Course Syllabus
BIOSTAT 719 Generalized Linear Models
Fall 2024
3 Units/Credits
Instruction Mode: In person

Course Hours and Location:	Tuesdays and Thursdays, 3:05 PM – 4:20 PM (ET) @Hock 10089
Instructor:	**Hwanhee Hong**, Ph.D. Associate Professor Department of Biostatistics & Bioinformatics
Office:	Hock Plaza, Room 11041
Email:	hwanhee.hong@duke.edu
Office Hours:	Thursdays, 4:20pm - 5:00pm (ET) @Hock 11041
Teaching Assistants and TA office hours:	**Annabel Settle**, second year B&B PhD student annabel.settle@duke.edu Tuesdays, 4:20pm - 5:00pm (ET) @Hock 10092 **Qi Wang**, first year B&B PhD student qi.wang@duke.edu Wednesdays, 9:00am - 9:40am (ET) @Hock 10092

I. Course Overview

Medical research requires ability to relate explanatory variables to an outcome variable. This course will provide tools needed to analyze such relationships for various types of outcome variables within a unified framework of generalized linear models. Observations are not always independent, and the course will discuss methods used to account for correlations within clusters of observations.

Prerequisites: BIOSTAT 701, BIOSTAT 702, BIOSTAT 705, BIOSTAT 706/706A or permission of Director of Graduate Studies.

II. Course Objectives and Student Learning Outcomes (SLOs)

The primary objectives and their SLOs are as follows:

Objective 1: *To have students understand theory and develop practical skills to perform regression analyses relating outcome with explanatory variables for a variety of outcome types.*

- ***SLO 1.1:*** Students should understand the basic theory of exponential family and generalized linear models (GLM) and their use in real data applications with various outcome types.
- ***SLO 1.2:*** Students should be able to perform practical GLM analyses using R and SAS, identify parameters of interest, and carry out model-fitting, estimation, testing, and interpretation.

Objective 2: *To have students understand theory and develop practical skills to perform statistical analysis with correlated observations.*

SLO 2.1: Students should recognize longitudinal/clustered data in which correlations within clusters are expected.
SLO 2.2: Students should understand the basic theory of generalized linear mixed models (GLMM) often used to account for correlations within data clusters and perform practical analyses of clustered data with GLMM using R and SAS.

III. Course Textbooks/Software:

- **Required Textbook:** *An Introduction to Generalized Linear Models*, Fourth Edition (Chapman & Hall/CRC Texts in Statistical Science), Annette J. Dobson and Adrian Barnett. 2018.
- **Recommended/Future Reading:**
  Foundations of Linear and Generalized Linear Models, First Edition (Wiley Series in Probability and Statistics), Alan Agresti. 2015.
- **Software:** R, SAS.

IV. Learning Outcome Assessment and Letter Grading Scale

Assessment Methods	Points
Attendance and pop quiz	5%
Homework	20%
Midterm Exam	25%
Final Exam	30%
Group Research project	20%
Total	**100%**

Letter Grade	Point Proportion
A+	97 – 100
A	95 – 97
A-	90 – 95
B+	85 – 90
B	80 – 85
C	70 – 80
D	50 – 70
F	< 49

V. Course Requirements and Expectations

- Lectures will be conducted *in person* at regularly scheduled class times.
- There will be 9 weekly homework assignments. Homework assignments will be posted in Canvas. Homework is due at the beginning of the class on the due day and submitted through Canvas. If you cannot upload your homework to Canvas on time (e.g., due to medical or family emergency), email your homework to the TAs and instructor.
- The midterm and final exams will be in-class and allow one or two-page cheat sheet only.
- The group research project will require analysis of a data set, final report, and final presentation.

- Pop quizzes will be given without prior notice for the purpose of checking attendance and key topics of lectures. They will be in a simple form, such as yes/no or multiple-choice questions that you can answer in one minute. If you miss a class due to medical or family emergency, please let your TAs and instructor know it then pop quiz will be given via email.

VI. Course Policies

- Microsoft Teams will be used as a communication tool.
- Co-operation between students and learning from each other is encouraged. However, homework assignments are to be prepared *independently* by each student.
- Midterm and final exams *should NOT be discussed with anyone*.
- Late assignments will receive *10% deduction*. If one cannot make a deadline, please let the instructor know in advance. Late assignments should be sent to the TAs and instructor via email.
- Students are *NOT permitted to share* any video from the class outside of the class.
- Students are *NOT permitted to solicit* class materials in previous years. This violates Student Conduct.
- Early or late exams will only be allowed for medical or family emergency (e.g., funerals) and you will need to provide valid proof.

VII. Artificial Intelligence Policies

- Students are allowed to use advanced automated tools (artificial intelligence or machine learning tools such as ChatGPT) on assignments in this course if that use is properly documented and credited.
- Students using AI should be transparent about their use. AI use must be acknowledged if you use it to assist with your homework or writing beyond spell-check or grammar suggestions. state it clearly.
- For example, text generated using ChatGPT-3 should include a citation such as: “Chat-GPT-3. (YYYY, Month DD of query). “Text of your query.” Generated using OpenAI. https://chat.openai.com/” Material generated using other tools should follow a similar citation convention. See other examples here.
- Students using AI should conduct fact-checking and provide proper sources (e.g., academic papers) about AI-generated contents.
- AI-generated contents without proper acknowledgements or references will be treated as plagiarism.

VIII. Academic Integrity and Misconduct Statement

- Academic Integrity Statement
  - https://gradschool.duke.edu/policies-forms/standards-conduct/
  - https://studentaffairs.duke.edu/conduct/about-us/duke-community-standard
- Harassment, Sexual Misconduct, and Discrimination (Duke Office for Institutional Equity)
  - https://oie.duke.edu/
  - https://students.duke.edu/get-assistance/
- ASA Ethical Guidelines for Statistical Practice
  - https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx

IX. Accommodation of Student Needs

Links to additional Information about Duke policies, updates related to COVID-19 and wellness resources:

- https://coronavirus.duke.edu/
- https://students.duke.edu/wellness/

X. Class and Department Emergency Checklist

Duke has developed guidelines for how to respond to dangerous incidents, including while teaching in a classroom or working in a laboratory.

- Class and Department Emergency Checklist
- Duke's Emergency Management Website

X. Course Schedule

(T=Tuesday; Th=Thursday; LN=Lecture Note; H=Homework)

Class No.	Date	Topics	Readings	Assignments
1	Aug 27 T	Introduction	LN0	
2	Aug 29 Th	Review of Linear Models	Chapter 1 LN1	H1
3	Sep 3 T	Exponential Family	Chapters 2 & 3 LN2	
4	Sep 5 Th	Exponential Family properties Generalized Linear Model (GLM)	Chapter 3 LN2, LN3	**H1 due** H2
5	Sep 10 T	Newton-Raphson algorithm and Method of Scoring	Chapter 4 LN3	
6	Sep 12 Th	GLM estimation	Chapter 5 LN3	**H2 due** H3
7	Sep 17 T	GLM estimation/inference	Chapter 5 LN3	
8	Sep 19 Th	Deviance and hypothesis testing	Chapter 5 LN3	**H3 due** H4
9	Sep 24 T	Deviance and hypothesis testing	Chapter 5 LN3	
10	Sep 26 Th	Logistic regression	Chapter 7 LN4	**H4 due** H5, H6*
11	Oct 1 T	Logistic regression	Chapter 7 LN4	
12	Oct 3 Th	Logistic regression further topics	Chapter 7 LN4	**H5 due**
13	Oct 8 T	Logistic regression further topics	Chapter 7 LN4	
14	Oct 10 Th	Poisson regression; Dispersion	Chapter 9 LN6	**H6 due** Practice Midterm, H7*
Fall break Oct 14 - 15				
Oct 17 (Thursday) 3:05PM – 4:20PM		**Midterm Exam (in-class with one doble-sided page of cheat sheet)**		

16	Oct 22 T	Midterm review Group Research project overview		**Group project out**
17	Oct 24 Th	Poisson regression; Dispersion	Chapter 9 LN6	
18	Oct 29 T	Multilevel/Hierarchical/Mixed effects models	Chapter 11 LN7	
19	Oct 31 Th	Multilevel/Hierarchical/Mixed effects models	Chapter 11 LN7	**H7 due** H8
20	Nov 5 T	Multilevel/Hierarchical/Mixed effects models	Chapter 11 LN7	
21	Nov 7 Th	Multilevel/Hierarchical/Mixed effects models	Chapter 11 LN7	H9*
22	Nov 12 T	Nominal logistic regression	Chapter 8 LN5	
23	Nov 14 Th	Ordinal logistic regression	Chapter 8 LN5	**H8 due** **Group project report and slides due Nov 17**
24	Nov 19 T	Research project presentations		
25	Nov 21 Th	Research project presentations		**H9 due**
26	Nov 26 T	**NO CLASS**		
Reading period Nov 27 – Dec 10 Two TA sessions will be offered During the reading period. Time and location TBD.				
Dec 15 (Sunday) **2 PM – 5 PM**		**Final Exam (in-class with two doble-sided pages of cheat sheet)** https://registrar.duke.edu/student-resources/exam-schedules/		

*This homework is long.