PittPublicHealth

BIOST 2063
Bayesian Data Science
Spring 2019

Instructor:	Robert Krafty, PhD
Email:	rkrafty@pitt.edu
Office:	7125 Public Health
Office Hours:	Wednesdays, 2:00 – 3:00pm
Lecture Times:	Mondays and Wednesdays, 11:30am – 12:55pm
Lecture Location:	A216 Public Health
Teaching Assistant:	Shu Wang
Email:	shw97@pitt.edu
Office:	A443 Public Health
Office Hours:	Tuesdays, 1:00 – 2:00pm

Description:

This is a course in Bayesian methods for applied statistics and data science whose broad goal is to provide students with the skills needed to be able to select, conduct, report and interpret appropriate Bayesian analyses for a wide variety of applied problems. General topics covered include Bayesian concepts of statistical inference, Markov chain Monte Carlo and other computational methods, linear, hierarchical and generalized linear models, model selection and diagnostics, and Bayesian learning. The course explores the use of the popular and free software packages R, JAGS and Stan for conducting Bayesian analyses.

Specific Course Objectives

At the end of this course, a student should be able to:

1. Describe fundamental concepts of Bayesian inference.
2. Identify Bayesian analyses appropriate for addressing specific research questions.
3. Describe fundamental concepts of Markov chain Monte Carlo and other popular sampling techniques.
4. Use R, JAGS and Stan to conduct Bayesian analyses.
5. Report and interpret results of Bayesian analyses.

Credit hours: 3

Prerequisites

The prerequisites are:

1. A passing grade in or exemption from BIOST 2041 or BIOS 2039.
2. Either a passing grade in, exemption from, or current enrollment in BIOS 2049.

Students who do not meet these prerequisites but who have mastered the material in these courses through other courses will be able to register with the instructor's permission after providing copies of appropriate transcripts and syllabi.

Required Textbook

- John K. Kruschke,
  Doing Bayesian Data Analysis: A Tutorial with R, JAGS and Stan,
  2nd edition, Academic Press, 2015.
  [Note that the first edition can be accessed electronically for free through the University Library. However, the course will be based on the second edition, as it more comprehensively discusses modern computing tools and high-quality programs.]

Optional/Supplementary Textbooks

- Michael Crawley,
  The R Book,
  2nd edition, Wiley, 2012
  [A comprehensive and popular reference for R. It can be accessed electronically for free through the University Library.]

- Emmanuel Lesaffre and Andrew B. Lawson,
  Bayesian Biostatistics,
  Wiley, 2012
  [More advanced than Kruschke; a good resource for students who desire moderately more mathematical motivation. It can be accessed electronically for free through the University Library.]

- Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari and Donald B. Rubin,
  Bayesian Data Analysis,
  3rd edition, CRC Press, 2014.
  [Arguably the most popular reference for Bayesian data analysis. It can be accessed electronically for free through the University Library.]

Software and Computing

The course will make extensive use of R, RStudio, TeX/LaTeX, JAGS and Stan, all of which can be downloaded for free:

- R – https://cran.r-project.org/
- RStudio - https://www.rstudio.com/
- MikTeX - https://miktex.org
- JAGS - http://mcmc-jags.sourceforge.net/
- Stan - http://mc-stan.org/
- R interface to Stan - http://mc-stan.org/rstan/

Students should have access to a laptop or desktop computer that has access to the aforementioned programs. A student computer lab with a limited number of Windows

machines is located in Public Health A438.

CourseWeb

CourseWeb (http://courseweb.pitt.edu) will be used to distribute course notes, sample code, data, homework assignments, turn in assignments, display grades and make announcements.

Grading

Course grades will be based on the weighted average of:

- Homework   50%
- Project   50%

Course grades will be assigned based on the following minimal scale. Note that this is a minimal scale, and, in certain circumstances, grades can be curved up.

[93%, 100%]   A;   [90%, 93%), A-;
[87%, 90%)   B+;   [83%, 87%) B;   [80, 83) B-;
[77%, 80%)   C+;   [73%, 77%) C;   [70, 73) C-;
[67%, 70%)   D+;   [63%, 67%) D;   [60, 63) D-;
< 60%   F

Homework:

Homework will be assigned approximately weekly. Assignments will be posted on and students will submit solutions through CourseWeb. Assignments will be graded based on appropriateness of analyses selected, technical execution, reproducibility, interpretation of results and clarity of reporting.   Students are required to use R Markdown to complete assignments and submit both (1) a written report in PDF format knitted in R Markdown and (2) the R Markdown RMD file used to generate the PDF report and that can be used to replicate results.

Project:

Students will be required to complete a course project in collaboration with a small group of fellow students.   The project can either be (1) a comprehensive data analysis of a real-world data set using techniques discussed in class or (2) an exploration of a method that was not covered in class.   Students will be required to submit a prospectus (3 pages or less), give an in-class presentation (15 minutes), submit a written report, and provide a well-documented script file and all necessary data to allow results to be replicated.   The grade for the project will be based on a weighted average of:

- Prospectus   10%
- Presentation   35%
- Write-up   35%
- Reproducibility   20%

The dates for milestones related to the project are:

Wednesday,	February 20	Groups are assigned
Wednesday,	March 13	Prospectus' are due
Wednesday - Wednesday,	April 10, 15, 17	Presentations
Wednesday,	April 24	Write-up due

Important Administrative Dates:

Monday,	January 7	First day of class
Friday,	January 18	Add/drop period ends
Monday,	January 21	No class (MLK Day)
Monday,	March 11	No class (Spring Break)
Wednesday,	March 13	No class (Spring Break)
Wednesday,	March 14	Deadline to withdraw
Wednesday,	April 24	Last day of the course

Specific Topics

Note that the exact topics and number of lectures per unit are subject to change.

Topic	# of Lectures
Unit 1: Introduction and Preliminaries	
Broad concepts, Software	1
Probability, Bayes Rule (Chapters 4 & 5)	1
Unit 2: Fundamentals for a Binomial Probability	
Exact Inference (Chapter 6)	1
Concepts of and Using MCMC (Chapters 7 & 8)	2
Hierarchical Models (Chapters 9 & 10)	2
Testing, Power and Sample Size (Chapters 11, 12 & 13)	2
HMC and Stan (Chapter 14)	1
Unit 3: The Generalized Linear Model	
Introduction and Overview (Chapter 15)	1
One and Two Groups of Continuous Outcomes (Chapter 16)	1
Regression for Continuous Outcomes (Chapters 17 & 18)	2
Variable Selection	2
ANOVA and Variants (Chapters 19 & 20)	2
Categorical/Nominal Outcomes (Chapters 21 & 22)	2
Ordinal Outcomes (Chapter 23)	1
Count Outcomes (Chapter 24)	1
Unit 4: Student Presentations	3

Accommodation for Students with Disabilities:

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, 140 William Pitt Union, 412-648-7890 as early as possible in the term.

Diversity Statement:

The University of Pittsburgh Graduate School of Public Health considers the diversity of its students, faculty, and staff to be a strength and critical to its educational mission. Pitt Public Health is committed to creating and fostering inclusive learning environments that value human dignity and equity. Every member of our community is expected to be respectful of the individual perspectives, experiences, behaviors, worldviews, and backgrounds of others. While intellectual disagreement may be constructive, no derogatory statements, or demeaning or discriminatory behavior will be permitted.

If you feel uncomfortable or would like to discuss a situation, please contact any of the following:
- the course instructor;
- the Pitt Public Health Associate Dean for Diversity at 412-624-3506 or nam137@pitt.edu;
- the University's Office of Diversity and Inclusion at 412-648-7860 or https://www.diversity.pitt.edu/make-report/report-form (anonymous reporting form).

Academic Integrity Statement:

All students are expected to adhere to the school's standards of academic honesty. Cheating/plagiarism will not be tolerated. The Graduate School of Public Health's policy on academic integrity, which is based on the University policy, is available online in the Pitt Public Health Academic Handbook www.publichealth.pitt.edu/home/academics/academic-requirements. The policy includes obligations for faculty and students, procedures for adjudicating violations, and other critical information. Please take the time to read this policy.

Sexual Misconduct, Required Reporting and Title IX:

The University is committed to combatting sexual misconduct. As a result, you should know that University faculty and staff members are required to report any instances of sexual misconduct, including harassment and sexual violence, to the University's Title IX office so that the victim may be provided appropriate resources and support options. What this means is that as your professor, I am required to report any incidents of sexual misconduct that are directly reported to me, or of which I am somehow made aware. There are two important exceptions to this requirement about which you should be aware:

- A list of the designated University employees who, as counselors and medical professionals, do not have this reporting responsibility and can maintain confidentiality, can be found here: www.titleix.pitt.edu/report/confidentiality.
- An important exception to the reporting requirement exists for academic work. Disclosures about sexual misconduct that are shared as part of an academic project, classroom discussion, or course assignment, are not required to be disclosed to the University's Title IX office.

If you are the victim of sexual misconduct, Pitt encourages you to reach out to these resources:
- Title IX Office: 412-648-7860
- SHARE @ the University Counseling Center: 412-648-7930 (8:30 A.M. TO 5 P.M. M-F) and 412-648-7856 (AFTER BUSINESS HOURS).

If you have a safety concern, please contact the University of Pittsburgh Police, 412-624-2121. Other reporting information is available here: www.titleix.pitt.edu/report-0