ICPSR Summer Program 2021 | **Short Course**

Multilevel/Hierarchical Modeling using R

Johannes Karreth

Assistant Professor, Department of Politics & International Relations
Ursinus College
Email: jkarreth@ursinus.edu

Course website: umich.instructure.com and jkarreth.net/mlm-icpsr.html (with links to course materials)
Course meetings: June 21–July 2, 2021 / 10:00am–1:00pm Eastern Time (GMT-4; this is the core course time; asynchronous options are available)

Course description and goals

This course uses the free, open-source software R to train participants in the analysis of multilevel, hierarchical, or structured data. These data are ubiquitous in all of the social sciences and include observations that are nested in higher-level units, such as groups of survey respondents in different countries, students in different schools, or country-level observations at repeated time points. The multilevel framework allows measuring changes and effects at different units of analysis and over time. Participants learn how to appropriately estimate quantities of interests as effects that vary across units and/or time; how much of a change in an outcome of interest is associated with individual- or group-specific features; and varying effects of individual-level characteristics across groups.

Covering both theory and practical applications, the course will explore the following topics:

- Characteristics of multilevel data structures
- Data management for multilevel data
- Fixed and random effects
- Multilevel regression for continuous and categorical outcomes
- Multilevel regression for time-series cross-sectional data
- Multilevel regression and post-stratification
- Post-estimation (incl. marginal effects and predicted probabilities)
- Model assessment and comparison

Upon completion of this course, participants will be able to:

- Understand how multilevel/hierarchical/structured data challenge the assumptions of pooled, i.e., standard OLS regression models
- Distinguish the concepts of fixed and random effects in the context of multilevel data
- Estimate regression models with varying slopes and varying intercepts

- Generate such post-estimation quantities as marginal effects, predicted probabilities, etc. from multilevel regression models
- Use graphical tools to present results from multilevel regression models

The course content will be reinforced through regular hands-on exercises. Participants will learn how to use R to manage multilevel data, analyze their own multilevel data, and to communicate their results to a broader audience. Practical examples and applied exercises form an integral part of the course. Participants will have the opportunity to use data from their own projects for assignments during the course if applicable. A course project allows participants to complete a research project from start to end using multilevel modeling.

Remote Learning Setup

To recreate the in-person experience as much as possible, I created virtual labs as frequent opportunities for hands-on work and immediate feedback from me. Lectures will run on Zoom. Participants can interrupt and ask questions any time. I will also be available for virtual office hours during every day of the workshop, both during fixed times and by appointment.

Prerequisites

Participants should be familiar with basic statistics and the core concepts of linear regression. However, even participants with limited prior experience will be able to effectively participate as the course begins with a review of regression as it relates to multilevel data.

Literature

Participants should have access to:

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

The following book is recommended as a background companion:

- Snijders, Tom A.B. and Bosker, Roel J. 2012. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, Second Edition*. London: SAGE Publications.
- Raudenbush, Stephen W. and Bryk, Anthony S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks: Sage Publications.
- Hox, Joop. 2010. *Multilevel Analysis: Techniques and Applications, Second Edition*. Quantitative Methodology Series. New York: Routledge.

As a general primer for R, I recommend:

- Fox, John and Weisberg, Sanford. 2011. *An R Companion to Applied Regression, Second Edition*. Thousand Oaks: Sage.
- Teetor, Paul. 2011. R Cookbook. Sebastopol, CA: O'Reilly Media.

As a background guide for mathematical concepts discussed in this short course, I recommend:

- Moore, Will H. and Siegel, David A. 2013. *A Mathematics Course for Political and Social Research*. Princeton, NJ: Princeton University Press.

Additional readings will be made available to participants during the course.

Software and Preparation

Before the start of the course, participants should try to install the following programs on their laptops:

1. R is an open-source software package and available for download at http://www.r-project.org.
2. RStudio is a convenient integrated development environment for R and available for free at http://www.rstudio.com.

We will go over how to use these programs on the first day of the course, using a detailed tutorial with step-by-step instructions. We will also have time to catch up on installation problems on the first day.

Course outline

The following time slots and topics may be modified as the course proceeds. The most current version of this document can be found at http://www.jkarreth.net/mlm-icpsr.html and on Canvas.

- **Lectures** are self-contained mini-units mixing lecture and discussion, with slides provided on Canvas.
- **Labs** are guided tutorials with documented scripts available on Canvas.
- **Assignments** are problem sets that participants may complete to reinforce the material learned in the course on that respective day.

Day	Unit	Topic
June 21	Lecture 1	Why consider multilevel regression?
June 21	Lecture 2	Review: Probability and hypothesis testing
June 21	Lecture 3	Review: Regression methods for statistical inference
June 21	Lecture 4	Characteristics of multilevel data structures
June 22	Lab 1	Introduction to R and the lme4 package
June 22	Lecture 5	Inspecting multilevel data
June 22	Assignment 1	Data exploration
June 23	Lecture 6	The linear multilevel model
June 23	Lecture 7	Fixed and random effects: what comparisons are we making?
June 23	Workshop	Your research project in this course: Initial idea
June 24	Lecture 8	Hypothesis testing; Power considerations: how many groups are enough?
June 24	Lab 2	Constructing and processing multilevel data
June 24	Assignment 2	Linear multilevel model
June 25	Lecture 9	Multilevel regression for binary outcomes
June 25	Assignment 3	Binary logit/probit multilevel model
June 28	Lecture 10	Multilevel regression for ordered, categorical, and count outcomes
June 28	Lab 3	Postestimation for multilevel regression
June 28	Lecture 16	Multilevel regression and post-stratification (MRP)
June 29	Lecture 11	Multilevel regression for time-series cross-sectional data
June 29	Assignment 4	Linear multilevel model for time-series cross-sectional data
June 29	Workshop	Your research project in this course: Data and model
June 30	Lecture 12	Model assessment and comparison
June 30	Lecture 13	Bayesian methods for multilevel regression
July 1	Lecture 13, ctd.	Bayesian methods for multilevel regression
July 1	Lecture 14	Missing data in multilevel structures
July 1	Lecture 15	Tricks of the trade
July 1	Lab 4	Model presentation and communication
July 2	Workshop	Project presentations and feedback

For each day, the core reading usually provides substantial details for the units discussed on that day. Background readings typically address questions you may have during and after course. Sample applications demonstrate the techniques encountered on the respective day.

Day 1 Monday, June 21

Lecture 1 Why consider multilevel regression?
Lecture 2 Review: Probability and hypothesis testing
Lecture 3 Review: Regression methods for statistical inference
Lecture 4 Characteristics of multilevel data structures

Core reading:

- Gelman & Hill, chapters 1, 2, 3

Background reading:

- Steenbergen, Marco R. and Jones, Bradford S. 2002. "Modeling Multilevel Data Structures." *American Journal of Political Science* 46 (1): 218–237.

Day 2 Tuesday, June 22

Lab 1 Introduction to R and the lme4 package
Lecture 5 Inspecting multilevel data
Assignment 1 Data exploration

Core reading:

- Gelman & Hill, chapter 11

Background reading:

- Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (10): 1–23.

Day 3 Wednesday, June 23

Lecture 6 The linear multilevel model
Lecture 7 Fixed and random effects: what comparisons are we making?
Workshop Your research project in this course: Initial idea

Core reading:

- Gelman & Hill, chapters 12 and 13

Background reading:

- Mummolo, Jonathan and Peterson, Erik. 2018. "Improving the Interpretation of Fixed Effects Regression Results." *Political Science Research and Methods* 6 (4): 829–835.

Day 4 Thursday, June 24

Lecture 8 Hypothesis testing; Power considerations: how many groups are enough?
Lab 2 Constructing and processing multilevel data
Assignment 2 Linear multilevel model

Core reading:

- Gelman & Hill, chapter 20

Background reading:

- Stegmueller, Daniel. 2013. "How Many Countries for Multilevel Modeling? A Comparison of Frequentist and Bayesian Approaches." *American Journal of Political Science* 57 (3): 748–761.
- Elff, Martin et al. 2020a. "Multilevel Analysis with Few Clusters: Improving Likelihood-Based Methods to Provide Unbiased Estimates and Accurate Inference." *British Journal of Political Science*: 1–15.
- Stegmueller, Daniel. 2020. "Comment on Elff et al." *British Journal of Political Science*: 1–6.
- Elff, Martin et al. 2020b. "Rejoinder to Daniel Stegmueller's Comments." *British Journal of Political Science*: 1–3.
- Bryan, Mark L. and Jenkins, Stephen P. 2016. "Multilevel Modelling of Country Effects: A Cautionary Tale." *European Sociological Review* 32 (1): 3–22.
- McNeish, Daniel and Stapleton, Laura M. 2016a. "Modeling Clustered Data with Very Few Clusters." *Multivariate Behavioral Research* 51 (4): 495–518.
- McNeish, Daniel M. and Stapleton, Laura M. 2016b. "The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration." *Educational Psychology Review* 28 (2): 295–314.
- McNeish, Daniel. 2017. "Small Sample Methods for Multilevel Modeling: A Colloquial Elucidation of REML and the Kenward-Roger Correction." *Multivariate Behavioral Research* 52 (5): 661–670.

Day 5 Friday, June 25

Lecture 9 Multilevel regression for binary outcomes
Assignment 3 Binary logit/probit multilevel model

Core reading:

- Gelman & Hill, chapter 14

Background reading:

- King, Gary, Tomz, Michael, and Wittenberg, Jason. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44 (2): 347–361.
- Hanmer, Michael J. and Kalkan, Kerem Ozan. 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.

Day 6 Monday, June 28

Lecture 10 Multilevel regression for ordered, categorical, and count outcomes
Lab 3 Postestimation for nonlinear multilevel models
Lecture 16 Multilevel regression and post-stratification (MRP)

Core reading:

- Gelman & Hill, chapters 14 and 15

Background reading and applications of MRP:

- Park, David K., Gelman, Andrew, and Bafumi, Joseph. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." *Political Analysis* 12 (4): 375–385.
- Lax, Jeffrey R. and Phillips, Justin H. 2009. "How Should We Estimate Public Opinion in The States?" *American Journal of Political Science* 53 (1): 107–121.
- Kastellec, J. P., Lax, J. R., and Phillips, J. H. (2019). Estimating State Public Opinion With Multi-Level Regression and Poststratification using R. *Working paper*. Available at https://scholar.princeton.edu/sites/default/files/jkastellec/files/mrp_primer.pdf.
- Selb, Peter and Munzert, Simon. 2011. "Estimating Constituency Preferences from Sparse Survey Data Using Auxiliary Geographic Information." *Political Analysis* 19 (4): 455–470.
- Warshaw, Christopher and Rodden, Jonathan. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" *Journal of Politics* 74 (1): 203–219.
- Buttice, Matthew K. and Highton, Benjamin. 2013. "How Does Multilevel Regression and Poststratification Perform with Conventional National Surveys?" *Political Analysis* 21 (4): 449–467.
- Toshkov, Dimiter. 2015. "Exploring the Performance of Multilevel Modeling and Poststratification with Eurobarometer Data." *Political Analysis* 23 (3): 455–460.
- Flores, Andrew R, Herman, Jody L, and Mallory, Christy. 2015. "Transgender inclusion in state non-discrimination policies: The democratic deficit and political powerlessness." *Research & Politics* 2 (4).
- Leemann, Lucas and Wasserfallen, Fabio. 2017. "Extending the Use and Prediction Precision of Subnational Public Opinion Estimation." *American Journal of Political Science* 61 (4): 1003–1022.
- Broockman, David E. and Skovron, Christopher. 2018. "Bias in Perceptions of Public Opinion among Political Elites." *American Political Science Review* 112 (3): 542–563.
- Downes, Marnie and Carlin, John B. 2020. "Multilevel Regression and Poststratification Versus Survey Sample Weighting for Estimating Population Quantities in Large Population Health Studies." *American Journal of Epidemiology* 189 (7): 717–725.
- Kiewiet De Jonge, Chad P, Langer, Gary, and Sinozich, Sofi. Forthcoming. "Predicting State Presidential Election Results Using National Tracking Polls and Multilevel Regression With Poststratification (MRP)." *Public Opinion Quarterly*.
- Smith, Candis Watts, Kreitzer, Rebecca J., and Suo, Feiya. 2020. "The Dynamics of Racial Resentment across the 50 US States." *Perspectives on Politics* 18 (2): 527–538.

Day 7 Tuesday, June 29

Lecture 11 Multilevel regression for time-series cross-sectional data
Assignment 4 Linear multilevel model for time-series cross-sectional data
Workshop Your research project in this course: Data and model

Core reading:

- Bell, Andrew, Fairbrother, Malcolm, and Jones, Kelvyn. 2019. "Fixed and random effects models: making an informed choice." *Quality & Quantity* 53 (2): 1051–1074.

Background reading:

- Hazlett, Chad and Wainstein, Leonard. 2020. "Understanding, Choosing, and Unifying Multilevel and Fixed Effect Approaches." *Political Analysis*.
- Bell, Andrew and Jones, Kelvyn. 2015. "Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data." *Political Science Research and Methods* 3 (1): 133–153.
- Clark, Tom S. and Linzer, Drew A. 2015. "Should I Use Fixed or Random Effects?" *Political Science Research and Methods* 3 (2): 399–408.
- Bell, Andrew, Fairbrother, Malcolm, and Jones, Kelvyn. 2019. "Fixed and random effects models: making an informed choice." *Quality & Quantity* 53 (2): 1051–1074.
- Singer, Judith D. and Willett, John B. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press.

Day 8 Wednesday, June 30

Lecture 12 Model assessment and comparison
Lecture 13 Bayesian methods for multilevel regression

Core reading:

- Gelman & Hill, chapters 18, 24, 25

Background reading:

- Vehtari, Aki, Gelman, Andrew, and Gabry, Jonah. 2016. "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC." *Statistics and Computing*: 1–20.

Day 9 Thursday, July 1

Lecture 13, ctd. Bayesian methods for multilevel regression
Lecture 14 Missing data in multilevel structures
Lecture 15 Tricks of the trade
Lab 4 Model presentation and communication

Core reading:

- Gelman & Hill, chapters 18, 21, 23, Appendix A

Background reading:

- Bell, Andrew, Jones, Kelvyn, and Fairbrother, Malcolm. 2018. "Understanding and misunderstanding group mean centering: a commentary on Kelley et al.'s dangerous practice." *Quality & Quantity* 52:2031–2036.
- Grund, Simon, Lüdtke, Oliver, and Robitzsch, Alexander. 2018. "Multiple Imputation of Missing Data for Multilevel Models: Simulations and Recommendations." *Organizational Research Methods* 21 (1): 111–149.

Day 10 Friday, July 2

Workshop Project presentations and feedback