Statistics 618/SPA 696: Bayesian Statistics for Social and Biomedical Sciences, Fall 2018 Seminar

Wednesday, 8:10-11:00 AM, Kerwin Hall 107.

Course Description: Principles and applications of modern statistical decision theory, with a special focus on Bayesian modeling, data analysis, inference, and optimal decision making. Prior and posterior; comparison of Bayesian and frequentist approaches, including minimax decision making and elementary game theory. Bayesian estimation, hypothesis testing, credible sets, and Bayesian prediction. Introduction to Bayesian computing software and applications to diverse fields. Grading: A-F only. Prerequisite: STAT-514 or permission of instructor.

Learning Outcomes: By the end of this course, students will be able to:

1. *Demonstrate* a basic understanding of Bayesian model specification, Bayesian posterior inference, and model assessment and comparison.
2. *Use* this understanding of Bayesian statistics to specify and estimate Bayesian multilevel (hierarchical) models with linear and nonlinear outcomes, treat missing data in a principled and correct manner using multiple imputation, gain facility in the R and bugs statistical languages, know how to compute the appropriate sample size and power calculations for Bayesian models, gain exposure to Bayesian approaches including MCMC computation, and be able to assess model reliability and fit in complex models.
3. *Apply* this understanding of Bayesian statistics to data in the social and biomedical sciences.
4. *Convey* analytical results from these models to both lay and technical audiences clearly in both writing and speech.

Prerequisite Details: This course assumes a knowledge of basic statistics as taught in a first year undergraduate or graduate sequence. Topices should include: probability, cross-tabulation, basic statistical summaries, and linear regression in either scalar or matrix form. Knowledge of R, basic matrix algebra and calculus is helpful.

Course Grade: The final grade will be based on two components: weekly attendance and participation (20%) and exercises (80%). Graduate students will have one additional component of their exercise grade that constitutes 30 points out of the 80 points total: submission of an analysis of real research using a multilevel model applied to data in their field along with 5-10 pages of discussion to include a description of the data, model diagnostics, and the subsequent findings. Consider this assignment to be the start of a research manuscript to be eventually submitted to a an academic journal. Graduate students will still submit all exercises assigned below in addition to this work.

Office Hours: By appointment.

Incompletes: Due to the scheduled nature of the course, no incompletes will be given.

Teaching Assistant: Simon Heuberger, sh6943a@american.edu. Office Hours: Friday 2-4PM, Kerwin Hall 109-J.

Required Reading: Gelman and Hill, "Data Analysis Using Regression and Multilevel/Hierarchical Models" (Cambridge University Press 2007). Some papers will be available at jstor.org or distributed by the instructor. Readings should be completed before class.

Topics (subject to minor change):
August 29: No class meeting (conference commitment). Reading: R for Beginners (to make sure you are fluent on the basics).

September 5: Introducing Bayesian Inference. Reading: Gelman & Hill, Chapters 1 and 2, MLE Review , Intro code from the lecture, Bayesian mechanics slides.

September 12: Linear Model Theory Review. Reading: Gelman & Hill, Chapters 3 and 4, Chapter 3-4 code from the lecture, Binomial PMF likelihood grid search, lecture slides (do not print!). Anaemia data. Exercises due: Gelman & Hill 2.2, 2.3, 2.4.

September 19: Multilevel Structures and Multilevel Linear Models: the Basics. Reading: Gelman & Hill, Chapters 11 and 12, Introductory Chapter (Gill and Womack, from the SAGE Handbook of Multilevel Modeling). Chapter 11-12 code from the lecture. Exercises due: Gelman & Hill 3.4, 4.4, 5.4, 6.1.

September 26: Multilevel Linear Models: Varying Slopes, Non-Nested Models and Other Complexities. Reading: Gelman & Hill, Chapter 13, Chapter 13 code from the lecture. Exercises due: Gelman & Hill 11.4, 12.2, 12.5.

October 3: Multilevel Logistic Regression, Multilevel Generalized Linear Models. Reading: Gelman & Hill, Chapter 14 (skip Section 14.3), Chapter 15, Chapter 14 code from the lecture. Exercises due: Gelman & Hill 13.2, 13.4, 13.5.

October 10: Multilevel Modeling in Bugs and R: the Basics, MCMC Theory. Reading: Gelman & Hill, Chapter 16, Bayesian Estimation Case Study (Gill and Witko 2012), R to JAGS code for the model (get data from here), Chapter 16 code from the lecture. Exercises due: Gelman & Hill 14.5, 14.6, 15.1, 15.2.

October 17: Fitting Multilevel Linear and Generalized Linear Models in Bugs and R, MCMC Coding. Reading: Gelman & Hill, Chapter 17, Chapter 17 code from the lecture. Exercises due: Gelman & Hill 16.1, 16.2, 16.3, 16.8.

October 24: TBD.

October 31: Likelihood and Bayesian Inference, Computation, MCMC Diagnostics and Customization. Reading: Gelman & Hill, Chapter 18, Chapter 18 code from the lecture. Exercises due: Gelman & Hill Rerun 16.3 using the instructions in 17.2 and 17.3, 17.5.

November 7: Treatment of Missing Data. Reading: Gelman & Hill, Chapter 25, Paper by van Buuren and Groothuis-Oudshoorn, Chapter 25 code from the lecture. Exercises due: Gelman & Hill 18.1, 18.2, 18.4.

November 14: Understanding and Summarizing the Fitted Models, Multilevel Analysis of Variance. Reading: Gelman & Hill, Chapter 21 and 22, Chapter 21 and Chapter 22 code from the lecture. Exercises due: missing data problems.

November 21: Thanksgiving Holiday.

November 28: Model Checking and Comparison. Reading: Gelman & Hill, Chapter 24, Chapter 24 code from the lecture. Exercises due: 21.1, 21.3, 21.4, 22.1.

December 5: Sample Size and Power Calculations. Reading: Gelman & Hill, Chapter 20, Chapter 20 code from the lecture. Exercises due: 24.1, 24.4.