APPLIED MULTILEVEL MODELS

2012 ICPSR FOUR-WEEK INTENSIVE SUMMER WORKSHOP AT THE UNIVERSITY OF MICHIGAN

JULY 16 – AUGUST 10, 2012

Professor Patricia McManus
Department of Sociology
Indiana University
pmcmanus@indiana.edu

Teaching Assistant
Michael Vasseur
mvasseur@indiana.edu

Description
Multilevel models are regression models that exploit data that are clustered within multiple units. Examples include data on children within classrooms, data on patients within hospitals, data on survey respondents within countries, data on crime within neighborhoods, data on firms over time, and repeated observations on individual persons. Multilevel models are also known as mixed models, hierarchical linear models, and random coefficients models. A special class of longitudinal multilevel models involve allow growth curve analysis of repeated observations on the same unit over a period of time. Statistical techniques that exploit the within-group correlation structures of clustered data offer powerful advantages over conventional regression analysis. This course uses an intensive mix of lecture and hands-on computer sessions to introduce students to the specification, estimation, and interpretation of models for the analysis of multilevel data. Class time and lab assignments will emphasize examples using Stata, although some syntax may be provided for SAS PROC MIXED/GLIMMIX and SPSS MIXED. Students with experience in SAS, SPSS or R will find the learning curve to Stata is short. The lab instructor, Mike Vasseur, has experience teaching Stata to students with no prior background in statistical software.

The workshop begins with an overview of applications and examples of multilevel modeling across academic disciplines and in applied research. In the first week, essential terminology and notation are introduced using a review of the general linear regression model in the presence of heteroskedasticity and correlated errors. The second week covers practical applications of the two-level hierarchical linear model from assumptions and specification to estimation, interpretation and visualization. Week three provides an introduction to longitudinal models and growth curve analysis. The final week is devoted to special topics, including data-recentering, three-level models, crossed random effects, and multilevel models for binary and count outcomes. The course assumes familiarity with the linear regression model.

Tentative Daily Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:00-1:50 PM</td>
<td>Lecture</td>
</tr>
<tr>
<td>1:50-2:15 PM</td>
<td>Lab</td>
</tr>
<tr>
<td>2:20-3:00 PM</td>
<td>Lab Review</td>
</tr>
<tr>
<td>3:15-5:00 PM</td>
<td>Office hours</td>
</tr>
</tbody>
</table>

Monday - Thursdays or by Appointment
Required Texts
- Volumes I & II are available as a single package: ISBN 978-1-59718-108-2. This text is referenced in the required readings for the workshop as MLMUS.

- These notes will be available in Ann Arbor on or before the first day of class. The notes contain lecture material, overheads, lab notes and exercises for the course. Bring these materials to class, and to get the most out of the class, review the day’s lecture notes in advance.

Recommended Texts
- This is a comprehensive guide to project management for quantitative data analysis using Stata, but the principles apply more generally. It is especially valuable for researchers with complex projects, collaborative projects, multiple projects.

- The second volume is as thorough as the first, and covers models beyond the scope of a single semester course in multilevel models.

Multilevel modeling is widely used in survey research, field experiments, meta-analytic studies, longitudinal studies, spatial studies, neighborhood studies and family studies. See pp.3-7 in MLMUS for a review of recommended texts in different subfields. Additional applications, textbooks, and data sources are listed in the Acknowledgements section of MLMUS, immediately following pp. xxviii.

Assignments and Course Materials
The course includes short lab exercises and a weekly assignment. Each of the assignments includes data analysis exercises using Stata. Data for the assignments will be provided in the course folder, and there will be a folder for Lab Notes for the computer exercises. The exercises are designed to give you hands-on experience with multilevel data and analytical techniques, the assignments are intended to provide a template for a systematic approach to model-building, estimation, interpretation and presentation.

Course Grades
We will review the short lab exercises and weekly assignments during class time, and students can get additional feedback by visiting office hours. Students who are interested in receiving a grade for this course must a) submit all course assignments with written responses and interpretations, using any of the datasets provided except the example dataset used to review the assignment; b) produce a short paper (5-8 pages) using the techniques learned during the workshop; and c) submit all assignments on or before the due date, and the short paper before the
final day of class. Note that you should design these analyses so that the results reveal interesting relationships, with statistically significant results.

**Workshopping your Project**

As time permits, students with multilevel data and projects may present their work to the class for advice and feedback on the methodological approach and results. Given time constraints, the number of presentations will be likely be limited. Selection will be based on background preparation and our joint assessment of the benefits to be gained by presenting the project in class. Consult with me in the first week if you are interested in presenting.

**Course Outline**

All topics will be covered, but the time spent on each topic will vary depending on student background and interests.

**Week 1**

**Overview and Introduction to Multilevel Models**

1. Overview: introduction to multilevel data analysis, advantages, disadvantages, examples, common errors, terminology and notation (*Day 1*)
2. Review of general linear model: assumptions of single-equation linear regression model in the presence of heteroskedastic or correlated error structures. Estimation, hypothesis testing, model comparisons using OLS, FGLS, MLE. (*Day 2*)
3. Review of interpretations in the general linear model: nonlinear functional form and interaction terms. (*Day 3*)
4. The two-level model: level-1 vs. level-2 covariates, the unconditional means model and intraclass correlation (*Day 4*)
5. Week 1 Lab exercises: data structures, wide vs. long form, describing clustered data, estimation commands for single-level and multilevel models, post-estimation commands for specification-testing.

**Week 2**

**Hierarchical linear models**

1. Random intercept models: specification, assumptions, estimation, goodness of fit, hypothesis testing, model comparisons.
2. Random intercept models with level-2 covariates
3. Random intercept models with level-1 and level-2 covariates.
4. Empirical Bayes predictions and residual diagnostics
5. Random coefficient models: specifying the covariance structure
6. Cross-level interactions
7. Block diagonal models
8. Heteroskedastic errors
9. [Tentative] Project presentation
10. Week 2 Lab exercises: random intercept and linear mixed-model estimation commands; post-estimation commands for hypothesis testing and model comparisons; post-estimation commands for residual diagnostics, post-estimation commands for interpretation and visualization.
**Week 3**

**Longitudinal models and growth curve analysis**

1. Time-varying and time-invariant predictors
2. Covariance structures for conventional error-components models
3. The within (fixed-effects) and between estimators, first-difference models
4. Hausman and Mundlak tests for endogeneity of the level-2 error component.
5. Specification of temporality and the growth curve: empirical growth plots, piecewise linear models, polynomial models
6. Basic models for unconditional means and unconditional growth
7. Specification and testing of the conditional growth curve model
8. Plotting population averaged and prototypical growth curves
9. Latent growth curves
10. {Tentative} Project presentation
11. Week 3 Lab exercises: comparisons of the within, between, first difference and random intercept models; systematic approaches to growth curve analysis; graphing prototypical growth curves.

**Week 4**

**Topics and Extensions**

1. Centering strategies
2. Crossed random effects
3. Three-level models
4. A brief introduction to multilevel models for binary response models and other categorical outcomes.
5. Project presentations
6. Week 4 Lab exercises: alternative approaches to centering; estimating and interpreting the two-way random effects model; estimating and interpreting the three-level random effects model.