EMORY
ROLLINS SCHOOL OF PUBLIC HEALTH

DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: BIOS 709 **SECTION NUMBER:**

CREDIT HOURS: 4 **SEMESTER:** Spring

COURSE TITLE: Generalized Linear Models

INSTRUCTOR NAME: Yijuan Hu

INSTRUCTOR CONTACT INFORMATION

EMAIL: yijuan.hu@emory.edu

PHONE: 404-712-4466

SCHOOL ADDRESS OR MAILBOX LOCATION: 1518 Clifton Rd NE, Atlanta, GA, 30322

OFFICE HOURS: GCR 368, Tue 3 - 4pm or by appointment

Teaching Assistant(s): Zhengyi Zhu

COURSE DESCRIPTION

We will introduce the general framework of generalized linear models (GLMs), the theory of maximum likelihood estimation, and inference for GLMs. We will then survey the use of GLMs for categorical data including logit models for nominal and ordinal data and loglinear models for contingency tables and count data. We will also examine several extensions of GLMs including quasi-likelihood methods for overdispersion, generalized estimation equations (GEE) and generalized linear mixed models (GLMM) for longitudinal or correlated data.

It is a required course for the BIOS PhD program. The students take it in Spring of their 2nd (1st) year if they are on the regular (expedited) track.

The prerequisites include BIOS 508, BIOS 511, and BIOS 707 (or their equivalents).

CONCENTRATION COMPETENCIES:

- $B_{PhD}1$: Conduct independent research in the fields of biostatistics and its application.
- $B_{PhD}3$: Develop and assess new statistical methods to address a broad range of complex biomedical or public health problems.
- $B_{PhD}4$: Conduct complex and statistical analyses for a broad range of applications.
- $B_{PhD}5$: Teach statistical theory or methodology at multiple levels.

EVALUATION

- Class participation: 10%
- Homework (assigned after finishing every topic): 20%
- Project 1 (assigned close to Mid-term): 20%
- Project 2 (assigned close to Final): 20%
- Final exam (in class, open book): 30%

[90, 100]: A

[87, 90): A-

[85, 87): B+

[80, 85): B

[75, 80): B-

[70, 75]: C

< 70: F

For the grade basis of Sat/Unsat, a grade of B or better is considered Sat and a grade of B- or worse is considered Unsat.

COURSE STRUCTURE

Homework and final exam: the students will receive questions and be asked to apply the general GLM theory to specific applications (e.g., different types of data).

Competency addressed: Develop and assess new statistical methods to address a broad range of complex biomedical or public health problems. Teach statistical theory or methodology at multiple levels.

Projects: the students will receive two real datasets and scientific questions. The students will select appropriate methods, conduct comprehensive analyses of data, interpret and summarize results in language that a non-statistician can understand, and submit a report mimicking an article published in a clinical journal.

Competency addressed: Conduct independent research in the fields of biostatistics and its application. Conduct complex and statistical analyses for a broad range of applications

COURSE POLICIES

The students are required to attend the class. The students are encouraged to ask questions at any time (before, during, after) of the class. The class materials will be distributed before the class as well as being posted on Canvas. The class will be taught using the slides, whiteboard, and sample R (SAS) programs. Homework, project, and exam policies will be discussed individually before each assignment.

Required books:
Agresti, A. Categorical Data Analysis , 3rd Edition, John Wiley & Sons (2012)
- The book's website http://www.stat.ufl.edu/aa/cda/cda.html contains a large number of examples using SAS, R and other software packages

Stokes, M.E., Davis, C.S., and Koch, G.G. Categorical Data Analysis Using SAS, 3rd Edition, SAS Institute (2012)

As the instructor of this course I endeavor to provide an inclusive learning environment. However, if you experience barriers to learning in this course, do not hesitate to discuss them with me and the Office for Equity and Inclusion, 404-727-9877.

RSPH POLICIES

Accessibility and Accommodations

Accessibility Services works with students who have disabilities to provide reasonable accommodations. In order to receive consideration for reasonable accommodations, you must contact the Office of Accessibility Services (OAS). It is the responsibility of the student to register with OAS. Please note that accommodations are not retroactive and that disability accommodations are not provided until an accommodation letter has been processed.

Students who registered with OAS and have a letter outlining their academic accommodations are strongly encouraged to coordinate a meeting time with me to discuss a protocol to implement the accommodations as needed throughout the semester. This meeting should occur as early in the semester as possible.

Contact Accessibility Services for more information at (404) 727-9877 or accessibility@emory.edu. Additional information is available at the OAS website at http://equityandinclusion.emory.edu/access/students/index.html

Honor Code

You are bound by Emory University's Student Honor and Conduct Code. RSPH requires that all material submitted by a student fulfilling his or her academic course of study must be the original work of the student. Violations of academic honor include any action by a student indicating dishonesty or a lack of integrity in academic ethics. *Academic dishonesty refers to cheating, plagiarizing, assisting other students without authorization, lying, tampering, or stealing in performing any academic work, and will not be tolerated under any circumstances.*

The RSPH Honor Code states: "Plagiarism is the act of presenting as one's own work the expression, words, or ideas of another person whether published or unpublished (including the work of another student). A writer's work should be regarded as his/her own property."
(http://www.sph.emory.edu/cms/current_students/enrollment_services/honor_code.html)

COURSE OUTLINE

- Review: Linear Algebra, Relevant Distributions, MLE, and Linear Regression
- Topic 1: GLM Model Formulation, Estimation, Inference and Diagnostics
- Topic 2: GLM Examples, Grouped v.s. Ungrouped Data, Case-Control Studies, Overdispersion
- Topic 3: Logit Models for Multinomial Responses
- Topic 4: Loglinear Models for Contingency Tables
- Topic 5: Conditional Logistic Regression
- Topic 6: Quasi-Likelihood Approach
- Topic 7: Analysis of Correlated Data using Generalized Estimating Equations (GEE)
- Topic 8: Analysis of Correlated Data using Linear Mixed Models (LMMs)
- Topic 9: Analysis of Correlated Data using Generalized Linear Mixed Models (GLMMs)
- Topic 10: Diagnostics