PHP 2514: Applied Generalized Linear Models

Stavroula A. Chrysanthopoulou, PhD

Fall 2025

Instructor: Stavroula A. Chrysanthopoulou, PhD
Email: stavroula_chrysanthopoulou@brown.edu Web: https://canvas.brown.edu

Class Hours: Tue & Thu 2:30 - 3:50pm Office Hours: TBD
Class Room: School of Public Health / TBD **TBD** Office: 7th floor - Room 706

Teaching Assistants:
TBD

Course Description

This course provides a foundation in the statistical theory and practical implementation of Generalized Linear Models (GLM) for analyzing data in clinical and population health research. Students will study applied GLMs for outcome types commonly encountered in public health, including continuous, binary, count, and survival data. Topics will include exploratory analysis, estimation, as well as model building, diagnostics, and predictions. Emphasis is placed on understanding how theoretical assumptions and modeling choices affect real-world data analysis and scientific conclusions. The course requires extensive use of statistical programming to implement GLM methods, with data and examples drawn from medical and pharmaceutical research, public health, and the social sciences. Designed for graduate and advanced undergraduate students, the course equips participants with a practical, hands-on regression toolkit alongside a solid understanding of the theoretical foundations of GLMs. The course also introduces a critical evaluation and important considerations of using AI tools in the application of GLMs to real-world research problems.

Design of Course / Prerequisites

This graduate-level course is designed and will be offered only in person. Students should have prior knowledge of probability and statistical inference (e.g., PHP2510/1510) and regression analysis (e.g., PHP 2511/1511). Familiarity with R or other similar programming language is also advisable. Students who wish to take the course may also contact the instructor to determine whether they have the required knowledge and ask for an override code. The course will move at a quick pace and will focus more on the applications of GLMs as opposed to theoretical derivations.

Required Readings

Agresti, A. (2018), An Introduction to Categorical Data Analysis, 3rd edition,Wiley, ISBN: 978-1-119-40526-9

Frank E. Harrell, Regression Modeling Strategies: With Applications, to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, 2nd ed. Springer Series in Statistics. ISBN: 978-3-319-19424-0

Recommended Readings

Agresti, A. (201), Categorical Data Analysis, 3rd edition,Wiley, ISBN: 978-0-470-46363-5

Dobson, A.J., Barnett, A.G.,(2018). An Introduction to Generalized Linear Models (4th edition). Chapman and Hall/CRC. ISBN-13: 978-1-584-88950-2

Software

Students are at the liberty to use any statistical software with which they feel comfortable for answering homework and exam questions. Lecture notes will include examples of data analysis using R. The book also provides code for several analysis examples using mainly R and STATA.

Students can find and use a variety of free-of-charge statistical software at the Brown University Software Catalog.

Technology

All the material for this course will be shared through Brown Canvas. I am committed to ensuring access to online course resources by students. If you have any concerns or questions about access or the privacy of any of these platforms, please reach out to the IT Service Center (https://it.brown.edu/get-help), which provides many IT Services including remote assistance.

Course Objectives

After successful completion of this course students will be able to explore, select the appropriate GLM, conduct a comprehensive analysis and interpret results from fitting GLMs to real data.

In particular, students will obtain the following capabilities:

1. Perform Exploratory Data Analysis to inform decision on the appropriate GLM to fit available data.
2. Understand fundamental aspects of parameter estimation and model fitting.
3. Conduct a comprehensive model selection and validation (including model diagnostics) procedure to select the GLM(s) that best fits available data.
4. Understand, interpret and present results from Linear and Generalized Linear Models.

Course Competencies

The purpose of this course is to provide training necessary for a solid foundation in statistical methods for analyzing data using GLM, with an emphasis on applications. The main goal is to help student acquiring the necessary competencies to work collaboratively in an environment that values skill in data analysis using GLM for answering important problems in biomedical research, (biology, public health and clinical medicine), including:

1. Demonstrate a foundation in statistical theory and methods for standard designs and analyses encountered with biomedical data.
2. Identify and implement statistical techniques and models for analysis of data.
3. Acquire knowledge and skills in research methodologies to collaborate with substantive investigators.
4. Recognize key research designs and be able to assist in developing plans for their implementation.
5. Apply programming skills to analyze data and develop simulation studies
6. Attain proficiency in management, and documentation of study data for use in practical statistical analysis.
7. Formulate a public health question in statistical terms.
8. Develop proficiency in making oral, written and poster presentations of work to statistical and non-statistical colleagues.
9. Review and evaluate the use of biostatistical methods in public health or biomedical field of study.

Overall Course Expectations

Students in this course are expected to do the following:

1. Attend all lectures and actively participate in class discussions.
2. Complete and turn in all assignments on time. Solutions to homework must be clearly written with appropriate tables and figures included.
3. Demonstrate an understanding on material on examinations.
4. Respect each other, each others questions and each others discussion.

Evaluation

Students will be evaluated based on:

Grade Category	Percentage
Participation	10%
Homework	60%
Final Exam	30%

Evaluation Category Details

Participation

This course will move in a fast pace; active participation in class discussions is a crucial component for success in the course. Unexcused absences are not allowed and may result in a loss of percentage points or even an incomplete (INC) at the end of the semester.

Homework

Students will need to complete a set of homework assignments that will require them to use statistical software to conduct data analyses using statistical methods studied in the course, and interpret the results. The output of each homework assignment will be a report in which students will summarize the methods used and present results of the statistical analysis they conducted in response to the homework questions. It is expected that all assignments will be completed on time. **No late assignments will be accepted.**

Final Exam

The final exam will be a take home assignment for which students will have to submit a complete report by the due date. The final exam will be comprehensive in nature requiring students to plan and carry out an in depth analysis of public health data using methods they have learned in this class.

Semester Hours

Over the course of a semester it is estimated that students will spend on average the following amount of time (varying by student) on each task:

Task	Hours Spent on Task
Class Time	38
Reading	52
Homework	75
Exams (Take Home)	20
Total	185

Students with Special Needs

Brown University is committed to full inclusion of all students. Students who, by nature of a documented disability, require academic accommodations should contact the professor. Students may also consult with the Student Accessibility Services (SAS) Office of Brown University (SAS@brown.edu).

Diversity Statement

This course is designed to support an inclusive learning environment where diverse perspectives are recognized, respected and seen as a source of strength. It is our intent to provide materials and activities that are respectful of various levels of diversity: mathematical background, disability, gender, age, religion, socioeconomic status, ethnicity, race, and culture.

Student Responsibilities and Academic Integrity

The Brown University's Academic Code

Please carefully review Brown University's Academic Code.

In summary, the Academic Code states:

"*Academic achievement is evaluated on the basis of work that a student produces independently. A student who obtains credit for work, words, or ideas which are not the products of his or her own effort is dishonest. Such dishonesty undermines the integrity of academic standards of the University. Infringement of the Academic Code entails penalties ranging from reprimand to suspension, dismissal or expulsion from the University. Brown students are expected to tell the truth. Misrepresentation of facts, significant omissions or falsifications in any connection with the academic process (including Change of Course permits, the academic transcript, or applications for training or employment) are violations of the Code. This policy also applies to alumni, insofar as it relates to Brown transcripts and other records of work at Brown. Misunderstanding the Code will not be accepted as an excuse for dishonest work. If a student has questions on any aspect of the Academic Code as it relates in a particular course or as it may be interpreted in practice, he or she should consult the instructor in the course or one of the deans of the Graduate School so as to avoid the serious charge of academic dishonesty." (Academic Code, p.4)*

Artificial Intelligence (AI) tools Usage Guidance

The Brown's Academic Code states,

A student's name on any exercise (e.g., a theme, report, notebook, performance, computer program, course paper, quiz, homework assignment report, or examination) is regarded as assurance that the exercise is the result of the student's own thoughts and study, stated in his or her own words, and produced without assistance, except as quotation marks, references, and footnotes acknowledge the use of printed sources or other outside help.

Students are expected to abide by the Academic Code.

One of the main objectives of this course is to help students enhance their own critical thinking in using appropriate Statistical Methods to conduct comprehensive data analyses, and interpret and present results from these analyses.

Please also keep in mind that AI-generated reports can be inaccurate and totally misleading.

Therefore AI-generated submissions are **not permitted**, unless specifically indicated in the assignment Question instructions.

In some assignments, you may be asked to explore the use of AI in the statistical analysis of public health data, with an emphasis on what AI can and cannot do well. Such cases will be clearly indicated in the relevant assignments.

Students are strongly encouraged to talk to the Professor if they need any clarification regarding the use (or not) of AI-tools to complete course assignments.

Tentative Class Schedule

NOTE: Class readings are subject to change, contingent on mitigating circumstances and the progress we make as a class. Students must attend lectures and check the course website for updates.

Week	Contents
1	Syllabus & Course Overview Review - Intro to GLM - The Exponential Family
2	Statistical Inference for Continuous Outcomes – Comparing Means Review of Linear Models
3	Multiple Linear Regression Linear Models - Model Selection
4	Linear Models - Assessing model fit Linear Models - Regression Diagnostics
5	Linear Models - Prediction Intro to Logistic Regression
6	Logistic Regression - link functions / interpretation Logistic Regression - GoF / Regression Diagnostics
7	Models for count Data - Poisson Regression Poisson Regression GoF
8	Contingency Tables / Log linear Models Log Linear models
9	Models for Nominal Data Models for Nominal Data

Tentative Class Schedule (*continue*)

Week	Contents
10	Models for Ordinal Data
	Models for Ordinal Data
11	Intro to Survival Analysis
	Cox Regression
12	Cox Regression
	Cox Regression
13	Parametric Survival Models
	No class (Thanksgiving)
14	Parametric Survival Models
	Other topics
15	Review Session (Reading Period)
	Review Session (Reading Period)
	Final Examination Period: **Dec 12 - Dec 20** (Sunday inclusive)
	Final Exam (take-home): **Release Date**
	Final Exam: **Due Date**