# **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 Namboordiri text.

**Required textbooks:**

Raudenbush, S., & Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Thousand Oakes, CA: Sage. [This book is denoted R&B in the readings list below.]

Namboodiri, K. (1984). *Matrix Algebra: An Introduction*. Thousand Oaks, CA: Sage Publications. [This book is denoted N in the readings list below.]

**Other useful textbooks:**

De Leeuw, J., & Meijer, E. (2010). *Handbook of multilevel models.* Springer.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Goldstein, H. (2011). *Multilevel statistical models*. Wiley.

Hox, J.J. (2002). *Multilevel analysis: Techniques and applications*. Lawrence Erlbaum.

Longford, N. T. (1993). *Random coefficient models*. Oxford University Press.

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* Oxford University Press.

Snijders, T.A.B. & Bosker, R.J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Sage.

**Required readings:**


Bryk, A. S., & Raudenbush, S. W. (1988). Toward a more appropriate conceptualization 0f research on school effects: A three-level hierarchical linear model. American Journal of Education, 97, 65–108.

Carlson, D., Borman, G.D., & Robinson, M. (2011). A multi-state district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis, 33,* 378-398.

Hedges, L. V., & Vevea, J. L. (2008). Fixed- and random-effects models in meta -analysis. *Psychological Methods, Vol 3,* 486-504.

Huttenlocker, J. Haight, W., Bryk, A. Seltzer, S., Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236-248.

Konstantopoulos, S. (2008). Do small classes reduce the achievement gap between low and high achievers? Evidence from Project STAR. *Elementary School Journal, 108,* 275-291.

Konstantopoulos, S. (2011). How consistent are class size effects? *Evaluation Review, 35,* 71-92.

Konstantopoulos, S., & Hedges, L. V. (2008). How Large an effect can we expect from school reforms? *Teachers College Record, 110,* 1613-1640.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. Sociology of Education, 62, 172–192.

Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: The case 0f school effects. Educational Psychologist, 35, 125–141.

Nye, B., Hedges, L.V. & Konstantopoulos, S. (2000). The effects of small classes on achievement: The results of the Tennessee class size experiment. *American Educational Research Journal, 37,* 123-151.

Nye, B., Konstantopoulos, S, & Hedges, L.V. (2004). How Large are Teacher Effects? *Educational Evaluation and Policy Analysis, 26,* 237-257.

Rumberger, R. W., & Palardy, G. J. (2005). Test Scores, Dropout Rates, and Transfer Rates as Alternative Indicators of High School Performance. *American Educational Research Journal, 2005, 42,* 3–42.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. Sociology of Education, 59, 1–17.

Sampson, R. J., & Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277(5328),* 918-924.

Seltzer, M.H. (1994). Studying variation in program success: A multilevel modeling approach. *Evaluation Review, 18*(3), 342-361.

Seltzer, M., Frank, K. & Bryk, A. (1994). The metric matters: The sensitivity of conclusions concerning growth in student achievement to choice of metric. *Educational Evaluation and Policy Analysis, 16,* 41-49.

Seltzer, M.H., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis, 25*(3), 263-286.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics, 23,* 323-355.

**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/).*
- *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
- *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**

| | |
|---|---|
| Class 1 (1/10) | Course overview; Rationale for using multilevel models. Reading: R&B Chapter 1. Review on Regression. First homework assignment. |
| Class 2 (1/17) | Introduction to the logic of HLM. Reading: R&B Chapter 2; N-all. Connections between HLM and regression models; Matrix algebra. |
| Class 3 (1/24) | HLM model-Estimation. Reading: R&B Chapter 3; Lee & Bryk (1989), Raudenbush & Bryk (1986). **First homework assignment due.** Second homework assignment. |
| Class 4 (1/31) | Introduction to two-level models. Reading: R&B Chapter 4; Konstantopoulos & Hedges, 2008; Lee, 2000. Computer Lab: Introduction to the HLM software. |
| Class 5 (2/7) | Using HLM in evaluation studies. Reading: R&B Chapter 5; Carlson et al. (2011); Seltzer (1994). **Second homework assignment due.** Third homework assignment. |
| Class 6 (2/14) | Computer Lab: HLM, SAS, SPSS, STATA. Reading: Singer 1988; HLM manual (pdf file with student version). |
| Class 7 (2/21) | The use of HLM in growth models. Reading: R&B Chapter 6. Huttenlocher et al. (1991); Seltzer et al. (1994); Seltzer et al. (2003). **Third homework assignment due.** Fourth homework assignment. |
| Class 8 (2/28) | Using HLM in meta-analysis. Reading: R&B Chapter 7, Hedges & Vevea, 1998; Konstantopoulos 2008, 2011. Computer lab in class. |
| Class 9 (3/14) | Introduction to three level models. Reading: R&B Chapter 8; Bryk & Raudenbush (1988); Nye et al 2000, 2004. Computer lab in class. |
| Class 10 (3/21) | Three level models continued. R&B Chapter 8. **Fourth Homework Assignment Due.** Fifth homework assignment. |
| Class 11 (3/28) | Multilevel models for binary data. Reading: R&B Chapter 10; Rumberger & Palardy (2005). Use of HLM in criminology. Reading: Sampson et al. (1997). Fifth Homework Assignment. Computer Lab in class. |
| Class 12 (4/4) | Assessing the adequacy of multilevel models. Reading: R&B Chapter 9. **Fifth homework assignment due.** |

Class 13 (4/11) Cross-Classified random effects models. R&B Chapter 12.

Class 14 (4/18) **Power Point Presentations**.

Class 15 (4/25) **Power Point Presentations**.

**Final paper due 5/3/2013 10:00 AM (hardcopy).**