CEP 935—Advanced Topics in Multivariate Data Analysis II

Syllabus

Instructor: Spyros Konstantopoulos

Time: 3:00 - 6:50 PM Thursdays
Room: 228 Erickson Hall
Office: Erickson Hall 450
Email: spyros@msu.edu
Office Hours: By appointment (please email me ahead of time to schedule an appointment)

Teaching Assistants: Xin Luo, Li Ma

Email: luoxin1@msu.edu
Email: mali2@msu.edu

Office Hours: Mondays and Tuesdays TBD (Erickson Hall graduate student lounge)

Course Content:

Data with nested or hierarchical structure are common in education and the social sciences. A typical example stems from education, where students are nested within classrooms, classrooms are nested within schools, schools are nested within districts and so forth. Another example stems from organizational research, where employees are nested within different departments of firms, and these departments are in turn nested within firms and so forth. In addition, in longitudinal studies (with repeated measures for each individual) observations are nested within individuals, who are within larger units such as schools or firms etc.

It is important whenever the data have a nested structure to conduct analyses that will take the clustering of the data into account. The last two decades, much progress has been made in the development of appropriate methods and software for analyzing multilevel data (Goldstein, 1987; Bryk and Raudenbush, 1992). Statistical models for hierarchically structured data are known as random or mixed effects, variance components, multilevel, or hierarchical linear models. The term HLM is more common in education.

This course focuses on the logic and use of multilevel models in education and the social sciences. Specifically, the primary objective of the course is to illustrate ways in which multilevel models help addressing questions in teacher and school effects research, in organizational research, in evaluation research, in the study of individual change. Although most of the examples that will be discussed in this course are related to education, the methods are applicable to any situation where the data are nested. For
example, the models can be used to examine differences in achievement among classrooms, schools, or to determine how family background and educational experiences relate to differences in rates of students’ cognitive growth.

This course will also illustrate the use of multilevel models in meta-analysis where effect size estimates are nested within studies, and studies and nested within investigators or labs and so forth. Finally, the course will portray the use of multilevel models for binary outcomes, or for ordinal, count, or multinomial data.

The primary required text for the course will be Raudenbush and Bryk’s *Hierarchical Linear Models: Applications and Data Analysis Methods* (Second Edition, 2002). This book will be supplemented with various articles that discuss applications of the methods. The software programs SPSS and HLM will be used in the class to demonstrate the usefulness of the methods. The student version of HLM 7 for windows is freely available on the web (http://www.ssicentral.com/hlm/student.html). The student version should include the HLM 7 manual as a pdf file. Other specialized software is MLWin. SAS, SPSS, and STATA can also be used. SAS and SPSS should be available on most campus computers.

**Requirements:**

1. Five homework assignments (50% of the course grade). Some homework assignments will involve data analysis using pre-specified datasets. We encourage you to work in groups of two or three on your homework assignments. All teams of students should turn in a single hard copy of the group homework with all names listed; all group members will receive the same grade. All homework assignments must be typed (12 point font 1 inch margin). The assignments are due on the beginning of the class the day they are due (hard copies).

2. Discussion of assigned readings (papers) in class (10% of the grade).

3. Final paper (40% of the course grade, which includes the power point presentation of the paper). The final paper should demonstrate the usefulness of multilevel methods and should be like a short publishable quality paper. The paper should include an introduction that discusses the significance of the research question, a short literature review of the main papers in the area, a detailed data analysis plan, and presentation and discussion of the results. References, tables, figures, and appendixes (if applicable) should also be included. APA style is recommended. Data will be available ahead of time for use. Alternatively, students are encouraged to use their own datasets if available. Again, we encourage you to work in groups of two or three on your final paper. The text must not exceed 12 pages (double spaced, 12 point font 1 inch margin). Students are expected to do a power point presentation of their projects and discuss their methods and findings at the end of the semester (last class period). The presentation’s quality should be comparable to a conference’s presentation. The presentation will take about 15 minutes and there will be five minutes for questions/discussion. Students can use any software (e.g., HLM, SAS, SPSS, STATA, etc.) to analyze the data. **The final paper will due**
Friday 5/3/2013 at 10:00 am (hardcopy). Please return the final paper to Erin Johnson at 455 Erickson Hall.

Grades are criterion-referenced. That is, grades will be assigned based on the percent of the total possible points that you receive on the examinations and the assignments [4.0 > 90%, 3.5 > 85%, 3.0 > 80%, 2.5 > 70%, 2.0 > 60%].

Prerequisite:

Students should have completed CEP 933 and CEP 934. Students who have not completed such courses must have a very good understanding of linear models and multiple regression (including estimation formulae, assumptions, inferences, interpretation, etc). Although knowledge of matrix algebra is not required as a prerequisite for the course, students should have some background in matrix algebra. This notation simplifies the presentation of complex models. Therefore, it is expected that students will familiarize themselves with the contents of the Namboodiri text.

Required textbooks:


Other useful textbooks:


Required readings:


### University and Class Policies

- **Michigan State University** seeks to ensure that its programs are accessible to all persons. Students in need of special assistance or an accommodation regarding any of the course requirements as outlined in the syllabus and discussed in class are advised to notify me immediately. We will meet privately to discuss a resolution of your issue, which may or may not include an appropriate referral. Confidentiality will be maintained regarding your special needs. To arrange for accommodation, students with disabilities should contact the Resource Center for People with Disabilities (RCPD) ([http://www.rcpd.msu.edu/](http://www.rcpd.msu.edu/)).

- **Academic honesty**: academic dishonesty, including plagiarism, may result in a zero grade in the course and removal from the program. If students are unclear about the Academic honesty policy, they are encouraged to consult the following link [https://www.msu.edu/unit/ombud/academic-integrity/plagiarism-policy.html](https://www.msu.edu/unit/ombud/academic-integrity/plagiarism-policy.html)

- The instructor reserves the right to make any changes he considers academically advisable. Changes will be announced in class, it is your responsibility to keep up with any changed policies, schedules, and assignments.

- You are responsible for any materials covered in classes that you miss.
### Tentative Schedule of Topics

<table>
<thead>
<tr>
<th>Class</th>
<th>Date</th>
<th>Topic</th>
<th>Reading</th>
<th>Homework Assignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>1/10</td>
<td>Course overview; Rationale for using multilevel models. Reading: R&amp;B Chapter 1. Review on Regression. First homework assignment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 2</td>
<td>1/17</td>
<td>Introduction to the logic of HLM. Reading: R&amp;B Chapter 2; N-all. Connections between HLM and regression models; Matrix algebra.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 3</td>
<td>1/24</td>
<td>HLM model - Estimation. Reading: R&amp;B Chapter 3; Lee &amp; Bryk (1989), Raudenbush &amp; Bryk (1986). <strong>First homework assignment due.</strong> Second homework assignment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 4</td>
<td>1/31</td>
<td>Introduction to two-level models. Reading: R&amp;B Chapter 4; Konstantopoulos &amp; Hedges, 2008; Lee, 2000. Computer Lab: Introduction to the HLM software.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 5</td>
<td>2/7</td>
<td>Using HLM in evaluation studies. Reading: R&amp;B Chapter 5; Carlson et al. (2011); Seltzer (1994). <strong>Second homework assignment due.</strong> Third homework assignment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 10</td>
<td>3/21</td>
<td>Three level models continued. R&amp;B Chapter 8. <strong>Fourth Homework Assignment Due.</strong> Fifth homework assignment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class 12</td>
<td>4/4</td>
<td>Assessing the adequacy of multilevel models. Reading: R&amp;B Chapter 9. <strong>Fifth homework assignment due.</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Class 14 (4/18)  Power Point Presentations.


Final paper due 5/3/2013 10:00 AM (hardcopy).
Multivariate data consist of individual measurements that are acquired as a function of more than two variables, for example, kinetics measured at many wavelengths and as a function of temperature, or as a function of pH, or as a function of initial concentrations, and so forth, of the reacting solutions. The absence of measurements in multivariate data sets is a persistent problem in chemistry and other fields. Obviously, when an entire row or column of a data matrix is missing, there is little recourse, but when a limited number of isolated measurements are missing in a more dispersed pattern, some information is still retained in the remaining data and exclusion of entire measurement vectors is generally not necessary.

It is important whenever the data have a nested structure to conduct analyses that will take the clustering of the data into account. The last two decades, much progress has been made in the development of appropriate methods and software for analyzing multilevel data (Goldstein, 1987; Bryk and Raudenbush, 1992). Statistical models for hierarchically structured data are known as random or mixed effects, variance components, multilevel, or hierarchical linear models. Multivariate Data Analysis - Science topic. Explore the latest questions and answers in Multivariate Data Analysis, and find Multivariate Data Analysis experts. Questions (428). Publications (213,596). Questions related to Multivariate Data Analysis. 1. 2. I have a data frame of multivariate abundances, measured from sites under two different treatments. These sites have been sampled each summer for ~15 years, and thus are repeated measures. I am modelling my data using the mvabund package in R, which fits Generalized Linear Models.