ICPSR Summer Program in Quantitative Methods of Social Research
Summer Workshop
Multilevel Analysis with R

July 21-25, 2025

Instructor

Ling Zhu
Professor
Department of Political Science
Email: lzhu4@central.uh.edu
Website: ling-zhu.com

Online Meeting via Zoom

- Zoom ID and link are available on Canvas.

Workshop Description

This five-day workshop introduces participants to various statistical models suitable for multilevel data. Multilevel analysis is a widely used approach for dealing with nested data, which enables researchers to consider both social context and micro-level characteristics in their statistical models. Multilevel analysis also provides advantages and flexibility for modeling longitudinal data. This course will provide participants with a comprehensive roadmap for multilevel analysis, covering topics such as setting up and managing multilevel data, identifying data structures, selecting the appropriate model specification, evaluating fixed and random effects, and interpreting and visualizing statistical results. We will begin with applications in the context of hierarchical linear regression models, and then discuss several extensions suitable for time-series cross-sectional data and categorical dependent variables.

Learning Objectives

Multilevel models can be used for various inferential goals, including descriptive analysis, causal inference, and prediction. This workshop is designed to help participants learn:

- How to analyze structured data;
- How to achieve more efficient inference for regression parameters;
- How to use multilevel analysis to include predictors at different levels;

- How to make sound inferences about treatment effects that vary;
- How to effectively convey complex data information using data visualization.

Prerequisites

Participants should be familiar with the basics of linear regression models and linear algebra; however, prior experience in multilevel data analysis is not required for this workshop.

Recommended Readings

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press. (An excellent/comprehensive introduction to multilevel analysis, with plenty of useful applications in R.)

Luke, Douglas A. 2020 *Multilevel Modeling: Quantitative Applications in the Social Sciences*. Second Edition, Sage Publication. (An accessible Sage monograph on the basics of multilevel modeling.)

Finch, William Holmes, Jocelyn E. Bolin, and Ken Kelley. 2019. *Multilevel Modeling Using R*. Second edition. Chapman and Hall/CRC. (An accessible introduction to various generalized linear models with `R`.)

Rabe-Hesketh, Sophia and Anders Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata*. Third Edition. Stata Press. (A well-stocked toolbox of how to implement multilevel models using `Stata`.)

Wickham, Hadley, and Garrett Grolemund. 2017. R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. 1 edition. Sebastopol, CA: O'Reilly Media. (An accessible introduction to R for data science.)

Wickham, Hadley. 2009. *ggplot2: Elegant Graphics for Data Analysis.* Springer. (An in-depth introduction to data visualization using ggplot2)

Additional journal articles are listed along the workshop calendar to show various examples of multilevel analysis in political science and other social science disciplines. They will be available via the Canvas site for this workshop.

Statistical Software

During this workshop, all lectures and statistical labs will be conducted with an emphasis on `R`. `R` is a statistical environment and high-level programming language for data analysis and visualization. It is the GNU version of the `S` language. `R` is a free and open-source software.

The current (October 2023) version of R is R version 4.3.2 (Eye Holes). R is an object-oriented language. Unlike Stata and most other statistical packages, it operates by assigning values to objects in the workspace. In lecture notes, lab handouts, etc., R commands and outputs will be marked by "R Code" and "R Output", respectively.

———————————-R Code———————————

```
Age<-cps2011$age
summary(Age)
```

———————————-R Output———————————

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.00   15.00   34.00   34.99   52.00   85.00
```

The Comprehensive R Archive Network (CRAN) is the go-to spot for all things related to R. You may also consider using RStudio, which is an integrated development environment for R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management.

Workshop Calendar

The workshop calendar is based on Eastern Standard Time (EST). Each of the following three days will be divided into short morning and afternoon sessions, which include lectures on the theoretical foundations of various multilevel models and data labs on empirical examples of multilevel models. The focus will be given to teaching participants how to specify proper multilevel models for different data structures and the substantive interpretation of statistical results. Lecture notes and R-code for data labs will be distributed through the ICPSR (University of Michigan) Canvas site (weblink TBA) for this workshop.

July 21: Multilevel Analysis with Cross-Sectional Data 1

1. 9:00–11:00 am: Lecture 1. Multilevel Analysis: Basic Concepts
   Recommended Readings
   - Gelman and Hill, Chapter 1.
   - Marc Hooghe, Tim Reeskens, Dietlind Stolle, Ann Trappers. 2009. "Ethnic Diversity and Generalized Trust in Europe: A Cross-National Multilevel Study." *Comparative Political Studies*, 42(2): 198-223.
2. 12:00-2:00 pm, Lab 1. Getting Started with R
   - Lab 1 notes and R script.
   - This is a "warming-up" lab session, which introduces R and RStudio.

July 22: Multilevel Analysis with Cross-Sectional Data 2

1. 9:00–11:00pm: Lecture 2. Varying Intercepts, Varying Slopes, and Non-Nested Models

   Recommended Readings

   - Galmen and Hill, Chapters 13–15
   - Guo, Guang. 2000. "Multilevel Modeling for Binary Data." *Annual Review of Sociology*, 26: 441–462.
   - Austin, Peter C., Henrik Stryhn, George Leckie, and Juan Merlo. 2017. "Measures of Clustering and Heterogeneity in Multilevel Poisson Regression Analyses of Rates and Count Data." *Statistics in Medicine*, 37(4): 572–589.
   - Kim, Jun et al. 2017. "Income, Financial Barriers to Health Care and Public Health Expenditure: Multilevel Analysis of 28 Countries." *Social Science & Medicine*, 176: 158–165.

2. 12:00–2:00pm: Lab 2. Fitting and Interpreting Multilevel Models

   - Lab 2 notes and R script
   - Fitting a multilevel linear regression model
   - Multilevel generalized linear models for binary dependent variable and count data

July 23: Taking Time Seriously: Using Multilevel Analysis to Model Dynamics

1. 9:00–11:00am: Lecture 3. Frailty Models: Cox Regression Models with Random Effects

   Recommended Readings

   - Bienen, Henry and Nicolas van de Walle. 1992. "A Proportional Hazard Model of Leadership Duration." *The Journal of Politics* 54(3): 685–717.
   - Box-Stefensmeier, Janet M. and Christopher Zorn. 2001. "Duration Models and Proportional Hazards in Political Science." *American Journal of Political Science* 45(4): 972–988.
   - Keele, Luke. 2010. "Nonproportionally Difficult: Testing for Nonproportional Hazards in Cox Models." *Political Analysis* 18(2):189–205.
   - Zorn, Christopher. 2000. "Modeling Duration Dependence." *Political Analysis* 8(3):367–380.
   - Hanagal, David D. 2017. "Frailty Models in Public Health." *Handbook of Statistics*, 37: 209–247.

2. 12:00-2:00pm: Lab 3. Estimating and Interpreting Frailty Models

   - Lab 3 notes and R script
   - Survival data and Cox regression: Overview
   - Duration dependence
   - Frailty models: Adding random effects and modeling clusters in survival data

July 24: Multilevel Models for Panel Data

1. 9:00–11:00pm: Lecture 4. Multilevel Dynamic Models for Repeated Measures Data

   Recommended Readings

   - Alejo, Javier, Gabriel Montes-Rojas, and Walter Sosa-Escudero. 2018. "Testing for Serial Auto-Correlation in Hierarchical Linear Models." *Journal of Multivariate Analysis*, 165(2):101–116.
   - Zhu, Ling. 2013. "Panel Data Analysis in Public Administration: Substantive and Statistical Considerations." *Journal of Public Administration Research and Theory*, 23(2): 395–428.

2. 12:00–2:00pm: Lab 4. Multi-Level Panel Data Models: Fixed-Effects, Random-Effects, and Mixed-Effects

   - Lab 4 notes and R script
   - Fixed-effects, random-effects, and population average effects
   - Multilevel mixed effects for repeated noncontinuous measures

July 25: Extensions: More on Multilevel Models for Panel Data

1. 10:00–11:00am: Lecture 5. Growth Curve Models and Other Extensions

   Recommended Readings

   - Plutzer, Eric.2002. "Becoming a Habitual Voter: Inertia, Resources, and Growth in Young Adulthood." *The American Political Science Review*, 96(1): 41–56.
   - Bernick, Ethan M., and Nathan Myers. 2008. "Treatment or Placebo: Are State Programs Decreasing the Proportion of Uninsured?" *Policy Studies Journal*, 36(3): 367–384.
   - Tamborini, Christopher R., Changhwan Kim, and Arthur Sakamoto. 2015. "Education and Life Time Earnings in the United States." *Demography*, 52: 1383–1407.
   - Huang, Jim and Ellen Barnidge. 2016. "Low-income Children's Participation in the National School Lunch Program and Household Food Insufficiency." *Social Science &Medicine*, 150: 8–14

2. 12:00–2:00pm: Lab 5. Growth Curve Analysis and Other Extensions

   - Lab 5 notes and R script
   - Growth curve models
   - Other dynamic multilevel panel data models

Caveat: The aforementioned workshop schedule may be subject to change.