Multilevel/Hierarchical Modeling

PLS 900, Michigan State University

Spring 2018

Instructor: Constanza F. Schibber

Time and Location: Monday 3-5:50 PM, South Kedzie Hall 104

Office Hours: TDB

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. More than 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. (Selected chapters.)

See the schedule for a additional readings and the reading list.

Grading

Your grade will be structured as follows:

- Participation & Attendance: 5%
- In-Class Lab Sessions: 10%
- In-Class Discussion: 15%
- Assignments: 30%
- Research Paper: 25%
- 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, 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 50 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 `JAGS` code presented during the lecture. Students will also have to complete a few exercises. The instructor will be available to help students complete the tasks. Students are allowed (and encouraged) to collaborate. At the end of the lab session, each student will fill in a short Quizz 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: There will be three assignments 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. The assignments have to be prepared using `RStudio` and the `R` library `knitr` (instructions will be provided separately). Unless otherwise noted, all assignments should be completed by the time of class, 3 PM, and uploaded to the class 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 three 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.

Students will submit a research paper proposal by February 26, 2018.[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 either 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 February 30th, 2018. If a proposal is rejected, the student will submit a new proposal by February 30, 2018.

Students will submit a brief update on their project on March 12, 2018.[2]

The final project is due on Wednesday May 2, 2018, 5 PM. Students should plan to meet with the instructor *at least once* to discuss their project after turning in an update on their project (3/12) and their final project (5/2).

Presentation of the Research Paper: Each student will create a poster summarizing their research paper. A poster session will be held jointly with *PLS 900, R Programming* on a date TBD. 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 16, 2017. The instructor will prove feedback on the poster and students will have to provide a written response to each comment along with a new draft of the poster on (or before) April 21, 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.

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

[2] Instructions will be uploaded to the course website.

The following provides helpful advice about structuring and organizing a good poster: 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.

Academic Honesty & Integrity

We will abide by MSU's statements and policies regarding academic honesty, as detailed at the ombudsman's website at https://ombud.msu.edu/academic-integrity/index.html. No cheating will be tolerated, and this includes handing in a homework that is the same as someone else's homework or using computer software to automatically calculate solutions to an assignment that you are supposed to solve yourself without any assistance.

Installing `R`, `RStudio`, and `knitr`

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 should also download `RStudio` and `knitr`. You will have to use them to submit your assignments and final project.

Installing `JAGS`

For Bayesian modeling we will be using *Just Another Gibbs Sampler* (`JAGS`). `JAGS` is not part of R, it is 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

Monday	
Jan 8th 1. **Introduction to the Course. Why Multilevel Models?** 2. **Linear and Generalized Linear Models Review** Reading: Gelman and Hill, Chapters 2, 3, 4, 5, 6, & 7	**1**
15th **Martin Luther King Day**	
22nd 3. **Multilevel Linear Models: The Basics** Reading: (1) Gelman and Hill, Chapters 11 & 12.	**2**
29th 4. **Multilevel Linear Models: Varying Slopes, Non-nested Models, and Other Complexities** Reading: (1) Gelman and Hill, Chapter 13 (2) Western, Bruce. 1998. Causal Heterogeneity in Comparative Research: A Bayesian Hierarchical Modelling Approach, *American Journal of Political Science* 42:1233-1259.	**3**
Feb 5th 5. **Multilevel Logistic Regression** 6. **Multilevel Generalized Linear Models** Reading: Gelman and Hill, Chapters 14 & 15 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. **Homework 1 Due**	**4**

Monday
12th **5**

7. **Visualization and statistical inferences**
   Readings: (1) Gelman and Hill, Chapters 7 & 8; (2) Gill, Chapter 9; (3) Hanmer, Michael J. and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57(1):263- 277

8. **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.
- Benjamin, Daniel J., James O. Berger, Magnus Johannesson, Brian A. Nosek, E. J. Wagenmakers, Richard Berk, Kenneth A. Bollen, et al. 2017. "Redefine Statistical Significance." *Nature Human Behavior* 1-18.

19th **6**

9. **Frequentist Multilevel Models**

Discussion 1: The Fixed Effect-Random Effect Debate
- Beck, Nathaniel and Jonathan N. Katz. 2007. Random coefficient models for time-series cross-section data, *Political Analysis* 15(2): 182-195.
- Bell, Andrew and Kelvyn Jones. 2014. Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data, *Political Science Research and Methods* 3(1): 133-153
- Fairbrother, Malcom. 2014. Two Multilevel Modeling Techniques for Analyzing Comparative Longitudinal Survey Datasets, *Political Science Research and Methods* 2(1), 119-140.
- Clark, T. and Linzer, D. 2015. Should I Use Fixed or Random Effects?, *Political Science Research and Methods* 3(2), 399-408.
- Alexander W. Schmidt-Catran, Malcolm Fairbrother. 2016. The Random Effects in Multilevel Models: Getting Them Wrong and Getting Them Right, *European Sociological Review* 2(1): 23-38.

Discussion 2:
TBD (Discussion of selected papers based on students' interests)

Monday	
26th	**7**
10. **An Introduction to Bayesian Statistics** Reading/Discussion: - Jeff Gill, selected chapters. - Daniel Stegmueller, "How many countries for multilevel modeling? A comparison of Frequentist and Bayesian approaches", *American Journal of Political Science*, 101(474): 409-423. **Project Proposal Due**	
Mar 5th **Spring Break**	
12th	**8**
11. **Bayesian Inference & Multilevel Modeling in** `Bugs/JAGS` 12. **Bayesian Multilevel Linear Models** Reading: (1) Gelman and Hill, Chapters 16, 17, & Appendix C, (2) Gill, Chapter 15 **Homework 2 Due** **Project Update Due**	
19th	**9**
13. **Bayesian Multilevel Generalized Linear Models** 14. **Bayesian inference and computation** Reading: (1) Gelman and Hill, Chapters 17, 18, & 19, (2) Gill, Chapter 12. 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.	

Monday	
26th	**10**

15. **Understanding and summarizing the fitted models**

16. **Model checking and comparison**

Reading: (1) Gelman and Hill, and Chapters 21 & 24, (2) Gill, Chapter 7.

Discussion: Bayesian Model Comparison.
- George, Edward I. and McCulloch, Robert E. 1993. Variable Selection via Gibbs Sampling, *Journal of the American Statistical Association* 88(423): 881-889
- Quinn, Kevin M., Andrew D. Martin and Andrew B. Whitford. 1999.Voter Choice in Multi-Party Democracies: A Test of Competing Theories and Models, *American Journal of Political Science* 43(4): 1231-1247.
- Kruschke, John K. 2011. Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison, *Perspectives on Psychological Science* 6(3):299-312.

Homework 3 Due

Apr 2nd	**11**

17. **Analysis of variance**

Reading: Gelman and Hill, Chapters 22

Discussion: On Priors
- Gelman, Andrew. 2006. Prior distributions for Variance Parameters in Hierarchical Models (Comment on article by Browne and Draper). *Bayesian Analysis* 1(3):515-534.
- Gill, Jeff and Walker, Lee D. 2005. Elicited priors for Bayesian model specifications in political science research, *Political Analysis* 67(3): 841-872.
- Traunmller, Richard and Murr, Andreas and Gill, Jeff. 2015. Modeling latent information in voting data with dirichlet process priors, *Political Analysis* 23(1):1-20

9th	**12**

18. **Missing-Data Imputation**

Reading: Gelman and Hill, Chapters 25
Additional readings TBD

Monday	
16th	**13**

19. **Time**

Discussion:
- Brandt, Patrick T and Freeman, John R. 2006. Advances in Bayesian time series modeling and the study of politics: Theory testing, forecasting, and policy analysis, *Political Analysis* 14(1): 1-36.
- Pang, Xun. 2010. Modeling heterogeneity and serial correlation in binary time-series cross-sectional data: A Bayesian multilevel model with AR(p) errors. *Political Analysis* 18(4); 470-498.
- Pang, Xun. 2014. Varying Responses to Common Shocks and Complex Cross-Sectional Dependence: Dynamic Multilevel Modeling with Multifactor Error Structures for Time-Series Cross-Sectional Data, *Political Analysis* 22(4); 464-496.
- Cranmer, Skylar J., Rice, Douglas R., and Siverson, Randolph M. 2017. What To Do About Atheoretic Lags. *Political Science Research and Methods*, 5(4), 641-665.

Draft Poster Due

23rd	**14**

20. **TBD**