SYLLABUS

Multilevel Analysis of Survey Data
Summer Institute in Survey Research Techniques

University of Michigan: SurvMeth 988.202
University of Maryland: SURV699K

Note: This is a video course taught simultaneously at the University of Michigan and the University of Maryland. Although the syllabus is the same for each location, registration, on-site faculty, staff, and facilities are different. Check the syllabus and the individual institutions for details.

Logistical Information

Where and When

Class Time: June 29-July 17, 2015, 9:30-12:15 p.m.

Location: The location for lectures and labs at the University of Michigan is room 300 in the Perry Building; location for lectures and labs at the University of Maryland is room 2208 in Lefrak Hall. These are video-conference rooms that we will use for all sessions (lectures and labs).

<table>
<thead>
<tr>
<th>Staff</th>
<th>University of Michigan</th>
<th>University of Maryland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructors:</td>
<td><strong>Associate Prof. Robert Henson</strong></td>
<td><strong>Associate Prof. Robert G. Croninger</strong></td>
</tr>
<tr>
<td></td>
<td>Educational Research Methodology</td>
<td>Teaching and Learning, Policy and</td>
</tr>
<tr>
<td></td>
<td>TBA</td>
<td>Leadership</td>
</tr>
<tr>
<td></td>
<td>University of Michigan</td>
<td>2115 Benjamin Building</td>
</tr>
<tr>
<td></td>
<td>Ann Arbor, MI 48106</td>
<td>University of Maryland</td>
</tr>
<tr>
<td></td>
<td>Phone: TBD</td>
<td>College Park, MD 20742</td>
</tr>
<tr>
<td></td>
<td><a href="mailto:rahenson@uncg.edu">rahenson@uncg.edu</a></td>
<td>(301) 405-2927</td>
</tr>
<tr>
<td></td>
<td></td>
<td><a href="mailto:croninge@umd.edu">croninge@umd.edu</a></td>
</tr>
<tr>
<td>Office Hours:</td>
<td>Wed 12:30-2:30, Room TBD</td>
<td>Wed 1:00-3:00 pm, Benjamin 2110</td>
</tr>
</tbody>
</table>
Graduate Volunteer:  

**Matt Griffin**  

0313A Benjamin Building  

University of Maryland  

griff23@umd.edu

Office Hours: TBD

---

**Registration**

The course is offered through the University of Michigan's Summer Institute in Survey Research Techniques at the Institute for Social Research and the Joint Program in Survey Methodology at the University of Maryland. Dr. James Lepkowski directs the Summer Institute at the University of Michigan; Jill Esau is education programs administrator (jesau@isr.umich.edu). The Institute’s office is in room 4050 in the ISR building at the University of Michigan (734-647-4620). Fred Conrad directs JPSM; Rupa Jethwa Eapen is the assistant to the director (rjeapen@survey.umd.edu). The JPSM office is in room 1218 in Lefrak Hall (301-314-7911).

The course enrolls both degree and non-degree students. Degree students may take the course either for credit or as an audit; non-degree students are all auditors (“summer scholars” at UM). Students at the University of Michigan register for SurvMeth 988.202 and pay tuition through the Summer Institute; students at the University of Maryland register for SURV 699K and pay tuition as part of the University’s summer schedule.

---

**Course Overview**

**Purpose and Aims**

A problem. Although the multilevel modeling is applicable to many fields of study, our course focuses on applications in education. A major phenomenon of interest in educational research is the learning of individual students. Two educational realities -- (1) that learning implies change in individuals over time but (2) learning occurs in organizational settings (classrooms, schools) -- cause persistent and troublesome methodological problems in quantitative research. One problem is how to measure change. Another involves how to accommodate the grouped nature of the phenomenon (often referred to as "the unit of analysis problem"), as education almost always occurs in groups (classrooms, schools) while its effects are expected to accrue to individuals. These methodological problems have distinct, long-standing, and non-overlapping literatures. In a sense, the problems share a common cause: traditional statistical techniques are not adequate to model the hierarchical nature of either phenomenon.

Applicability. Education illustrates a clear dilemma in this regard, and much educational research is flawed because of it. However, the problem is not limited to that field. Both measuring change and estimating effects on individuals when the "treatment" takes place in groups are common phenomena in social science and evaluation research. Therefore, although many of the examples that we use in this course come from education, the usefulness of the techniques learned in the
course have broad applications. In past years, students have come from many social science and professional fields.

**Focus.** The course is meant to introduce students to multilevel methodology and to help them become skilled in its use. We begin by discussing the need for multilevel methods in survey research, as well as the circumstances under which these methods are appropriate and necessary. We ask students to read articles, in which the method is employed, both to become familiar with the standard terminology and notation and to review the substance of a variety of work where these methods are applicable.

**Target audience and prerequisites.** Although the course focus is on quantitative methodology, our orientation is more toward applications than statistical derivations. We expect that students have a working knowledge of general linear models and non-experimental multivariate methods. We also expect that students have considerable experience with statistical computing in a social science setting. Generally, this means that students have completed at least two semesters of graduate training in quantitative research methods and are capable of formulating, conducting, and interpreting results for analysis of variance (ANOVA) and OLS regression with multiple variables. Students also should be reasonably skilled in and comfortable with using common general-purpose statistical software (e.g., SPSS, SAS, STATA) for data manipulation.

**Objectives.** Hopefully, students in this class will learn:

1. To formulate multilevel research problems or questions.
2. To pose and test hypotheses about various parameters of multilevel models and to write clearly about both the statistical and substantive meaning of these results.
3. To examine critically the application of multilevel methods to particular problems in light of its statistical assumptions.
4. To analyze a meaningful two-level linear model, write up the results, and draw substantive conclusions from the analysis.
5. To gain a reading knowledge of applications of multilevel models beyond the standard two-level organizational model.

**Organization**

**Format.** We teach the course in a workshop format, which alternates between lectures and labs. Both lectures and labs are required, as this is a "hands-on" course. Lectures and labs are informal, meant to provide opportunities for questions and discussions. A culminating activity in the course is the final project (both a paper and an oral presentation). The purpose of presenting students' "work in progress" is to provoke discussion around the methodology employed in a variety of contexts (data, research questions). In past years, several final projects have been developed into conference presentations and gateway research activities.

The class structure and format is meant to accommodate some variety in the backgrounds and expectations students bring to the class. Lectures provide the theory and purpose of multilevel models, discussion of analyses using them, as well as a vehicle for understanding what is being done, why it might be done, and what conclusions may be drawn from the results. The labs offer "how-to" experiences, where students receive instruction in how to use the statistical software,
HLM (Hierarchical Linear Models). Although out-of-class work involves both readings and weekly write-ups of computing projects, the majority of out-of-class time is spent at a computer, learning to work with HLM and to prepare data and variables to use in assignments. There is no exam.

Requirements. Requirements include attending all lectures and labs, completing assigned readings, writing up two short and one longer data analysis, and making an oral presentation about the final project. As the course is designed around learning to use multilevel modeling, all students (including auditors) are expected to complete the assignments, in order to learn the methodology, how to apply it to data, and to receive feedback on their work. The assignments are meant to be progressive. Thus, it is impossible to do the later assignments without mastering the skills involved in the earlier ones. If for some reason students need to miss a class, they are still expected to make arrangements to review the materials covered and to complete the assignments (most lectures and labs are video taped).

The course finale is the presentation of final projects, each focusing on a topic of personal and substantive interest. This will include both a written paper of 8-10 pages and a short presentation to the class (10-15 minutes). Students may work individually or in pairs on all assignments. Assignments 1-2 are each worth 30% of the final grade, and the final project is worth 40% of the grade (15% oral presentation; 25% paper).

Data. We make use of the first two waves of the Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K). However, class members are welcome to use their own data (if the data have an appropriate multilevel structure) for assignments. If you plan to do this, you should discuss it early with the instructors.

The ECLS-K data are available from the National Center for Education Statistics (NCES) in the U.S. Department of Education. Many NCES datasets exist in two forms: public-use and restricted. ECLS-K is available in both forms, but we will make use of the public-use version. You should feel free to obtain a copy of these data from NCES on your own (go to: http://nces.ed.gov/ecls/). For the class, we will supply students with SPSS systems files (children and schools) made from ECLS-K, where we have sampled down the numbers of children, schools, and variables. The dataset contains the first two waves of ECLS-K data, the Fall 1998 wave and Spring 1999 wave when children where kindergarteners. Because the sampling frame for ECLS-K involved a two-stage process (first schools were sampled, then a random sample of kindergarten children in those schools was drawn), the data structure is quite appropriate for school-effects studies using multilevel statistical methods.

Computer usage. We’ll provide each of you with a copy of the latest version of HLM for Windows (HLM 7.01). This version has been provided to us by the authors and programmer for instructional purposes and will “expire” at the end of the class. Students who do not use Windows as an operating system should consider purchasing software that creates a virtual platform for running Windows (e.g., Fusion or Parallel). If you wish to purchase your own copy of HLM software, the cost is about $495, with modest additional charges for future upgrades. It is also possible to do a “trial-run” and rent HLM for specific periods of time. The HLM software is sold by Scientific Software Incorporated (SSI). Check the website for more specific information www.ssicentral.com.
Materials

A. Books:


Raudenbush, S.W., Bryk, A., Cheong, Y.F., Congdon, R.T., & du Tort, M. (2012). HLM 7: Hierarchical Linear and Nonlinear Modeling. Chicago, IL: Scientific Software International. [This is the handbook or users guide for the software we will use in class. It is only available online as a pdf file; it can be downloaded at the class C-Tools website. It takes a little time to open so be patient.]


B. Articles:

We have supplemented readings from Raudenbush and Bryk with articles. Some of these articles address technical issues while others provide examples of the use of multilevel modeling in research studies. These readings are available on the class C-Tools site.

C. Websites

There are numerous websites in Europe and the United States with lectures and training materials on multilevel models. We do not specifically recommend any of these sites, but you may want to explore some of them to see what is useful and available to you. We list a few here: the Scientific Software International website, a site at the University of Bristol (UK), another at the University of California Los Angeles, and a site at the University of Michigan for Optimal Design.

www.ssicentral.com. [There are FAQs and a number of “how-to” postings.]

http://www.cmm.bristol.ac.uk/learning-training/ [Some video lectures and examples, mostly associated with MLwiN; follow some of the links to find lists of publications, recommended textbooks, and how to join a multilevel discussion group.]

http://www.ats.ucla.edu/stat/ [No videos about HLM but examples and datasets from some of the major textbooks. Some good videos of lectures about other techniques and SPSS, too.]

http://sitemaker.umich.edu/group-based/optimal%20design%20software [Optimal design, web-based resources for doing power analyses using multilevel models.]
Schedule

The pace of the course is fast. Each week includes an assignment that involves posing a multilevel research question, data manipulation, computing, interpreting, and writing. We expect all students (credit or no-credit) to complete each assignment and present their final project at the end of the course. The assignments build on one another, becoming progressively more complex. Because of the fast pace, we discourage requests for extensions on assignments. It is important that students practice the skills covered in lab and master them promptly, as knowledge is cumulative. The major time spent on the course is at the computer, particularly in the first two weeks.

The topics and reading assignments listed for each class are approximate, based on the pace we think we will follow and reflecting the order of topics covered class by class. Although the progression will be the same as listed in the syllabus, the days on which each topic is covered depends, to some extent, on how the course progresses. Assignment deadlines will not change.

The final project and its presentation represent the culminating experience of the course. Format for the project should resemble articles from scholarly journals (e.g., Presentation of a meaningful problem or issue (roughly 1-2 pages), presentation of research questions (1 page), discussion of methods used to answer the research questions 1-2 pages, presentation of results (2 pages), and discussion of the substantive meaning of the findings (2-3 pages). You need not conduct a literature review on your chosen topic.

We expect that students read the assigned readings for each class before the class meets. Some of the readings for the course are quite technical, and you may find them difficult to fully understand. Some students have found it useful to revisit the more complex readings after the lecture, concentrating on the particular issues covered in the class.

We strongly encourage students to work in pairs. In past years, many students have told us that the collegiality of working with a partner greatly enhanced their experience in this class. Because research out of class settings is almost always collegial, we have found that two minds focused on something new often produce more than the sum of the parts.

WHY DO WE NEED MULTILEVEL MODELS?

Monday, June 29: Introduction to the multilevel approach. Why use multilevel modeling?
We’ll discuss complex data structures and the challenges these structures pose for research, including aggregation bias and misidentification of potential sources of variance. The optional readings for today describe the dataset that we have provided to you for your course assignments.

Readings
Chapter 1, Hierarchical linear models, 2nd edition (Raudenbush & Bryk)

Multilevel models: When and why (Hox)

Optional Readings
Early childhood longitudinal study (Westat)

Working with ECLS-K (powerpoint)
LOGIC OF MULTILEVEL MODELS

Tuesday, June 30: Logic and start of HLM Lab Boot Camp. Today we’ll discuss the logic of multilevel models, starting with specification of a fully unconditional model and the calculation of the intra-class correlation (ICC). The articles by Kim et al., McCoach, Roberts and Peugh provide alternative introductions to multilevel modeling.

Your first assignment is due Monday July 6th. Write a 3-4 page paper that includes a set of multilevel research questions, a brief description of the data and why multilevel modeling is the appropriate method to use with the data, calculate the ICC for the dependent variable that you chose, and interpret the results. You will have an opportunity to work on the assignment in lab.

Readings
Chapter 2, Hierarchical linear models, 2nd edition (Raudenbush & Bryk)

Applying hierarchical linear modeling to social work administration data (Kim et al.)

Optional Readings
Dealing with dependence (McCoach)

An introductory primer on multilevel and hierarchical linear modeling (Roberts)

A practical guide to multilevel modeling (Peugh)

Wednesday, July 1: Continue HLM Bootcamp and Group Activity. We’ll continue discussing the logic of multilevel models and orienting you to the datasets and software used in the class. We’ll spend time interpreting the various files used by HLM, such as input files, output files, stat files, MDM, and we’ll run a fully unconditional model and interpret the ICC. No optional readings for today.

Readings
Chapter 1-2, HLM7: Hierarchical linear and nonlinear modeling (Raudenbush et al.)

BASIC MULTILEVEL MODELS

Thursday, July 2: Examples of basic models. We’ll discuss basic multilevel models -- how to specify them and interpret them. The Lee et al. article provides an example of a two-level, random intercept model using the ECLS-K dataset while the optional readings provide additional examples of models.

Readings
Chapter 4, Hierarchical linear models, 2nd edition (Raudenbush & Bryk)

Full-day v. half-day kindergarten (Lee et al.)

Optional Readings
Predicting teacher commitment using principal and teacher efficacy variables (Ware & Kitsantas)

Multilevel models for school effectiveness research (Rumberger & Palardy)
Friday July 3: No class (off for 4th of July observance).

Monday, July 6: Random intercepts and slopes. We’ll look more closely at two different types of models – a random intercept model and a random intercept and slopes models. We’ll also begin a more in-depth discussion of centering and how level-1 centering decisions influence the interpretation of the level-2 coefficients. The Enders and Tofighi and the Preacher et al. articles discuss centering and cross-level interactions; the optional readings also discuss centering options, a sometimes-controversial topic in multilevel modeling.

Readings
Centering predictor variables in cross-sectional multilevel models (Enders & Tofighi)
Computation tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis (Preacher et al.)

Optional Readings
Centering decisions in hierarchical linear models (Hoffman & Garvin)
Effects of different forms of centering in hierarchical linear modeling (Kreft et al.)
Understanding and estimating the power to detect cross-level interaction effects (Mathieu et al.)

Tuesday, July 7: LAB: When building a multilevel model, we specify the level-1 model before specifying the higher levels of the model. We’ll discuss model-building strategies, random v. fixed effects, the meaning of lambda, centering options, and interpretation of HLM output. The optional reading discusses the use of multilevel modeling to estimate contextual effects.

Second Assignment due Friday, July 10th. 4-5 page paper that presents a level-1 model, discusses the logic of the model, and interprets each coefficient (build on your first assignment, if possible). Include at least one random slope in your model. Note: we will review everything that you’ll need to be able to do to complete your second assignment during this session.

Readings
Chapters 1-2 (review), HLM7: Hierarchical linear and nonlinear modeling (Raudenbush et al.)

Optional Readings
Using hierarchical linear modeling to study social contexts (Lee)

Wednesday, July 8: Model assumptions. We’ll discuss some of the assumptions for multilevel models, the consequences of violating those assumptions, and how to think about model fit. We’ll introduce strategies for checking model assumptions and exploring model fit. Optional readings highlight issues associated with the violation of assumptions and power.
Readings
Chapter 9, Hierarchical linear models, 2nd edition (Raudenbush & Bryk)

Evaluation of model fit and adequacy (McCoach & Black)

Optional Readings
Influence of violations of assumptions on multilevel parameter estimates and standard errors (Maas & Hox).

Power, sample size, and design (Spybrook)

Power and sample size in multilevel models (Snijders)

Thursday, July 9: LAB: We’ll finish our discussion of the specification of two-level models and the interpretation of results. We’ll explore different strategies for presenting results, including the graphing options using HLM software. No optional readings for today.

Readings
Chapter 18, HLM7: Hierarchical linear and nonlinear modeling (Raudenbush et al.)

MODELING CHANGE

Friday, July 10: Introduction to modeling change. Modeling change presents unique challenge to research, including regression to the mean and multicolinearity. We’ll discuss prior conceptual and technical issues associated with modeling change, and the opportunities for modeling change presented by multilevel models (i.e., growth curves). The Huttenlocher article is the full article for the vocabulary growth model discussed in Chapter 6. In addition, we’ll begin the LAB where we discuss how you set up a growth curve MDM. The unit at level 1 is time rather than individuals, which requires a change in how we think about setting up our data. We’ll provide examples of how to create a MDM and do a growth curve analysis.

Readings
Chapter 6, Hierarchical linear models, 2nd edition, pp. 160-185 (Raudenbush & Bryk)

Measuring change more effectively by modeling individual growth (Willet)

Optional Readings
Early vocabulary growth (Huttenlocher et al.)

Monday, July 13: Finish modeling change and LAB from Friday. There are aspects of modeling change not discussed from the previous Friday such as piecewise models that will be discussed. No optional readings for today.

Readings
Chapter 6 (review), Hierarchical linear models, 2nd edition, pp. 160-185 (Raudenbush & Bryk)
Summarizing data through piecewise linear growth curve models (Chandrasekaran et al.)

**Tuesday, July 14: Three-level model with change.** We’ll extend the examples to three-level models. Conceptually, there is no limit to the possible levels that you might consider, though each additional level adds complexity. A common three-level model would be change at level 1, individual at level 2, and some organizational structure at level 3.

**Readings**
Chapter 8, Hierarchical linear models, 2nd edition, pp. 160-185 (Raudenbush & Bryk)

Higher order instructional goals in secondary schools (Raudenbush et al.)

**Optional Readings**
Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model (Bryk & Raudenbush)

A multilevel study of predictors of student perception of school climate (Kothe et al.)

**ADDITIONAL MODELS**

**Wednesday, July 15: Hierarchical generalized linear models.** Not all dependent variables are continuous. What if the dependent variable is binary, ordinal, count, or multinomial? We’ll introduce the idea of hierarchical generalized linear models, using binary outcomes as an example. The optional readings discuss additional hierarchical generalized linear models. The article by Lee and Burkam uses a binary outcome.

**Readings**
Chapter 10, Hierarchical linear models, 2nd edition (Raudenbush & Bryk)

**Optional Readings**
Dropping out of high school: The role of school organization and structure (Lee & Burkam)

Cross-classified and multiple membership structures in multilevel models (Fielding & Goldstein)

Neighborhoods and violent crime (Samson et al.)

An illustration of multilevel modeling for ordinal response data (O’Connell)

Beyond individual differences (Everson & Millsap)

**Thursday, July 16: Lab.** HGLM. We’ll help you think about how to run your multilevel models using the logic of hierarchical generalized linear models. These models are an extension of what we have done so far in class, where the assumed distribution of the dependent variable is normal and the “link” is identity. We’ll focus on binary outcomes as an example.

**Readings**
Chapters 7-8, HLM7: Hierarchical linear and nonlinear modeling (Raudenbush et al.)
STUDENT PRESENTATIONS

Friday, July 17: Presentation of final projects. This is a presentation of your final project for the class, which should be an expansion of your work for Assignment 2 (see the description of the final project on page 6). Depending on the number of presentations, you should plan on roughly a 10-15 minute power point presentation with some questions about your work (similar to a conference presentation). Include handouts of tables and graphs as appropriate.

Your final papers are due Monday, July 20th. Please submit them electronically. If you use citations to support the theory on which your work rests, please use APA or ASA format, and include all citations. Citing course readings is appropriate and recommended. Although these are relatively short papers, the optional reading by Dedrick et al. provides some insights into what to include in multilevel papers and how to format tables.

Optional Readings
Multilevel modeling (Dedrick et al.)

Estimating multilevel models using SPSS, Stata, SAS, and R (Albright & Marinova)

Revised June 9, 2015