Multilevel/Hierarchical Modeling using R

Johannes Karreth
Assistant Professor, Department of Politics & International Relations
Ursinus College
Email: jkarreth@ursinus.edu
Course website: http://www.jkarreth.net/mlm-icpsr.html (with links to course materials)
Course meetings: August 17–21, 2020 / 9:00am–5:00pm Eastern Daylight Time (core course time; asynchronous options available)

Due to Covid-19, this course will run in a virtual format in 2020. See below for more information. Further details on virtual tools will be posted at www.jkarreth.net/mlm-icpsr.html by August 15.

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 will 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, 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.

**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:


The following books are recommended as background companions:


As a general primer for R, I recommend:


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


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](http://www.r-project.org).
2. RStudio is a convenient integrated development environment for R and available for free at [http://www.rstudio.com](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-nus.html](http://www.jkarreth.net/mlm-nus.html) and in the course folder.

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

<table>
<thead>
<tr>
<th>Day</th>
<th>Unit</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>Lecture 1</td>
<td>Why consider multilevel regression?</td>
</tr>
<tr>
<td>Monday</td>
<td>Lecture 2</td>
<td>Review: Probability and hypothesis testing</td>
</tr>
<tr>
<td>Monday</td>
<td>Lecture 3</td>
<td>Review: Regression methods for statistical inference</td>
</tr>
<tr>
<td>Monday</td>
<td>Lecture 4</td>
<td>Characteristics of multilevel data structures</td>
</tr>
<tr>
<td>Monday</td>
<td>Lab 1</td>
<td>Introduction to R and the lme4 package</td>
</tr>
<tr>
<td>Monday</td>
<td>Lecture 5</td>
<td>Inspecting multilevel data</td>
</tr>
<tr>
<td>Monday</td>
<td>Assignment 1</td>
<td>Data exploration</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lecture 6</td>
<td>The linear multilevel model</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lecture 7</td>
<td>Fixed and random effects: what comparisons are we making?</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lecture 8</td>
<td>Power considerations: how many groups are enough?</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lab 2</td>
<td>Constructing and processing multilevel data</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Assignment 2</td>
<td>Linear multilevel model</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lecture 9</td>
<td>Multilevel regression for binary outcomes</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lecture 10</td>
<td>Multilevel regression for ordered, categorical, and count outcomes</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lab 3</td>
<td>Postestimation for multilevel regression</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Assignment 3</td>
<td>Binary logit/probit multilevel model</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lecture 11</td>
<td>Multilevel regression for time-series cross-sectional data</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lecture 12</td>
<td>Model assessment and comparison</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lecture 13</td>
<td>Bayesian methods for multilevel regression</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lecture 14</td>
<td>Missing data in multilevel structures</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lab 4</td>
<td>Working with clustered data in R</td>
</tr>
<tr>
<td>Thursday</td>
<td>Assignment 4</td>
<td>Linear multilevel for time-series cross-sectional data</td>
</tr>
<tr>
<td>Friday</td>
<td>Lecture 15</td>
<td>Tricks of the trade</td>
</tr>
<tr>
<td>Friday</td>
<td>Lab 5</td>
<td>Model presentation and communication</td>
</tr>
<tr>
<td>Friday</td>
<td>Lecture 16</td>
<td>Multilevel regression and post-stratification (MRP)</td>
</tr>
</tbody>
</table>

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, August 17

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

Core reading:
  · Gelman & Hill, chapters 1, 2, 3, 11

Background reading:

Day 2 Tuesday, August 18

Lecture 6  The linear multilevel model
Lecture 7  Fixed and random effects: what comparisons are we making?
Lecture 8  Power considerations: how many groups are enough?
  Lab 2  Constructing and processing multilevel data
Assignment 2  Linear multilevel model

Core reading:
  · Gelman & Hill, chapters 12, 13, and 20

Background reading:
Day 3 Wednesday, August 19

Lecture 9 Multilevel regression for binary outcomes
Lecture 10 Multilevel regression for ordered, categorical, and count outcomes
Lab 3 Postestimation for nonlinear multilevel models
Assignment 3 Binary logit/probit multilevel model

Core reading:
- Gelman & Hill, chapters 14 and 15

Background reading:

Day 4 Thursday, August 20

Lecture 11 Multilevel regression for time-series cross-sectional data
Lecture 12 Model assessment and comparison
Lecture 13 Bayesian methods for multilevel regression
Lecture 14 Missing data in multilevel structures
Lab 4 Working with clustered data in R
Assignment 4 Linear multilevel for time-series cross-sectional data

Core reading:
- Gelman & Hill, chapter 18, 19, 24, 25

Background reading:
Day 5 Friday, August 21

Lecture 15  Tricks of the trade
Lecture 16  Multilevel regression and post-stratification (MRP)
Lab 5  Model presentation and communication

Core reading:

- Gelman & Hill, chapter 21

Background reading:

- Bell, Andrew, Jones, Kelvyn, and Fairbrother, Malcolm. Forthcoming. “Understanding and misunderstanding group mean centering: a commentary on Kelley et al.’s dangerous practice.” *Quality & Quantity*.