Psychology 209

Longitudinal Data Analysis and Bayesian Extensions

Fall 2012

Instructor:

Sarah Depaoli (sdepaoli@ucmerced.edu)
Office Location: SSM 312A
Office Phone: (209) 228-4549 (although email will result in a faster response)
Office Hours: Wednesdays 4:00-6:00pm, or by appointment
Course Website: http://faculty1.ucmerced.edu/sdepaoli

Course Time/Location:

MW 6:00-7:15pm, COB 279

Course Description:

Psychology 209 focuses on the analysis of longitudinal data. This course is broken into several main sections of content all dealing with different aspects of longitudinal data analysis.

- First, a general introduction will be provided, which will include discussion on advantages and challenges of longitudinal studies.
- The second section of the course will specifically focus on analysis of variance approaches to longitudinal data analysis (e.g., repeated measures and MANOVA approaches).
- The third section will introduce several methods using mixed-effects regression models for outcomes that are continuous, binary, ordinal, nominal, and for count data.
- The fourth section of this course will introduce methods for analyzing longitudinal data that are part of the general latent variable framework. We will cover models that estimate continuous growth (e.g., growth curve modeling and growth mixture modeling) and categorical growth over time (e.g., latent transition analysis).
- The last section of this course will cover the Bayesian estimation framework as a tool for estimating various longitudinal models. Advantages and reasons for using this framework will be discussed, as well as the basic features of Bayesian estimation (e.g., estimation algorithms and samplers). We will cover various prior distributions needed to estimate longitudinal models. The issue of where priors come from (and various ways to specify them) will also be covered. Finally, we will discuss several advanced-Bayesian topics in relation to longitudinal data analysis, including model specification (and mis-specification), model comparison, and some special topics for longitudinal mixture models (e.g., label switching and problems with priors).

We will be working with several different computer programs throughout this course: M*plus*, R, and WinBUGS (OpenBUGS). You need not have prior exposure to M*plus* or WinBUGS (OpenBUGS), but it is assumed that you have had some prior exposure to R. If you are not familiar with the R programming environment, please see the instructor to discuss supplementary material that you might benefit from before the semester begins;

note that exposure through PSY 202a and 202b is sufficient.

Required Text:

Hedeker, D., & Gibbons, R. D. (2006). *Longitudinal Data Analysis.* Hoboken, NJ: John Wiley & Sons.

Other Required Readings (Instructor will Distribute):

****Please do not be intimidated by this additional reading list. In some cases, we will just be reading a portion of the assigned piece. This list has been incorporated into the course to help you grasp a deeper understanding of the topics while becoming more familiar with methodological literature.*

Collins, L. M., & Lanza, S. T. (2010). *Latent Class and Latent Transition Analysis.* Hoboken, NJ: John Wiley & Sons.

Congdon, P. (2006a). *Bayesian Models for Categorical Data.* Hoboken, NJ: John Wiley & Sons.

Congdon, P. (2006b). *Bayesian Statistical Modelling.* Hoboken, NJ: John Wiley & Sons.

Depaoli, S. (in preparation). The effect of data-driven priors on parameter estimate bias for continuous growth models.

Depaoli, S. (under review). Mixture class recovery in GMM under varying degrees of class separation: Frequentist versus Bayesian estimation.

Depaoli, S. (2012). The ability for posterior predictive checking to identify model misspecification in Bayesian growth mixture modeling. *Structural Equation Modeling.*

Grimm, K. A., & Ram, N. (2009). Nonlinear growth models in Mplus and SAS. *Structural Equation Modeling*, *16*, 676–701.

Jasra, A., Holmes, C. C., & Stephens, D. A. (2005). Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, *20*, 50–67.

Kaplan, D., & Depaoli, S. (2012). Bayesian statistical methods. In T. Little (Ed.), Handbook of quantitative methods (pp. TBD). Oxford: Oxford University Press.

Kaplan, D. & Walpole, S. (2005). A stage–sequential model of reading transitions: Evidence from the early childhood longitudinal study. *Journal of Educational Psychology*, *97*, 551–563.

Muthén, B. O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In L. M. Collins & A. G. Sayer (Eds.), New methods for the analysis of change (pp. 291–322). Washington DC: APA.

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), Handbook of quantitative methodology for the social sciences (pp. 345–368). Newbury Park, CA: Sage Publications.

Muthén, B. O., & Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In G. Fitzmaurice, M. Davidian, G. Verbeke, & G. Molenberghs (Eds.), Longitudinal data analysis (pp. 143–165). Boca Raton: Chapman & Hall/CRC Press.

Nylund, K. (2007). Latent transition analysis: Modeling extensions and an application to peer victimization. Doctoral dissertation, University of California, Los Angeles.

Nylund, K., Asparouhov, T. & Muthén, B. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, *14*, 535–569.

Sinharay, S. (2004). Experiences with Markov chain Monte Carlo convergence assessment in two psychometric examples. *Journal of Educational and Behavioral Statistics*, *29*, 461–488.

Tofighi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelson (Eds.), Advances in latent variable mixture models (pp. 317–341). Charlotte, NC: Information age Publishing.

Prerequisites:

- Successful completion of Psychology 202a/202b (or equivalent) or instructor consent.
- Understanding of ANOVA and regression analysis.
- Basic understanding of matrix algebra operations (material covered in 202b or akin is sufficient).
- Introductory knowledge of structural equation modeling (material covered in 202b or akin is sufficient).
- Basic familiarity with the R programming environment.

Course Goals:

- Build a strong foundation in analyzing longitudinal data.
- Learn how to address various problems that commonly arise with longitudinal data (e.g., missing data and estimation problems).
- Understand the different longitudinal models that can be used with different *types* of data.
- Gain comfort in reading methodological literature (even applied researchers should be comfortable with this by the end of this course).

- Recognize connections between some common longitudinal models.
- Gain exposure to the Bayesian estimation framework.
- Learn about different problematic features of Bayesian estimation of longitudinal models (e.g., label switching).
- Learn about various "model testing" and "model comparison" procedures under the frequentist and Bayesian modeling frameworks.
- Learn basic programming in M*plus*, R, and WinBUGS (OpenBUGS).

Course Outcomes:

You should be able to demonstrate proficient knowledge of longitudinal data analysis and estimation through the homework assignments and a final paper. Specifically, through these assignments you should:

- Be able to identify the most appropriate analysis procedure for a particular longitudinal question.
- Understand the differences/similarities between ANOVA, regression, and SEM longitudinal data analysis approaches.
- Understand assumptions that are made within each modeling framework we discuss.
- Understand some of the drawbacks to longitudinal procedures and how to address or circumvent common problems.
- Be able to understand and manipulate prior distributions on model parameters within the Bayesian framework.
- Be able to run convergence diagnostics for Bayesian estimation algorithms.
- Utilize model constraints to aid in the prevention of label switching within longitudinal mixture models.
- Be able to write code and estimate a variety of models within the M*plus*, R, and WinBUGS (OpenBUGS) programs.
- Be able to produce appropriate plots using all programs covered in this course.
- Understand methodological literature written about frequentist and Bayesian longitudinal data analysis methods.
- Build comfort with the mathematical notation used to represent longitudinal models, the algorithms used to estimate those models, and the statistics/criteria used to assess fit and compare models.

Course Grades:

Grading is based on: seven homework assignments (49%), a take-home final project (51%).

Homework assignments will be graded based on 10 points. The final project will be graded based on 100 points. A weighted linear combination will be taken at the end of the semester to determine the final course grade using the following criteria:

Percentage	Grade
101+	A+
95-100	A
92-94	A-
89-91	B+
85-88	B
82-84	B-
79-81	C+
75-78	C
72-74	C-
69-71	D+
65-68	D
62-64	D-
0-61	F

If a student is on the cusp between two letter grades, classroom participation will determine whether the student receives the higher or lower letter grade.

Assignments:

Assignments will be handed out in class and will incorporate a combination of hand calculations, data analysis, and conceptual problems. All datasets for these assignments will be provided by the instructor. You will typically have one week to complete these assignments. Assignments are due at the beginning of class on the date specified.

Final Paper:

You will be expected to write a paper in APA style using some of the techniques that you have learned this semester. I encourage you to use your own data if you have access to longitudinal data. If you do not have access, I will provide you with some resources where you can obtain free longitudinal data sets. The paper will be comprised of an introduction where you substantively justify your variable(s) and model(s). You will need to write a complete methods section which describes your variable(s) and design–this needs to also include a (brief) description of the frequentist and/or Bayesian methods that you will employ. This methods section will be followed by a results section where you will describe any relevant diagnostics and results based on frequentist model estimation and/or results based on Bayesian model estimation; you should include any relevant figures and tables as well. Finally, you will need to write a discussion section tying your findings and your introduction together. The final product that you hand in should be a manuscript of (near) publishable quality. I recognize that it is perhaps unreasonable to request a manuscript ready for submission at the end of a one-semester course, however this paper should be of the quality that only minor tweaking and/or expanding of thoughts are needed before submission. Grading will be based on several different components, including substantive development of the model, handling of model diagnostics/fit, correct interpretation of model results, description of estimation processes, and APA style. I will also grade on the professionalism and technical aspects of your writing. This assignment will be due on the day of our scheduled final. I will not "pre-read" any assignments (i.e., I am not "grading" your paper multiple times throughout the course), but I am more than happy to discuss your papers with you (e.g., direction of analysis, aspects of the introduction, etc.) at any point during the semester. More details on this assignment will be provided as the semester progresses. Specifically, I will hand out a more detailed assignment sheet which has due dates for certain "milestones" for the assignment (e.g., a date you need to hand in a short summary of your data source and topic, etc). (*Note: If you have missing data in your final project data, please see me early in the semester since we will not directly cover this topic in class until the end of the term.*)

Academic Integrity:

Students should be familiar with University policies on academic honesty. A general code of conduct for the University of California can be found at:
http://www.ucop.edu/ucophome/coordrev/ucpolicies/aos/uc100.html.
Basically, do not cheat or plagiarize. This will earn you a very uncomfortable meeting with me and a zero on the assignment.

- Students are encouraged to work together on computational aspects of the homework assignments using the various software/programming languages we will work with. However, it is expected that you work independently on hand-calculation, discussion, and interpretation portions of the assignments. The words you submit in your written assignments should be entirely your own. Likewise, I expect that the take-home project will be done independent of your classmates. If you have questions/issues arise, please come see me.

Late Assignments:

Late assignments will not be accepted and will result in a grade of zero unless a "reasonable justification" can be made to me. Acceptable reasons for late work might be that you could not turn in the assignment due to illness (which may require a doctor note) or due to conference or other professional travel. I will never accept the excuses of "I was too busy" or "I forgot". We are all busy and deadlines are important to respect. If you cannot turn in an assignment on time, please come and talk to me about it as soon as possible.

Course Outline:

**Note that I have never taught this course before so *nothing* on this schedule is set in stone.

Week 2:

8/27: Introduction, advantages, challenges, the simplest longitudinal models

- Reading: H&G Chapter 1

8/29: Analysis of variance approaches

- Reading: H&G Chapter 2

Week 3:
9/3: Labor Day, NO CLASS
9/5: Analysis of variance approaches cont.

- Assignments: HW 1 handed out, Term paper instructions handed out

Week 4:
9/10: Analysis of variance approaches cont.
9/12: Multivariate analysis of variance approaches

- Reading: H&G Chapter 3

Week 5:
9/17: Intro to mixed-effects regression models

- Reading: H&G Chapter 4
- Assignments: HW 1 due; HW 2 handed out

9/19: Mixed-effects regression models for continuous outcomes
Week 6:
9/24: MRMs cont., Introduction to the EM algorithm

- Assignments: HW 2 due

9/26: EM algorithm cont.
Week 7:
10/1: Finish EM algorithm, Mixed-effects polynomial regression models

- Reading: H&G Chapter 5
- Assignments: Term paper abstract due, HW 3 handed out

10/3: Mixed effects polynomial regression models cont., Binary outcome MRMs

- Reading: H&G Chapters 5, 9

Week 8:
10/8: Binary and Ordinal outcome MRMs cont.

- Reading: H&G Chapters 9-10

10/10: Some specific autoregressive models

- Reading: H&G Chapter 7; Nylund (2007) [select pages to be assigned]
- Assignments: HW 3 due

Week 9:
10/15: Growth curve modeling

- Reading: Muthen (2001); Muthen (2004); Muthen and Asparouhov (2008)

10/17: Intro to finite mixture modeling, growth mixture modeling

- Assignments: HW 4 handed out

Week 10:

10/22: The "correct" number of mixtures, confirmatory versus exploratory models

- Reading: Nylund, Asparouhov, & Muthen (2007); Tofighi & Enders (2008)

10/24: Non-linear latent growth models

- Reading: Grimm & Ram (2009)

Week 11:

10/29: Categorical LVMs, intro to LCA (not longitudinal, but necessary)

- Reading: Collins and Lanza (2010) [chapter 7]; Nylund (2007) [select pages to be assigned]
- Assignments: HW 4 due

10/31: Latent transition analysis and other important Markov models

- Reading: Kaplan & Wapole (2005)
- Assignments: HW 5 handed out (11/1)

Week 12:

11/5: Introduction to Bayesian estimation (estimation algorithm, samplers, priors)

- Reading: Kaplan & Depaoli (in press)

11/7: Introduction to Bayesian estimation cont, convergence diagnostics

- Reading: Sinharay (2004)
- Assignments: HW 5 due (11/9)

Week 13:

11/12: Veteran's Day, NO CLASS
11/14: Finish convergence, WinBUGS: "Someone save us all!!"

- Assignments: HW 6 handed out

Week 14:

11/19: Mommy, where do priors come from??

- Reading: Depaoli (under review); Depaoli (in preparation)

11/21: "Chopping Carrots" Day: NO CLASS

Week 15:

11/26: Bayesian longitudinal analysis

- Assignments: HW 6 due, HW 7 handed out

11/28: Bayesian model comparison and fit:
"All models are wrong"...and so are the model assessment procedures!

- Reading: Depaoli (2012), Kaplan and Depaoli (2012)

Week 16:

12/3: Bayesian model comparison and fit, Label Switching intro

- Reading: Jasra and Stephens (2005)

12/5: Problems: within- and between-chain label switching (mixture and non)

- Assignments: HW 7 due

FINAL EXAM WEEK

12/13: Scheduled Final Exam Day

- Assignments: Final paper due by 6:00pm!!