647-674. - Gelman, Andrew and Hal Stern. 2012. The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant, *The American Statistician* 60(4):328-331. - Nuzzo, Regina. 2014. Statistical Errors. P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume, *Nature* 506:150-152. - Editorial. 2016. The ASA's Statement on p-Values: Context, Process, and Purpose, *The American Statistician* 70(2):129-133. **Homework 0, complete by this date**	**3**
9th 5. **Multilevel Linear Models: The Basics** Reading: Gelman and Hill, Chapters 11 & 12	**4**
16th 6. **Multilevel Linear Models: Varying Slopes, Non-nested Models, and Other Complexities** Reading: Gelman and Hill, Chapter 13 **Homework 1 Due**	**5**
23rd 7. **Multilevel Logistic Regression** Reading: Gelman and Hill, Chapter 14 **Homework 2 Due**	**6**

Thursday	
Mar 2nd 8. **Multilevel Generalized Linear Models** Reading: Gelman and Hill, Chapter 15 **Final Paper Proposal Due** Discussion: Gelman, Andrew, Jeffrey Fagan, and Alex Kiss. 2012. An Analysis of the New York City Police Department's "Stop-and-Frisk" Policy in the Context of Claims of Racial Bias. *Journal of the American Statistical Association*, 102(479): 813-823.	**7**
9th **Spring Break**	
16th 9. **Comparing Frequentist and Bayesian Inference** 10. **An Introduction to Bayesian Statistics** Discussion: Jeff Gill, Chapters 1, 2, & 10 **Homework 3 Due**	**8**
23rd 11. **Bayesian Inference & Multilevel Modeling in `Bugs/JAGS` and `R`.** Reading: (1) Gelman and Hill, Chapters 16; (2) Gill, Chapter 11 Discussion: Jeff Gill, Chapters 5 & 7 **Homework 4 Due**	**9**
30th TBD	**10**
Apr 6th 12. **Bayesian Multilevel Linear Models** Reading: (1) Gelman and Hill, Chapter 17, (2) Gill, Chapter 15 **Homework 5 Due**	**11**

Thursday	
13th 13. **Bayesian Multilevel Generalized Linear Models** 14. **Bayesian inference and computation** Reading: (1) Gelman and Hill, Chapters 17, 18 and 19 (2) Gelman, Andrew Jennifer Hill, and Masanao Yajima. 2012. Why we (usually) don't have to worry about multiple comparisons, *Journal of Research on Educational Effectiveness,* 5:189-2011. Discussion: Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman. 2015. Forecasting elections with non-representative polls *International Journal of Forecasting*, 31(3): 980-991. **Homework 6 Due**	**12**
20th 15. **Understanding and summarizing the fitted models** Reading: Gelman and Hill, and Chapter 21 Discussion: - *"How many people do you know in prison?": Using overdispersion in count data to estimate social structure in networks*, Tian Zheng, Matthew J. Salganik, and Andrew Gelman **Homework 7 Due** **Draft Poster Due**	**13**
27th 15. **Analysis of variance** 16. **Model checking and comparison** Reading: Gelman and Hill, Chapters 22 and 24	**14**
May 4th Exam week	**15**
11th Exam week **Final Paper Due**	**16**