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
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 for time-series cross-sectional data
June 29	Workshop	Your research project in this course
June 30	Lecture 12	Missing data in multilevel structures
June 30	Lecture 13	Model assessment and comparison
June 30	Lecture 14	Bayesian methods for multilevel regression
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.
A detailed schedule with assigned readings will be posted closer to the course.