NYU | SCHOOL OF GLOBAL PUBLIC HEALTH

GPH-GU 2372/3372 Applied Bayesian Analysis in Public Health

Class Schedule: Monday, 6:45 PM-9:25 PM ET
Class Location: GCASL, Room 275
Semester and Year: Fall 2022

Professor: Hai Shu
Phone: 212-992-3811
Email: hs120@nyu.edu
Course Assistant: Sooyoung Kim
Course Assistant Email: sk9076@nyu.edu

Office: 708 Broadway, Room 759
Office Hours: Monday, 6:05 PM-6:35 PM ET

COURSE DESCRIPTION:
Bayesian analysis is one of the two major statistical paradigms; the other is Frequentist analysis. The course will briefly review the theory behand Bayesian methods and will focus on the practical implementation to public-health and biomedical data. Topics include comparison of Bayesian and Frequentist analyses, Bayesian inference of various one-parameter models and normal models, Markov Chain Monte Carlo algorithms, Bayesian (generalized) linear regression models, and Bayesian hierarchical models. Data analysis with the R software will be emphasized in the course. Upon successful completion of the course, students will be able to formulate Bayesian models for data analysis in public health and biomedicine, and will be able to implement the Bayesian inference using R.

COURSE OVERVIEW:
This course complements other introductory statistical courses by providing a Bayesian view into statistical inference and relating it to public health and biomedical problems. The course will render systematic training in Bayesian analysis with extensive examples and practice for analyzing public health and biomedical datasets throughout teaching materials, assignments and exams. Real-data topics will include but be not limited to diabetes, cancer, nutrition, and aerobics. Students will learn and practice in R programming for Bayesian data analysis with popular R packages such as *MCMCpack* and *coda*. This applied course will be a good preparation for advanced courses and for work in environments that use Bayesian analysis.

COURSE FORMAT:
3-credit course: This course will be offered in person. In total, students will receive 2,240 minutes (14 class sessions) of live in-person instruction and a 30-minute in-person office hour before each class session.

COURSE LEARNING OBJECTIVES AND RELATED COMPETENCIES AND COMPONENTS:

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

Learning Objective	**Competency**[1]	**Course component (lesson # & topic, assignment, etc.)**
1. Explain the difference between Bayesian and Frequentist analyses	Apply descriptive and inferential methodologies according to the type of study design for answering a particular research question. (BC-1) Harness basic concepts of probability, random variation and commonly used statistical probability distributions. (BC-2)	Lesson 1 HW 1
2. Apply appropriate Bayesian models to analyze public health and biomedical datasets and interpret the results	Apply descriptive and inferential methodologies according to the type of study design for answering a particular research question. (BC-1) Interpret results of statistical analyses found in public health research studies. (BC-6) Apply appropriate statistical methods and statistical software for optimal **analysis** of public health and biomedical research studies. (BIOS-3) Provide appropriate statistical **interpretation** of the results of data analyses of public health and biomedical studies. (BIOS-4)	Lessons 2, 3, 6, 7, 8, 10 & 11 All homework assignments Midterm exam Final exam Final individual project (for PhD students)
3. Apply appropriate prior distributions and conduct the posterior inference for Bayesian models based on data	Apply descriptive and inferential methodologies according to the type of study design for answering a particular research question. (BC-1) Harness basic concepts of probability, random variation and commonly used statistical probability distributions. (BC-2) Apply appropriate statistical methods and statistical software for optimal **analysis** of public health and biomedical research studies. (BIOS-3)	Lessons 2, 3, 6, 7, 8, 10 & 11 All homework assignments Midterm exam Final exam Final individual project (for PhD students)
4. Apply R software to implement Monte Carlo algorithms and Bayesian models for data analysis	Utilize relevant statistical software for data analysis. (BC-7) Apply appropriate statistical methods and statistical software for optimal **analysis** of public health and biomedical research studies. (BIOS-3)	Lessons 2 to 11 All homework assignments Final exam Final individual project (for PhD students)

[1] BC = Biostatistics Concentration Competency (MPH), BIOS = Biostatistics Concentration Competency (Doctorate).

PRE-REQUISITES:

- GPH-GU 2353 (Regression I: Linear Regression and Modeling) or equivalent.
- GPH-GU 2184 (Intermediate Statistical Programming in R) or equivalent.
- Calculus is recommended. Students who have not taken a Calculus course should contact the instructor before enrolling and will be directed to online materials.

COURSE REQUIREMENTS AND EXPECTATIONS:

Class attendance (10%): Students are expected to attend all lecture sections. Students are expected to come to class on time to prevent disrupting the lecture and classroom activities. Attendance will be taken at each class meeting and the percentage of classes attended will comprise the class participation portion of the grade. Excused absences are granted if student discusses with instructor prior to class.

Homework Assignments (30%): There will be 6 graded assignments. The assignments provide opportunities for students to understand Bayesian concepts and to implement the associated techniques using R. Homework assignments must be uploaded to the NYU Brightspace site **by 11:59 PM (US eastern time) on the due date.** Each assignment will be graded on a 100 point scale. There is a **20 point deduction for each day** an assignment is late (rounded up to a day[2]). Students are encouraged to discuss and work on the homework together but must hand in their own work, including their own computer programs and output. Do not copy code or any other work. Note that changing the font on someone else's computer work does not count as "your own". PhD students will have one extra question in each assignment.

Midterm Exam (25%): You will have the complete class length to complete the midterm examination. The exam is open book (and notes). The exam covers material through Lesson 6. The exam will consist of a set of short answer questions in which students are asked to explain, apply, interpret, and evaluate the concepts that have been taught. This will be a paper and pencil exam. There will be no computer work. PhD students will have an additional set of problems in the Midterm Exam.

Final Exam (35% for Master's students, 20% for PhD students): There will be a take-home final exam. The final exam will be released in the last day of class and must be uploaded to the NYU Brightspace site **by 11:59 PM (US eastern time) on the due date.** The final exam will be graded on a 100 point scale. There is a **10 point deduction for each hour** the final exam is late (rounded up to an hour[3]). The final exam will comprise 35% of the final grade. The exam is open book (and notes). The final exam will cover all the concepts covered during the course with a particular emphasis on the material we cover after the midterm. The final exam will consist of a set of short answer questions in which students are asked to explain, apply, interpret, and evaluate the concepts that have been taught. This will include questions for which each student will undertake data analyses in R and write up the results. Students are not allowed to discuss about the final exam with other classmates or other fellow students. Do not copy code or any other work. Note that changing the font on someone else's computer work does not count as "your own". PhD students will have an additional set of problems in the Final Exam.

Final Individual Project (only for PhD students: 15%): There will be a take-home individual project on analyzing a real-world public health dataset (a set of datasets will be curated for this purpose). Each PhD

[2] For example, if an assignment is 1 day and 5 minutes late, 40 points will be automatically deducted.
[3] For example, if the final exam is 1 hour and 5 minutes late, 20 points will be automatically deducted.

student is required to analyze the data with R using one or more Bayesian methods and models encountered in the course, and write a report about the Bayesian data analysis. The written report should contain sections including introduction, data description, model, results, and discussion. Detailed instructions will be given on Brightspace. The final individual project must be uploaded to the NYU Brightspace site **by 11:59 PM (US eastern time) on the due date.** Late submission will not be accepted.

Classroom environment: Ideally, everyone should be involved in classroom discussions. In order for everyone to feel comfortable presenting work and voicing opinions, questions, and suggestions, a climate of tolerance and respect is essential.

Course Syllabus Subject to Change: Every effort will be made to follow the syllabus content and schedule. If circumstances dictate, there may be modifications made during the semester and every effort will be made to notify students in a timely manner.

Statistical Software: This course uses R and RStudio. They are available for free at https://www.r-project.org/and https://rstudio.com, respectively.

Readings: Students should complete the readings before each class session.

ASSIGNMENTS:

There will be 6 graded homework assignments. Homework assignments can be found on NYU Brightspace in the Assignments section. Some homework questions will come from the textbook and others will relate to published articles or other public health scenarios. Homework assignments will be graded. Homework assignments should be submitted through Brightspace **by 11:59 PM (US eastern time)** on the due date. Each assignment will be graded on a 100 point scale. There is a **20 point deduction for each day** an assignment is late. All assignments must be neatly written or typed on unlined white paper. The student's name should appear at the top corner of each page and each page should be numbered. Problems should be submitted in the order in which they were assigned and numbered according to the assignment. All graphs should be properly labeled and only relevant R output should be included in homework. Homework solutions will be posted to Brightspace. Students are encouraged to discuss and work on the homework together but must hand in their own work, including their own computer programs and output. Do not copy code or any other work. Note that changing the font on someone else's computer work does not count as "your own". Detailed instructions and questions will be given to students for each homework assignment when it is assigned. Public health and biomedical datasets will be provided to students for use in the assignments. The 6 assignments will count equally, which will comprise 30% of the final grade.

Homework Assignment	Description	Due Date
HW 1: One-parameter models	For given data examples and one-parameter models, compute and plot the posterior distributions with conjugate priors, show the posterior mean, variance and confidence interval of the model parameter, and find and plot the posterior predictive distributions. Compare the results with the Frequentist's maximum likelihood estimates.	9/26
HW 2: The normal model and Monte Carlo approximation	Given a dataset, use Monte Carlo methods to conduct posterior comparison between the	10/11

	noninformative prior, conjugate prior, and semi-conjugate prior for the normal model.	
HW 3: The Markov Chain Monte Carlo method and the multivariate normal model	Given the HW2's dataset, implement the Gibbs sampling for posterior inference on the normal model with the semi-conjugate prior and compare with the results in HW2. For a multivariate version of the HW2's dataset, implement the Gibbs sampling for posterior inference on the multivariate normal model.	10/31
HW 4: Bayesian linear regression	Given a dataset, fit the Bayesian linear regression, compare the linear models yielded from different model selection methods, and interpret the results.	11/14
HW 5: The Metropolis-Hastings algorithm and hierarchical regression model	Given a dataset with a count outcome, fit an appropriate (generalized) linear mixed model by the Metropolis-Hastings algorithm, and interpret the findings.	12/5
HW 6: Probit regression	Fit and compare the logistic and probit regression models on a given dataset, and interpret the findings.	12/12

GRADING COMPONENTS:

Item:	**Percentage or Points:**
Class attendance	10%
Homework assignments (6)	30%
Midterm exam	25%
Final exam	35% for Master's students 20% for PhD students
Final individual project (only for PhD students)	15%

GRADING SCALE:

A:	94-100	C+:	77-79
A-:	90-93	C:	73-76
B+:	87-89	C-:	70-72
B:	83-86	D+:	67-69
B-:	80-82	D:	60-66
		F:	<60

BRIGHTSPACE:
Brightspace will be used extensively throughout the semester for assignments, announcements, and communication. Brightspace is accessible through at https://home.nyu.edu/academics

TECHNOLOGY POLICY:
[Clarification on use laptops, cell phones, etc. in the classroom. An example is below, but you are free to set your own policy:
Mobile device (e.g., smart phones, pagers, etc.) ringers will be turned off or placed on vibrate prior to class. Laptops and tablets can be used in the classroom to take notes, make calculations, and download/read course materials. Research suggests that non-academic use of the internet is associated with poorer learning outcomes.]

COURSE OUTLINE:

Date	Topics	Readings/Materials Due	Assignments Due
9/12	**Lesson 1: Introduction to Bayesian Inference** • Introduction to uncertainty and modelling • Basics of probability theory • Bayes' theorem, and prior & posterior distributions • Bayesian vs. Frequentist	Chapters 1&2 in Hoff (2009)	
9/19	**Lesson 2: One-parameter models** • Binomial model • Poisson model • One-parameter Exponential family models • The prior & posterior distributions and prediction of the above models • R implementation	Chapter 3 in Hoff (2009)	
9/26	**Lesson 3: The normal model** • Bayesian inference with known variance • Bayesian inference with known mean • Bayesian inference with unknown mean and variance • The prior & posterior distributions and prediction of the above cases. • R implementation	Chapter 5 in Hoff (2009)	HW 1
10/3	**Lesson 4: Monte Carlo approximation** • The Monte Carlo method	Chapter 4 in Hoff (2009)	

	• Sampling from predictive distribution • Posterior predictive model checking • Inference for non-closed form posterior distribution • R implementation		
10/11	**Lesson 5: Markov Chain Monte Carlo (MCMC)** • Sampling from the conditional distributions • Gibbs sampling • Monte Carlo vs. MCMC • MCMC diagnostics • R implementation	Chapter 6 in Hoff (2009)	HW 2
10/17	**Lesson 6: The multivariate normal model** • The multivariate normal density • Bayesian inference with known covariance matrix • Inverse Wishart distribution • Bayesian inference with unknown covariance matrix • R implementation	Chapter 7 in Hoff (2009)	
10/24	**Midterm Exam (in class)**		
10/31, 11/7	**Lesson 7: Linear regression** • Linear regression in a Bayesian context • Model checking for a normal linear regression model • Model selection: Bayesian variable selection, Bayesian model averaging, criterion-based methods • R implementation	Chapter 9 in Hoff (2009)	HW3 (10/31)
11/14	**Lesson 8: Hierarchical modeling** • Hierarchical models and exchangeability • The hierarchical normal model • Shrinkage • R implementation	Chapter 8 in Hoff (2009)	HW 4
11/21	**Lesson 9: Metropolis-Hastings algorithms**	Chapter 10 in Hoff (2009)	

	• Generalized linear models • The Metropolis algorithm • The Metropolis-Hastings algorithm • Combining the Metropolis and Gibbs algorithms • R implementation		
11/28	**Lesson 10: Hierarchical regression models** • Hierarchical linear regression models • Linear mixed model • Hierarchical generalized linear regression models • R implementation	Chapter 11 in Hoff (2009)	
12/5	**Lesson 11: Probit regression** • Logistic regression • Probit regression for binary data • Probit regression for ordinal data • R implementation	Chapter 12 in Hoff (2009)	HW 5
12/12	**Review Session**		HW 6
12/18			Final Exam (take-home)
12/20			Final individual project (only for PhD students)

READING/VIEWING LIST:

Textbook:

The following textbook is required:

- Hoff PD. *A First Course in Bayesian Statistical Methods*. Springer; 2009. Available through SpringerLink at the NYU library.

GPH DIVERSITY, EQUITY, and INCLUSION (DEI) STATEMENT:

The NYU School of Global Public Health (GPH) is committed to maintaining and celebrating a diverse, just, and inclusive environment for our students, faculty, and staff around the world. To foster this atmosphere and ideals of Diversity, Equity, and Inclusion (DEI), GPH promotes a welcoming learning environment that embraces cultural humility, and respects and values differences. These differences can include race, ethnicity, religion, gender identity, sexual orientation, physical, mental and emotional abilities, socioeconomic status, and other aspects of human diversity. In this course, we encourage students to share and discuss different perspectives, beliefs, and experiences while treating all with dignity and respect.

STATEMENT OF ACADEMIC INTEGRITY:

The NYU School of Global Public Health values both open inquiry and academic integrity. Students in the program are expected to follow standards of excellence set forth by New York University. Such standards include respect, honesty and responsibility. The SGPH does not tolerate violations to academic integrity including:

- Plagiarism
- Cheating on an exam
- Submitting your own work toward requirements in more than one course without prior approval from the instructor
- Collaborating with other students for work expected to be completed individually
- Giving your work to another student to submit as his/her own
- Purchasing or using papers or work online or from a commercial firm and presenting it as your own work

Students are expected to familiarize themselves with the SGPH and University's policy on academic integrity as they will be expected to adhere to such policies at all times – as a student and an alumni of New York University.

Plagiarism

Plagiarism, whether intended or not, is not tolerated in the SGPH. Plagiarism involves presenting ideas and/or words without acknowledging the source and includes any of the following acts:

- Using a phrase, sentence, or passage from another writer's work without using quotation marks
- Paraphrasing a passage from another writer's work without attribution
- Presenting facts, ideas, or written text gathered or downloaded from the Internet as your own
- Submitting another student's work with your name on it
- Submitting your own work toward requirements in more than one course without prior approval from the instructor
- Purchasing a paper or "research" from a term paper mill.

Students in the SGPH and SGPH courses are responsible for understanding what constitutes plagiarism. Students are encouraged to discuss specific questions with faculty instructors and to utilize the many resources available at New York University.

Disciplinary Sanctions

When a professor suspects cheating, plagiarism, and/or other forms of academic dishonesty, appropriate disciplinary action is as follows:

- The Professor will meet with the student to discuss, and present evidence for the particular violation, giving the student opportunity to refute or deny the charge(s).
- If the Professor confirms that violation(s), he/she, in consultation with the Chairperson or Program Director may take any of the following actions:
    - Allow the student to redo the assignment
    - Lower the grade for the work in question

- Assign a grade of F for the work in question
- Assign a grade of F for the course
- Recommend dismissal

Once an action(s) is taken, the Professor will inform the Chairperson or Program Director and inform the student in writing, instructing the student to schedule an appointment with the Senior Associate Dean for Academic Affairs, as a final step. The student has the right to appeal the action taken in accordance with the GPH Student Complaint Procedure.

STUDENTS WITH DISABILITIES:

Students with disabilities should contact the Moses Center for Students with Disabilities regarding the resources available to them, and to determine what classroom accommodations should be made available. More information about the Moses Center can be found here: https://www.nyu.edu/life/safety-health-wellness/students-with-disabilities.html. Students requesting accommodation must obtain a letter from the Moses Center to provide to me as early in the semester as possible.