ICPSR Summer Program | dates | time | location

Multilevel Models II: Advanced Topics

John Poe, University of Michigan
Office:
Office Hours:
jdpoe@umich.edu

Course Description

This course is designed to provide participants with the practical skills and workflow to deal with almost any kind of multilevel modeling problem no matter how complicated.

We begin with a review of the basic toolkit taught in standard multilevel, longitudinal, and panel data courses. The same set of models introduced here will be referenced repeatedly throughout the course so don't worry if you aren't an expert on the first day. After the general overview of tools, we spend the rest of the first week on the key theme of the class: that missing data, missing variables, and missing dependency structures can all lead to biased inferences. We have a variety of practical strategies to deal with each kind of missingness that we will explore throughout the class but it is important to understand that we have to deal with them all consistently and systematically.

In the second week, we step back and focus on practical implementation and understanding how to do just that. We begin with a general analysis workflow for how to think about dependency structures, omitted variable bias, and missing data in a practical sense. We then walk through debugging, diagnostics, and actually getting code to work from both the MLE and Bayesian perspective. We spend a day on model comparison and how to tell if you actually need a multilevel model or if a simpler approach will work. Finally, we go through how to explain and interpret these models.

In the last two weeks we put theory and coding skills into practice. Each day we will spend an hour thinking through the logic of a complicated multilevel modeling example.

We will build an analysis plan based on a checklist built from weeks 1 and 2. We will then work through how to actually put that logic into practice in software. This includes building simulated data, fitting initial models, interpreting error messages, diagnostics, and practical interpretation. Each of these examples have been chosen specifically to illustrate some important kind of problem or set of problems that you may run into in real life. That means the examples are relevant even if the subject-matter is unrelated to what you do.

Advanced Track

This course is designed to provide participants with a grounded and applied skill set that they can use to run hard models on real data. The goal is to learn how to actually sit down and solve advanced problems. The focus on practicality comes at a cost of some theoretical depth that can be frustrating to more advanced participants. The advanced track is an attempt to solve this trade-off and get into the weeds. Participants have the option of additional lectures about more abstract or theoretical extensions of a days topic. These lectures will range from discussions on Bayesian hierarchical models to more theoretical discussions of latent dependency structures to modern advanced causal inference and will come with additional notes and accompanying reading lists. These lectures occur right after the regular class and are completely optional. You don't need to do anything except stay after class.

Participants who do not feel that they are ready for the more advanced discussion or are simply more interested in other things will still gain an enormous amount of benefit from the applied core of the class. Participants who are more advanced and have a greater interest in methods as a field should find these added lectures invaluable.

Readings

Each lecture from the first week will have accompanying course notes that should be referenced throughout the workshop. Supplemental readings will be included for individual lectures. We will also have a working directory of online tutorials, blog posts, help files, and demos that we will reference constantly. I will provide an extended reading list for your future use should you want to become an expert on some specific aspect of multilevel modeling. You should skim it while you're here so that you can talk to us about particular readings.

In general, you should probably buy at least some of these books for your library. I recommend Hox (2017) and/or Snijders and Bosker (2011) for basic references. Most people in the social sciences recommend Gelman and Hill (2007) which is nice but dated and a new edition is supposedly in the pipeline. For advanced references, the three handbooks of multilevel modeling: De Leeuw, Meijer, and Goldstein (2008) Hox and Roberts (2011) and Scott, Simonoff, and Marx (2013) are all terrific. For broader (accessible) theories of how multilevel models fit into broader statistics, I *strongly* recommend Skrondal and Rabe-Hesketh (2004) and Hodges (2013). All of these books are available in the ICPSR library and I strongly advise you to go look at them to help you figure out what you want to buy.

Recommended Multilevel Books (Introductory/Intermediate):

- Gelman, Andrew, and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*: Cambridge University Press.
- Hox, Joop, Mirjam Moerbeek, Rens van de Schoot. 2017. *Multilevel analysis: Techniques and applications*. Routledge
- Snijders, Tom AB and Bosker RJ. 2011. *Multilevel analysis*. Springer
- Goldstein, Harvey. 2011. *Multilevel statistical models*. John Wiley & Sons.
- Stroup, W. W. 2012. *Generalized linear mixed models: modern concepts, methods and applications*, CRC press.

Recommended Multilevel Books (Advanced/Specialized):

- Skrondal, A. and S. Rabe-Hesketh. 2004. *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*, CRC Press.
- De Leeuw, Jan, Erik Meijer, and Harvey Goldstein. 2008. *Handbook of multilevel analysis*. New York: Springer.
- Wu, Lang. 2009. *Mixed effects models for complex data*: CRC Press.
- Congdon, Peter D. 2010. *Applied Bayesian hierarchical methods*: CRC Press.
- Hox, Joop, and J. Kyle Roberts. 2017. *Handbook of advanced multilevel analysis*. Psychology Press
- Fitzmaurice, Garrett M, Nan M Laird, and James H Ware. 2012. *Applied longitudinal analysis*. John Wiley & Sons.
- Scott, Marc A, Jeffrey S Simonoff, and Brian D Marx. 2013. *The SAGE handbook of multilevel modeling*: Sage.
- Hodges, J. S. 2013. *Richly parameterized linear models: additive, time series, and spatial models using random effects*, CRC Press.
- Lazega, Emmanuel, and Tom AB Snijders. 2015. *Multilevel Network Analysis for the Social Science*s.
- Grimm, K. J., et al. 2016. *Growth Modeling: Structural Equation and Multilevel Modeling Approaches*, Guilford Publications.

Software

I'm going to provide code for everything that we are doing using R and Stata. Fairly often it will be the case that some option or tool is available in one program but not the other. Part of this class is learning when that's catastrophic or just annoying and how to work around the software when you are pushing up against its limits.

Grades

No student is required to receive a grade to take this course. However, if you are planning to take the class for a grade then please let us know in the course survey sent out on the first day of class. Grades are determined based on a set of homework problems and the analytical portion of a paper. All homework will be available in the course Dropbox on the first day of class.

Assignments

- Open book exam focused on concepts in the first week
    - the basic toolkit
    - data structure
    - missing data
    - omitted variable bias
    - latent dependency structures
- Build an analysis plan and project workflow for your own project
    - A conceptual and diagnostic checklist
    - A data map including sources, sample sizes, and data structure
    - A coherent plan to do the analysis including hypotheses, models, diagnostic tests, and packages.
    - A reproducability plan
- Create an abbreviated analysis plan for 4 of applied problems in class
- Submit a final draft of the analysis for your final project (see the paper section below)

Paper

A key part of this course is the analytical portion of a seminar paper using advanced methods covered in this course. If you have some other advanced multilevel model that is not covered in this course but is required for your analysis you will need to consult with me about it in person. However, even if you plan to use a model covered in this course you should feel strongly encouraged to speak to me about it in person as well. Ideally, this is a paper that you have worked on previously where you want to build up the methods section but a new topic is fine. Do not try to do a project for which you do not currently possess data unless you plan to simulate data. Please turn in the relevant parts of the paper according to the deadlines below.

Wednesday, July 31

- The initial topic
- Hypotheses
- Descriptive statistics
- A stab at an analytical plan

Wednesday, August 7

- A full draft of the analytical plan
- Model results with interpretation (and code to generate them)
- A stab at diagnostics

Friday, August 16

- Final draft due by noon

WEEK 1: MULTILEVEL TOOLS & THEORY

Monday: The Basic MLM Toolkit

- Pooled Models
- Dummy Variables
- Random Intercepts
- Random Slopes

Reading

- The Basic MLM Toolkit Course Notes

Tuesday: Understanding Complex Data Structures

- How Many Groups Do You Have?
- How Many Observations Are In A Group?
- Can You Be In More Than One Group?
- Do You Have Different Types of Groups?
- Do You Care About The Groups?

Reading

- Complex Data Structures Course Notes

Wednesday: Missing Data & Selection Bias

- I Have All The Data (Bayes)
- I Have A Random Sample Of The Data (MCAR)
- I Have A Random Sample Conditional on Stuff (MAR)
- I Have Selection Bias...Help Meeee! (MNAR)

Reading

- Missing Data and Selection Bias Course Notes

Advanced Track: Bayesian Hierarchical Models vs Classical Multilevel Models

Thursday: Omitted Variable Bias

- Mundlak & Hausman
- Fixed vs Mixed Effects
- Random Coefficients

Required Readings

- Omitted Variable Bias Course Notes

Advanced Track: Basic Priors in Bayesian Hierarchical Models

Friday: Latent Dependency Structures

- Groups and Experiments
- Groups and Time
- Groups and Space
- Groups and Networks

Reading

- Latent Dependency Structures Course Notes

Advanced Track: Priors in Bayesian Hierarchical Models for Space, Time, and Networks

WEEK 2: PRACTICAL MODEL FITTING

Monday: How do I even do this?

- Building an Analysis Plan
- A Practical Workflow for MLMs

Reading

- Creating An Analysis Plan.
- Iegor Rudnytskyi (2019) Project-oriented workflow (with R)

Tuesday: Why won't stupid lme4 work?!?

- Convergence Diagnostics
- Random Effect Diagnostics

Reading

- Ben Bolker's FAQ on Mixed Effects Modeling
- Stata 16 ME documentation

Advanced Track: Optimization and Integration in Mixed Effects Models

Wednesday: Why won't %&*#$!@ brms work?!?

- MCMC Settings
- Divergent Transitions
- Prior Sensitivity

Reading

- Burkner PC (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms." The R Journal, 10(1), 395–411. brms help files
- Modrák, Martin (2018). Taming Divergences Stan Models
- Betancourt, Michael (2017).Diagnosing Biased Inference with Divergences

Advanced Track: Integration in Stan – The Theory of Hamiltonian Monte Carlo

Thursday: Do I actually need a MLM? like...really?

- Hausman and Mundlak Tests
- LR Tests and Bayes Factors
- Information Criteria Tests
- Cross-Validation
- The "Keep it Maximal" perspective

Reading

- Luke, S. G. (2016). "Evaluating significance in linear mixed-effects models in R." Behavior Research Methods.
- Ben Bolker's FAQ on Mixed Effects Modeling

Advanced Track: Alternatives to HMC - INLA and Variational Bayes

Friday: How on earth am I supposed to explain this to anyone?

- Caterpillar Plots
- Explaining within, between, and across-level variance
- Explaining technical details
- What to cite so people don't yell as much

WEEKS 3 & 4: SOLVING PROBLEMS

Weeks three and four are about practical problem solving exercises. These are chosen to illustrate real and difficult problems that you may run into in your own work. These applications are **crucial** to making good use of the class...even if they are not from your field...even if you couldn't care less about the subject matter.

I will provide you with a short write-up and summary of each of the applications that you are expected to read before the day that we cover it in class. I will also include reference papers and data. We will walk through the logic of how the authors did their analysis but we will not be replicating them specifically. The papers will act as a guide and inspiration.

Each day we will spend about an hour working through the logic of a problem and the design aspects from week 1 based on an evolving checklist. We will then spend about an hour working through a practical plan to do the analysis including code. This means that each day includes building a project workflow, building an analysis plan, choosing packages, and sketching out a way to frame the results for interpretation.

Each day will involve some mixture of the following

- Refining Research Questions
- Understanding Data Structure
- Selection Bias and Missing Data
- Evaluating Omitted Variable Bias
- Understanding Latent Dependency Structures
- Building a Project Workflow
- Building an Analysis Plan
- Building Fully Annotated Code
- Building Coherent Model Diagnostics
- Interpretation & Presentation

Monday: Explaining Test Scores

This is the basic students in classes in schools example on steroids. We will discuss differences between designs where we have some exogenous intervention for students (aka a causal inference model) and ones where we have observational data and can only really model correlations but have potentially very complex structures at work.

Main discussion topics:

- causal vs observational modeling
- modeling heterogeneity in effects
- complex data structures

Paper & Data

- since this is a running example throughout the first two weeks we won't bother with a paper but we will be working with a simulated data set based partly on the standard jsp data we have used in class until now

Advanced Track: Latent dependency structures and interference as a SUTVA violation

Tuesday: Muti-site RCTs for Microcredit Expansion

We will go though a Bayesian hierarchical analysis of 7 randomized controlled trials on micro-credit by Rachael Meager. This example is being used because of the low number of groups, the fact that there is a specific intervention (micro-credit trials), and the fact that this is an incredibly intimidating paper. While the work is exceptionally well done, it is not actually overly complex for the context of this class.

Main discussion topics:

- heterogeneous treatment effects
- dealing with small numbers of groups
- using ridiculously sophisticated published work as a template for your own

Paper and data

- Meager, R. (2019). "Understanding the average impact of microcredit expansions: A Bayesian hierarchical analysis of seven randomized experiments." American Economic Journal: Applied Economics **11**(1): 57-91.
- Data and code

Wednesday: Changes in Public Opinion Over Time

We will go through some basic and not-so-basic strategies to model dynamic changes in public opinion over time. We will mainly use a data set on Jewish Israeli public support for a two-state solution over a 20 year period while following a similar analysis of a paper (with different data) by Mark Peffley. This data provides an excellent illustration of time varying confounding also known as contextual effect moderation through time.

Main discussion topics:

- multilevel modeling of survey data
- latent dependency structures (time)

Paper

- Peffley, M., Hutchison, M. L., & Shamir, M. (2015). The impact of persistent terrorism on political tolerance: Israel, 1980 to 2011. American Political Science Review, **109**(4): 817-832.

Thursday: Modeling State Policy-Making

We will work out the logic of how to model environmental policy adoption in US states over time. We will mainly follow along with the design and analysis for my paper on state environmental policy adoption. This example highlights latent dependency structures (time, space, networks, latent classes) and complicated grouping structures.

Main discussion topics:

- complex data structures
- small numbers of groups
- latent dependency structures (time, space, networks, latent classes)
- what to do when the ideal best practice isn't technically possible

Paper

- Bromley-Trujillo, R., Butler, J. S., Poe, J., & Davis, W. (2016). The spreading of innovation: State adoptions of energy and climate change policy. Review of Policy Research **33**(5): 544-565.

Advanced Track: A general theory of latent dependency structures

Friday: State Voter Turnout

We will work through how to model state voter turnout in presidential elections as a function of same-day voter registration laws. This is particularly problematic from a causal inference standpoint because voter registration laws are not randomly assigned. We will follow along with the analysis by Yiqing Xu and see if we can't improve it as well.

Main discussion topics:

- causal vs observational modeling
- latent dependency structures (space, time)
- using ridiculously sophisticated published work as a template for your own

Paper & data

- Yiqing Xu. (2019) "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models." Political Analysis **25**: 57-78
- data and example

Advanced Track: Mixed Effects Models vs Complex Differences in Differences Designs

Monday: Modeling Political Systems and Interstate Conflict

We will look at the social relations model for militarized interstate disputes or MIDs. We will follow along in particular with a paper by Cassy Dorff and Michael Ward but the general approach has been used widely in the IR literature.

Main discussion topics:

- complex data structures
- latent dependency structures (time, space, networks, latent classes)

Paper and data

- Dorff and Ward (2013) "Networks, Dyads, and the Social Relations Model*" Political Science Research and Methods **1**159-178
- Correlates of War

Tuesday: Modeling Patient Outcomes in Complex Health Systems

We will look at how to model patient-level outcomes from electronic health record data. These data include millions of cross-classified groups at the patient, physician, hospital, community, and state all across time. We will discuss the computational difficulties involved in fitting a model where approximating random effects in standard ways is impossible. This will be useful for any kind of high dimensional models typically seen with big data analytics.

Main discussion topics:

- utterly ridiculous numbers of groups
- complex data structures
- latent dependency structures (space, time, networks)

Paper

- Bologa, C., et al. (2019). "Generalized Mixed Modeling in Massive Electronic Health Record Databases: what is a healthy serum potassium?"arXiv preprint arXiv:1910.08179.

Wednesday: Predicting the 2020 Presidential Election

We will pretend to be data scientists working for a political campaign in 2020 with the fate of the world depending on us!

Main discussion topics:

- multilevel modeling of survey data
- latent dependency structures (time, space, networks)
- MRP and local area estimation for election prediction
- causal vs observational modeling

Paper and data

- Kiewiet de Jonge, C. P., Langer, G., & Sinozich, S. (2018). Predicting State Presidential Election Results Using National Tracking Polls and Multilevel Regression with Poststratification (MRP). Public Opinion Quarterly, **82**(3):419-446.
- We will use the Cooperative Congressional Election Survey from 2016 since data from 2020 is obviously not available yet.

Thursday: Q&A

I'll provide a general overview and recap of the class and the main things that I'd like you to take with you when you go home. We will then have a Q&A session where you can ask any questions on multilevel modeling that you want.