EDF 7474
Multilevel Modeling
2011 Spring Term

Instructor: James Algina

Office: 119D Norman Hall

Phone: 273-4305

Email: algina@ufl.edu

Course Web Page: http://plaza.ufl.edu/algina/index.html

Reserve site: https://ares.uflib.ufl.edu/

- To help me keep track of email messages, please include **EDF 7474 and your last name** in the subject line of any correspondence sent via email.

- I have created a distribution list using your GATROLINK email address. If you do not use your GATROLINK account regularly, please make sure your GATROLINK account forwards to the account you do use regularly.

Office Hours: Tuesday 4:00 – 6:00 PM, Thursday 5:00 – 6:00 PM, and by appointment. (To set an appointment, send me an email indicating several times when you are free.)

Text (Recommended)

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd. Ed.). Sage: Newbury Park, CA.

Optional Resource

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2004). *HLM6 Hierarchical Linear and Non-linear Modeling*. Chicago: Scientific Software.

Topics

The following is a potential list of topics for the course. Time may not permit covering all of the topics.

1. Introduction to multilevel designs. Chapter 1

2. Multilevel models. Chapter 2, 3, 4, 5. (Chapter 3 is technical and can be skipped. Chapter 5 is most relevant to the discussion of centering.)

3. Designing Multilevel Studies.

4. Longitudinal data. Chapter 6.

5. Assumptions. Chapter 9.

6. Meta analysis. Chapter 7.

7. Three-level Models. Chapter 8.

8. Hierarchical models for categorical data. Chapter 10.

9. Models for cross-classified random effects. Chapter 12.

Class Attendance

Students are expected to attend class unless they have a valid excuse for missing a class. Examples of valid excuses are illness, attendance at professional conferences, and the like. The advanced topics are not covered by an exam. Nevertheless, students are expected to attend class when those topics are covered. Failure to do so without a valid excuse will result in a decrease in the final grade of a grade point category (e.g., an A will become an A-).

Computer Program

The main computer program for this course will be HLM 6. You can download a free student edition at www.ssicentral.com.

The student edition can run all the analyses the full version can in terms of models selected, statistical options, and output. Restrictions are, however, placed on the data used and the size of the model selected. The following restrictions apply in this edition:

- The DBMSCOPY utility used for the importation of data is not included. The student edition will only accept ASCII, SYSTAT, SPSS for Windows or SAS transport data files.
- For a level-2 model, the maximum number of observations is 7200 at level-1 and 350 at level-2 of the hierarchy.

- For a level-3 model, the maximum number of observations is approximately 7500 at level-1, 1700 at level-2, and 60 at level-3.

- No more than 5 effects may be included in any HLM equation at any level of the model, and the grand total of effects cannot be 25 or higher.

The regular version of HLM 6 is available for use in room 115 Norman Hall. This lab is usually open between 9:00 and 5:00 but the hours can be irregular on Fridays. The computers on

which HLM 6.0 is installed are also used by students who are not enrolled in EDF 7474, so plan to do your work in a timely fashion.

The data used to do the analyses reported on pages 69 to 85 in Raudenbush and Bryk are stored on the course web page. An introduction to HLM [Using HLM 6 (pdf format)] is also provided on the course web page. You should download the data and the introduction. As an introduction to using HLM, do the analyses reported on pages 69 to 85. One of the requirements in the course is to replicate these analyses on a different data set than was used by Raudenbush and Bryk.

Course Requirements

Summary of Due Dates

Task	Due date
Copy of Paper to be Reviewed	1/20
Data Analysis	2/17*
Proposal	2/24
Paper Review	3/3
Exam	3/24*
Final Paper	4/22

*Tentative

Examinations

An examination is tentatively scheduled for Thursday, March 24, 2011. It will cover the material in Chapters 1 to 5. A list of study questions will be provided. It is a good idea to keep up with answering the study question. Otherwise preparing for the exam can be challenging.

Papers and Projects

All work must be submitted electronically to algina@ufl.edu. Remember to include EDF 7474 and your last name in the subject line. With the exception of the final paper, all work is due by 3:00 PM on the due date. The final paper is due by 5:00 PM on the due date.

Data Analysis

Obtain a data set to which the hierarchical linear model can be applied and that includes (a) one quantitative level-1 dependent variable, (b) one quantitative level-1 independent variable, (c) one quantitative level-2 independent variable, and (d) one dichotomous level-2 independent variable. (If you do not have access to data I will show you how to find such data in the NELS88 or ECLS data sets, but you can use any data set that is available to you.) Use HLM to conduct analyses that are reported on pages 69 to 85 in Raudenbush and Bryk. Your report will consist of

- A brief opening section in which you describe the variables you will be analyzing, list the abbreviations you use in the HLM program, and explain what the level-1 and level-2 units are.

- Answers to questions that will be provided to you. All but the last two are from pages 69 to 85 in Raudenbush and Bryk.

- The entire HLM printout for each model.

- The *.sts file and descriptive statistics produced by SPSS for your level-1 and level-2 files. Either there should be no discrepancies between the results in the *.sts file and the descriptive statistics produced by SPSS or you should provide an explanation for the discrepancies.

Due two weeks after we complete the material on Chapter 4 in R&B. I estimate the due date to be Thursday, February 17, 2011.

Paper Review

Review one paper in which hierarchical **linear** models were used to analyze the data. (Do not review papers that focus exclusively on the following hierarchical models: logistic, multicategory logistic, ordinal, Poisson, survival.)

Due Thursday, March 3, 2011.

I need to review your selected paper to make sure they are suitable for the project. Submit a copy of the paper you have selected no later than Thursday, January 20, 2011. Keep a copy of the papers for your work.

In your review

(a) Describe the data that were collected, **including stating what the level-1 and level-2 sampling units are.**

(b) Report on the models that were estimated. We will study the following types of multilevel models: random effects ANOVA, means as outcomes, random intercepts, random regression coefficients, intercepts as outcomes, and intercept and slopes as outcomes. **If the authors report more than one application of a particular type of model, select one of the applications for your report.** [***More than one application of a particular type of model might be included in your article because multiple dependent variables were used or because different sets of independent variables were used***.] Do not include models estimated by OLS. For each type of model the authors use, do the following:

1. Present the equation used. Include the level-1 and level-2 equations. Also include the equation for the combined model. You can present the equation either symbolically, in list form (see pages 11-12 for an example of list form) or in a table (see page 12 for an example of a tabular form). If you present the equation in symbolic form and the equation for the residual in the combined model is complex, you can exclude the residuals from the equation, for example,

$$E(Y) = \gamma_{00} + \gamma_{01}Z_1 + \gamma_{02}Z_2 + \gamma_{10}X + \gamma_{11}XZ_1.$$

   You do not have to include expressions for the residual variances $\left(\text{e.g., } u_{0j} \sim N(0, \tau_{00})\right)$.

2. Classify each model with respect to the following categories: random effects ANOVA, means as outcomes, random intercepts, random coefficients, intercepts as outcomes, and intercept and slopes as outcomes.

3. For each model to which centering applies, state the centering of the level-1 variables. If the authors do not explicitly state the centering, try to infer it from the author's presentation of the models and/or results. But if the centering is unclear, just say so. If you can determine the centering for a model briefly discuss whether or not it was appropriate and why. If you cannot determine the centering, state what you believe is the correct centering.

4. Present the estimated fixed effect coefficients. You can do this in table form or equation form. Just make clear which parameter symbol from # 1 goes with which estimate. Sometimes an author will report only a subset of the estimates of the fixed effects for a particular model. If your author follows this practice, indicate which estimates of fixed effects are not reported.

5. Present the estimated residual variances. Use the symbols presented in class (i.e., $\hat{\sigma}^2$, $\hat{\tau}_{00}$, $\hat{\tau}_{11}$, etc.) to make clear which estimate goes with which parameter. Sometimes an author will report only a subset of the estimated residual variances for a particular model. If your author follows this practice, indicate which estimated residual variances are not reported.

6. Interpret of as many of the following estimates as the authors report: fixed effect coefficients, variance components, and pseudo R^2 statistics. Do not simply accept the author's interpretation. Write your own. The author may be wrong. For fixed effects coefficients do not simply state which are significant and non-significant. Explain what the coefficients tell you about the phenomena under investigation. Summarize the results of each model.

(c) As a culminating section, present an overarching summary of the results **all multilevel analyses** reported by the authors.

N.B. Late work will be subject to a grade point category penalty for each class period or

part thereof that the work is late.

Final Paper

This paper is due Friday, April 22, 2011. Paper submission after April 23, 2011 may result in a grade of I grade for the course. Papers submitted after April 22, 2011 will be subject to a grade point category penalty for day or part thereof that the work is late.

Choose one of the following four alternatives.

A. Conduct a research study in which the hierarchical linear model or hierarchical generalized linear model is the primary method of analysis. This kind of paper should be written in APA style and should include an introduction, method, results and discussion section at a minimum. The introduction should be a review of the literature that culminates in the research questions to be addressed and a rationale for investigating the questions.

I will evaluate the paper by using criteria typically used to judge papers submitted for publication.

1. Is there a good rationale for the research?

2. Is the data collection clearly and adequately described?

3. Are the results presented clearly and in appropriate detail?

4. Are the results interpreted correctly?

5. Are the conclusions clearly justified by the results?

Printouts used in preparing the paper should be appended to your paper when you submit your paper.

The *.sts file and descriptive statistics produced by SPSS for your level-1 and level-2 files (and level-3 files if necessary) should be appended to your paper. Either there should be no discrepancies between the results in the *.sts file and the descriptive statistics produced by SPSS or you should provide an explanation for the discrepancies.

B. Write a literature review on a methodological issue in hierarchical linear modeling.

The culminating section for a literature review must be either

1. A rationale for and design of a study to answer methodological problem that emerges in the literature review;

or

2. A guide to practice with regard to the methodological issue.

The latter would comprise decisions that need to be in an analysis, and criteria for making those decisions. Preparing a guide for practice should be undertaken only if the decisions and criteria are sufficiently complex to make the guide a contribution to the literature on multivariate analysis.

The criteria for evaluating the literature review section will be adequacy of coverage of the literature and the accuracy, clarity, and integration of the review. With regard to the last three criteria, you must give sufficient detail to permit the reader to know what were the purposes, methods and results of each study. Equally important, you must describe consistencies and inconsistencies that emerge across the papers.

A guide to practice will be evaluated by the criteria of importance, clarity, and correctness.

A proposed research study will be evaluated in terms of clarity and quality of the rationale and of the design.

C. Present and illustrate a multilevel modeling method that will not be presented in class. This paper will include a didactic presentation as well as analyses of at least one illustrative data set. If you select this type of project, you must report the project in the form of a paper. In particular do not present slides like those I use in class.

D. Make a comparison of two or more methods (including at least one multilevel modeling method) that can be used to analyze a given data set. The comparison should include a didactic presentation and comparison of the methods as well as analyses of at least one illustrative data set. If you select this type of project, you must report the project in the form of a paper. In particular do not present slides like those I use in class.

You must submit a brief proposal for your paper (two pages or less) <u>on or before Thursday, February 24, 2011</u>.

A proposal for a research study should include (a) a brief description of the method used to collect the data and the variables, (b) sample size information for all levels that are relevant in the study, (c) the questions you plan to answer, and (d) the models you will estimate.

A proposal for a literature review should include a description of the methodological issue and a tentative list of papers you will include in the review.

A proposal for types C or D should identify the method(s) and data set(s) that will be used.

It may be helpful for you to talk with me about your paper before you submit the proposal.

I will respond to your proposal by (a) approving it, (b) asking for clarification, or (c) asking for a new proposal. In the event of (b) I will expect your response by the class period following the date on which I return the proposal. In the event of (c) I will expect your response by the fourth class period following the date on which I return the proposal. Getting a proposal approved sometimes takes several cycles.

Late proposals will be accepted as will revisions that do not meet the time lines presented above. However, late proposals and resubmissions may have a negative impact on the grade awarded to your paper.

Grading

All work will be scored by using the following 23-point scale:

Score	Grade
21-22	A
19-20	A-
17-18	B+
15-16	B
13-14	B-
11-12	C+
9-10	C
7-8	C-
5-6	D+
3-4	D
1-2	D-
0	E

The final grade will be determined using the following weights for the three projects that will be submitted:

Data Analysis	20%
Article Review	20%
Exam	20%
Paper	40%

and will use the following scale

Score	Grade
[21-22	A

[19-20)	A-
[17-18)	B+
[15-16)	B
[13-14)	B-
[11-12)	C+
[9-10)	C
[7-8)	C-
[5-6)	D+
[3-4)	D
[1-2)	D-
0	E

[21-22 means your average must be 21 or above to be awarded an A.

Academic Honesty

Students are expected to do their own work. You should not solicit help with your exercises or your paper from other people, including students and faculty, nor should you provide help to other students. Soliciting help includes discussing with other people, including students and faculty, what the appropriate analysis of the data might be. Type the following statement on the cover sheet for the paper and exercises:

"**On my honor I have neither given nor received unauthorized aid on this assignment,**"

and sign the statement. By the University of Flrida honor code, even if you do not type and sign this statement, it is implied when you submit an exercise or paper.

Example of a symbolic expression:

Level-1

$$LC_{ij} = \beta_{0j} + \beta_{1j}SES + \beta_{2j}Gender + e_{ij}$$

Level-2

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Size + \gamma_{02}Climate + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Size + \gamma_{12}Climate + u_{1j}$$

$$\beta_{2j} = \gamma_{20}$$

Combined

$$\begin{aligned} LC_{ij} = {} & \gamma_{00} + \gamma_{01}Size + \gamma_{02}Climate + \gamma_{10}SES + \gamma_{20}Gender \\ & + \gamma_{11}SES \times Size + \gamma_{12}SES \times Climate + u_{0j} + u_{1j}SES + r_{ij} \end{aligned}$$

This is the model presented in list form:

The level 1 variables were SES and Gender and the level-1 equation was

- Equation for the intercept
    - Level-1 coefficients: β_{0j} (intercept), β_{1j} (SES), β_{2j} (Gender)
    - Residual: r_{ij}

. The fixed effects and random effects were

- Equation for the intercept
    - fixed effects: γ_{00} (Intercept), γ_{01} (Size), and γ_{02} (Climate)
    - random effect: u_{0j}
- Equation for the slope for SES
    - fixed effects: γ_{10} (Intercept), γ_{11} (Size), and γ_{12} (Climate)

- random effect: u_{1j}

- Equation for the slope for Gender
  - fixed effects: γ_{30} (Intercept), γ_{31} (Size), and γ_{32} (Climate)
  - random effect: none

In the combined model the following product terms were included: $\gamma_{11} SES \times Size$ and $\gamma_{12} SES \times Climate$

This is the model presented in a table:

	Equation				
	Level-1	Level-2 for Intercept	Level-2 for SES Slope	Level-2 for Gender Slope	Combined
Intercept	β_{0j}	γ_{00}	γ_{10}	γ_{20}	γ_{00}
SES	β_{1j}				γ_{10}
Gender	β_{2j}				γ_{20}
School Size		γ_{01}	γ_{11}		γ_{01}
School Climate		γ_{02}	γ_{12}		γ_{02}
SES × School Size					γ_{11}
SES × Climate					γ_{12}
Level-1 residual	e_{ij}				e_{ij}
Level-2 residual		u_{0j}	u_{1j}		u_{0j} and u_{1j}

For relatively simple model, the equation form is likely easier to prepare. For a complex model the list or table form may be easier to prepare. You are permitted to use ellipses $(\cdots$ or $\vdots)$ in equations and tables as long as you make clear what variables the ellipses run over.