**Applied Advanced Statistics**

LSAP 3595 – Fall 2018

Tuesdays 1:05 – 3:45

5520 Posvar Hall

Instructors

|  |  |
| --- | --- |
| Rip Correnti803 LRDC412-624-8765rcorrent@pitt.edu |  |

Baeksan Yu

806 LRDC

412-932-9410

 postcentre@pitt.edu

**Introduction:**

This course is designed as an advanced seminar for students who are interested in exploring statistical modeling with their own social science data. There will be two primary foci for the course. First, the course is designed to provide opportunities for students to examine their own data and gain experience interpreting and writing about their findings. In this sense the course attempts to build student knowledge about the use of advanced statistical techniques and provide individualized experience applying those methods to data that is directly relevant to the student (i.e., data they know they want to work with now or that mimic the anticipated structure of the data they will work on in the future). Please note, if the student does not have access to any data then the course instructor will provide an array of data students can choose from. The course is also designed as a seminar where students will gain exposure to different statistical models and also exposure to the issues that their classmates are working on. Therefore, a second goal of the course is also to provide opportunities for students to learn about a variety of advanced methods that they can consider in their future work.

**Prerequisites:**

PSYED 2410 (Applied Regression) or equivalent

**Recommended Text**s:

Raudenbush, S. W. Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*, 2nd edition. Newbury Park, CA: Sage.

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

Raudenbush, Bryk, Cheong, & Congdon (2004). *HLM6 (manual).* Lincolnwood, IL: SSI.

Throughout the class we will create supplementary source lists addressing different topics in Hierarchical Linear Models.

**Required Materials:**

All reading assignments will be available by e-mail or through Box.

For the analysis that we are going to run throughout the class we will mainly use the computer program—HLM V7.01—for this course. Students will be provided access to a computer with HLM software. Additional information about the software and HLM in general can be found at the Scientific Software International website (<http://www.ssicentral.com/hlm/hlm.htm>). Throughout this course, the instructor will demonstrate data file preparation for HLM using SPSS.

**Course Requirements and Grades:**

Students will complete one course assignment related to the lecture material. The course assignment will be worth 20% of the students’ grade for the semester. In addition, students will be evaluated on the basis of their progress and performance for their Final Project. Assignment 1 and assignment 2 will account for 30% of the total grade (each assignment is worth 15%), and the Final Project will account for the remaining 50% of the total grade; the *written* Final Project will account for 40% of the grade while a *class presentation* will account for the remaining 10%.

***Requirements for Assignment 1*** (Due Date: Sept. 27)

Provide HLM outputs for the models you run and one or two paragraphs in which you describe;

* Your study and dataset
* Why HLM is an analytical tool to examine the data
* Main dependent variable(s) of your study
* One or more potential hypotheses for your study
* What fully unconditional model tells you about variance for main dependent variable

***Requirements for Assignment 2*** (Due Date: Nov. 1)

Provide HLM outputs for the models you run and a paper in which you;

* Describe independent variable(s) and hypotheses of your study
* Choose a model and describe what each parameter means
	+ Check correlation matrix
	+ Check distributional assumptions
	+ Transform your dependent variable if necessary
	+ Create composite variable(s) if necessary (Factor Analysis, Item Response Theory)
* Write results
	+ Create tables
	+ Create graphs if necessary
	+ Describe coefficients
	+ Talk about statistical significance of the results
	+ Talk about effect size of the results
* Tell us how confident you are in the inferences drawn from the interpretations of the analysis
	+ Adequacy of your model
	+ Alternative ways to characterize your data

***Requirements for Final Assignment***  (Due Date: Dec. 14)

Write a research paper based on your data. The paper should focus primarily on the Methods and Results you have obtained with your data over the course of the semester. Guidelines for the paper write-up and the grading rubric are found in the **Appendix** at the end of the syllabus. The paper should include the following elements:

1. Rationale – brief theoretical background describing the importance of the construct(s) analyzed
2. Research Questions – direct, concise, and aligned with rationale and methods
3. Methods – a) model write-up, including description and purpose of the model chosen, b) descriptives table conveying only essential information for readers to orient to your analyses and results
4. Results – a) table of model results including fixed and random effects from the model, b) appropriate interpretation of findings that is accurate and interesting
5. Conclusions – findings discussed align with original RQs and hypotheses and discuss appropriate strengths of the data and findings as well as limitations

***The statements contained in this syllabus, other than the grade policies, may be subject to change with reasonable advance notice as deemed acceptable by the instructors***

**Departmental Grievance Procedures**

The purpose of grievance procedures is to ensure the rights and responsibilities of faculty and students in their relationships with each other. When a student in LSAP believes that a faculty member has not met his or her obligations (as an instructor or in another capacity) as described in the Academic Integrity Guidelines, the student should follow the procedure described in the Guidelines (See below) by (1) first trying to resolve the matter with the faculty member directly; (2) then, if needed, attempting to resolve the matter through conversations with the chair/associate chair of the department; (3) if needed, next talking to the associate dean of the school; and (4) if needed, filing a written statement of charges with the school-level academic integrity officer.

The more specific procedure for student grievances within LSAP is as follows:

1. The student should talk to the faculty member to attempt to resolve the matter.
2. If the matter cannot be resolved at that level, the student should talk to the relevant program coordinator (if the issue concerns a class) or his or her advisor.
3. If the matter remains unresolved, the student should talk to the chair of LSAP (currently Mary Kay Stein).
4. If needed, the student should next talk to the SOE associate dean of students. If the matter still remains unresolved, the student should file a written statement of charges with the dean’s designated Academic Integrity Administrative Officer.

**University Policies on Academic Integrity and Disability**

**Academic Integrity Code**

As members of a community of learners, all students are expected to meet the obligations of honesty and respect for the ethical standards of the University community and of their chosen field of study. You are therefore expected to familiarize yourself with the published rules and regulations governing *academic integrity*, a term meaning the ethicalstandards of integrity by which each student and faculty member is expected to operate. The College of General Studies Academic Affairs and Advising maintains an Academic Integrity Officer who handles disputes regarding the Academic Integrity Code for negotiation of grievances between faculty and students. You may read the full Academic Integrity Code online at <http://www.as.pitt.edu/faculty/policy/integrity.html>. If a student wishes to speak to the Academic Integrity Officer about a grievance they should contact the CGS Assistant Dean of Student Services at 412-624-6600.

**Plagiarism** is a serious offense and will be dealt with in accordance with the University procedures and guidelines located at <http://www.english.pitt.edu/resources/plagiarism.html>. While we encourage students to discuss ideas outside of class and collaborate with one another, it is expected that the written work turned in by each student represent their own thinking and ideas or with appropriate credit given to the origin of the ideas (with citation).

If you have a **disability** for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and Disability Resources and Services (DRS), 216 William Pitt Union, 412-648-7890 (412- 282-7355 for TTY) as early as possible in the term. DRS will verify your disability and determine reasonable accommodations for this course.

**Tentative Course Outline**

|  |  |  |  |
| --- | --- | --- | --- |
|  **Week**  | **Date** |  **Topic** |  |
| 1  | Aug. 30 | Introduction to Hierarchical Linear ModelingDiscussion of Individual Data Sets  | Lecture and ReadingsIndividual Consultation |
| 2 | Sep. 6 | Two-level Hierarchical Linear Models * Creating mdm/mdmt files

(Understanding mdmt files and learning how to manipulate mdmt files) | Building .mdm for HSB exampleIn-class Worksheet (FUM) |
| 3 | Sep. 13 | Two-level Hierarchical Linear Models * Run Models with Covariates
* Output interpretation
* Deciding on which random coefficients to keep
* Multicollinearity
* Effect size
 | In-class WorksheetCourse assignment 1:Create a .mdm and run a random coefficient model |
| 4 | Sep. 20 | Finish HSB Example* Explaining variance in slopes

Three-level HLM models as an extension of two level models | Course Ass. 1 due In-class worksheet |
| 5 | Sep. 27 | Linear Growth Models* Data Structure
* Centering
* Run 2-level Growth Model
 | **Final Project run FUM and 2-Para. Write-up (Ass. 1 due)**In-class Lab and worksheet |
| 6 | Oct. 4 | Quadratic Growth Models * Run 2-level Growth Model
* Lecture
* Lab
 | In-class Lab |
| 7 | Oct. 11 | Quadratic Growth Models* Run 3-level Growth Model

Alternative Growth Specifications* Lecture
 | In-class worksheet |
| 8 | Oct. 18 | Piecewise Growth Models  | In-class Lab and Worksheet |
| 9 | Oct. 25 | Non-Linear Models – Hierarchical Generalized Linear Models (Logistic; Poisson; Multinomial)* Lecture
* Logistic Example
 | In-class Lab and Worksheet |
| 10 | Nov. 1 | Non-Linear Models* Poisson Example
 | ***Final Project* outline of methods and results (Ass. 2) Due**  |
| 11 | Nov. 8 | Model checking assumptions | In-class Lab and Worksheet (tbd) |
| 12 | Nov. 15 | Power Analysis – Designing and Proposing Multilevel Studies* G-Power
* Optimal Design
 | LectureIn-Class Lab |
| ***Thanksgiving Recess*** |
| 13  | Nov. 29 | Multivariate Multilevel Models | In-class Lab |
| 14 | Dec. 6 | Multiple Imputation – Handling Missing Data * Lecture
 | *OMET* |
| 15 | Dec. 14 |  | ***Final Paper* (Ass. 3) Due** |

APPENDIX

Guide to Writing up Your HLM Results and Rubric for Grading

1. Research Questions:
* For final papers I want to see a specific hypothesis/research question (RQ)
* I also want to see a potential literature review or your thought process leading to the hypotheses and RQs – it is important that the front matter motivating the analyses is well aligned to the actual RQ and results (Note: This may be re-written from your first Final Project paragraph depending on how the analyses fit your initial RQ and analysis)
* RQs should be carefully crafted and as direct and concise as possible
* The front matter should also try to elucidate what is important about the numbers being shown in the results section…e.g., why should we care about the ICC? Whet the reader’s appetite for greek symbols and numbers by making them care about the substance of your RQ(s)
1. Descriptions of the sample used:
* What is the available population that the sample was drawn from
* How much missing data is there in the sample? Missing data is an important issue when you do an HLM analysis, especially since data missing at the 2nd or 3rd level automatically omits that case from the analysis
* Provide descriptive statistics in Tabular form
* Conduct a missing data analysis to compare your final sample with the full sample – this is important when you have cases that are excluded from analysis since it is important to know who you can generalize the findings to. Is it only units with no missing data? Was there a systematic reason why units were missing data? What defines the missing population? Are the results spurious due to that missing data? These are all questions the reader might have so try to address them before reviewers have a chance to comment and complain.
* If data comes from a prior study give the context for the data – is this a quasi-experimental study? is it an RCT?
1. Description of dependent and independent covariates:
* Provide the reader with a sense of the distribution of the dependent variable – this may have important consequences for the type of model chosen and it may also have important consequences for how the coefficients communicating your results are interpreted
* Provide the reader a table of the independent covariates – the style of the table can vary but it is often a good idea to include some basic demographic information such as means, SD, range, etc. In addition, in the text it is often a good idea to give the reader a sense of what a score on an individual item or composite means – e.g., if a variable is on a scale of 1 to 4 a score of “2” means…? For covariates of primary interest in your analysis it is often a good idea to give the reader some descriptive information about those covariates – help them understand the mean score and the distribution of scores around the mean.
* Unimportant covariates (e.g., controls) can be included in the table but do not necessarily need to be discussed in the text.
* When creating composites give information about the cronbach’s alpha or some empirical or theoretical justification for the composite. Name the composite something meaningful and help the reader understand what a high score on the construct means and what a significant effect of the composite might indicate about the effects of that construct.
* There is such a thing as providing too much information about covariates – try to be selective and highlight for the reader which covariates are of highest importance and focus discussion primarily on those covariates.
* If you find multicollinearity in your correlations in spss you can deal with this by dropping one of the variables from your analysis. An alternative approach is to think about a way to create a composite.
1. Model and methods write-up:
* Rather than list all covariates and presenting (only) the mixed model from the HLM output –use the general form of the models and center the readers’ attention not on the covariates included in the model, but more on the modeling decisions that will be made.
	+ For example, in a model examining interactions, you should be sure to include whether a particular covariate being examined is going to be fixed (meaning no variance estimated for that effect) or whether it is allowed to vary randomly or be free (meaning variance between higher level units on the effect of that parameter will be estimated).
	+ It is important to communicate to the reader when you make these modeling decisions – so they can understand the model being estimated.
	+ It is also often substantively interesting and should be cast in that light. One reason for examining an interaction is because you think there is a mediating or moderating relationship between some covariates and an effect of a different covariate. Walk the reader through this so they can understand the relationship being examined. That way when they get to the results they can anticipate what “success” might look like – what the meaning of a positive or negative effect is.
* Covariate names used in the models should be able to be understood by the reader and should match the variable names used and described in the tables of descriptives.
* In the write-up of the models it is appropriate to discuss how you will interpret the coefficients if you are using a logistic or Poisson model. For example, you might discuss the log link function – why this model is appropriate in this case and how the coefficients will be converted to meaningful units (e.g., predicted probabilities). This would also be a good place to describe for the reader the formula that will be used for those conversions from log-odds or event ratios to predicted probabilities or event counts.
* In some cases the methods being used to examine the data can be a contribution to the literature in and of itself. In these cases it is important to draw the readers’ attention to why this is the case and what the evidence contribute to our understanding of the world and what information can be provided to researchers and policymakers as a result.
1. Results
* In the text you do not need to discuss all significant coefficients. Focus first and foremost on the findings directly related to your hypotheses/RQs. Use a table to provide the reader results from all of the coefficients. This will allow them access to the information without bogging them down with unnecessary text. Brevity is important because length is often an issue and because extra text distracts the reader from the *main point* of your analysis.
* Variance components are interesting, but often do not rise to the level of being substantively important and are often not a primary focus of results sections (although in some cases they can be). Instead, consider discussing the results of variance components in the methods section when describing modeling decisions – such as when to fix or free a slope – or include the variance components in the table of the results. Once in the table the variance components can be referred to and discussed in just one or two sentences.
* Also, with the variance components available to the reader in a table it is easier for them to follow your calculations about variance that you want to convey in the text. For example, you might choose to present the results of a conditional model in one table and include the variance components from that model in the lower portion of the table. In the same table you could also report the variance from the fully unconditional model. In that way you could calculate for the reader the amount of variance explained through the prediction model (i.e., fully unconditional variance – prediction model variance/ fully unconditional model variance).
* We should focus less on the ICC and more on the percent of variance explained by our models in our write-ups. In general, the ICC is really only applicable for the Fully Unconditional Model – beyond that we are really more interested in the variance we can explain and in the significance level of our predictors.
* With respect to reporting significant covariates – we need to be clear about what to report in terms of coefficients and standard errors. Typically, we only report coefficient and SE because with sufficient power a t-ratio (coeff/se) of 1.96 or higher is significant at p<.05. So, an educated reader really just wants to see the coefficient and SE in the table so they can check for a coefficient twice the size of the SE.
* With respect to calculating effect sizes --- the basic formula is the “effect of the IV” (a 1 unit change in X) divided by “a standard deviation of the DV” (for now just use the SD from SPSS) – this would typically be reported earlier in a descriptives table. In the case of a logistic or Poisson model the ES is not as important as interpreting your log odds and event ratios correctly. The whole point of the exercise is to help the reader interpret what effects are not only statistically significant but what effects are also meaningful to think about and perhaps craft interventions or policy around.
1. Conclusions/Interpretations
* After reporting results it will be important to eventually take a step back and produce for the reader some preliminary conclusions from our analysis.
* Putting these in terms of the strengths and weaknesses of the analysis is good since it helps frame the trustworthiness of the findings for the reader. Thus, a good limitations section is often compelling for reviewers since it convinces them you are not trying to mislead by over-interpreting the results and making claims that the data does not warrant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Rationale**/ | **0 Points** | **1 point** |  **2 Points** |  |
| Theoretical backgroundNote: Length is not an issue | There is NO information/ thought process. | Information/ thought process provided is disconnected from the research and the analysis. | There is information/ thought process that leads reader to the hypothesis/ research questions and analysis conducted. |  |
| **Research Questions** | **0 Points** | **1 Points** |  **2 Points** | **3 Points** |
|  | There are no RQs/hypotheses,  | RQs/hypotheses are too broad and disconnected from the analysis.  | RQ/hypotheses connected to the analysis but too broad  | RQs/hypotheses were direct, concise and connected to the analysis. |
| **Methods** | **0 Points** | **1 Points** |  **2 Points** | **3 Points** |
| **Model write-** **up** | There is no model write-up | Model write-up attempted but numerous inaccuracies and little description of parameters  | Model write-up complete including equations and text write-up demonstrating appropriate understanding of outcome and model parameters | Model write-up complete including equations and text write-up demonstrating appropriate understanding of outcome and model parameters and coordinated with results write-up |
|  **Descriptives**  **Table** | There is no table describing the sample | A table of descriptives is provided but there is no context for the larger sample  | A descriptives table is provided and adequate information about the sample is provided so the reader can understand strengths and limitations of the sample  |  |
| **Results** | **0 Points** | **1 Points** |  **2 Points** | **3 Points** |
|  **Table of**  **model results** | There is no table describing the model results | A table is provided but it is difficult to read and understand | A table is provided which nicely summarizes the fixed and random effects results |  |
|  **Interpretation of**  **findings** | There is no interpretation of findings | There is a text description of the findings but it is difficult to follow and understand and it is disconnected from the table | The text description thoroughly covers the results and is coordinated with the table but it is difficult to find the main hypothesis test | The text description is clear and concise and makes use of and is coordinated with the table of results |
| **Conclusions** | **0 Points** | **1 Points** |  **2 Points** | **3 Points** |
|  | There are no attempts made to connect the research findings to the larger ideas motivating the research | There is an attempt to connect the research findings with the research questions and hypotheses | Conclusions about the research are aligned with the RQs and make use of the findings presented | Conclusions about the research are aligned with the RQs and make use of the findings presented and describe the strengths and limitations of the research as well as future directions |