DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: BIOS 507 **SECTION NUMBER:** 2

CREDIT HOURS: 4 **SEMESTER:** Spring 2020

COURSE TITLE: Applied Regression Models

CLASS HOURS AND LOCATION: Monday and Wednesday 10:00-11:50am CNR 5001

INSTRUCTOR NAME: Yi-An Ko

INSTRUCTOR CONTACT INFORMATION

EMAIL: yi-an.ko@emory.edu

PHONE: (404) 712-8128

SCHOOL ADDRESS OR MAILBOX LOCATION: RSPH GCR 208

OFFICE HOURS: TBD

Teaching Assistant(s):

TBD

COURSE DESCRIPTION

This is the first regression analysis course in applied statistics designed for MPH students. Both theoretical and applied aspects of linear regression and generalized linear regression modeling will be covered in this course. The emphasis will be on applications. The first part of the course covers the following topics: simple linear regression, multiple linear regression, analysis of variance, confounding and interaction, residual and influence diagnostics, variable transformations, multicollinearity, model selection and validation. The second part of the course includes: generalized linear models, introduction to maximum likelihood estimation, logistic regression, nominal and ordinal logistic regression, Poisson regression. Parameter interpretation and scientific interpretation of results will be emphasized throughout the course. Students are expected to use SAS (or R), when necessary, for homework assignments.

Prerequisites: Coursework in statistics up to and including an introduction to simple linear regression (BIOS 506 or equivalent). Familiarity with basic concepts of probability, statistical inference, and linear algebra (e.g., matrix inversion, some matrix algebra) is needed for successful completion of the course.

COURSE OBJECTIVES

The overall objective of this course is to help the student apply linear and generalized linear regression methods to scientific studies. The student will learn to identify the scientific goals of a study and to develop a statistical strategy appropriate for those goals. The student will learn to plan strategies for linear regression analysis and to implement these strategies. The student will learn to be aware of problems that arise in study design and data collection. The student will learn to interpret the results of regression analysis and effectively communicate the findings to the broad scientific community.

MPH FOUNDATIONAL COMPETENCIES:

- Analyze quantitative data using biostatistics, informatics, computer-based programming and software, as appropriate.
- Interpret results of data analysis for public health research, policy or practice

CONCENTRATION COMPETENCIES:

- Identify statistical issues in contemporary public health problems.
- Use statistical software for data management and exploratory data analysis.
- Apply regression modeling techniques for continuous, binary, categorical, and count data.
- Communicate the results of statistical analyses to a broad audience.

EVALUATION

Homework (30%) Homework assignments will be posted on Canvas and are to be turned in at the start of class on the due date unless otherwise noted. Students are permitted, and encouraged, to discuss homework assignments with others, but the final write-ups must be independent work. Without prior approval, no late homework will be accepted after the solutions have been posted. 20% in grade will be deducted for late homework. The homework assignments will also be graded for clarity and presentation.

Midterms (40%) There will be two midterm exams (open-book/notes, offline).

Final Project (30%) Written report covering a complete data analysis using techniques covered in the course. Data and details will be provided.

MPH Foundational Competency assessed	Representative Assignment
Analyze quantitative data using biostatistics, informatics, computer-based programming and software, as appropriate.	1. Bi-weekly homework assignment 2. Data analysis project

Interpret results of data analysis for public health research, policy or practice.	1. Bi-weekly homework assignment 2. Data analysis project
BIOS Concentration Competencies assessed	**Representative Assignment**
Identify statistical issues in contemporary public health problems.	1. Bi-weekly homework assignment 2. Midterm exams 3. Data analysis project
Use statistical software for data management and exploratory data analysis.	1. Bi-weekly homework assignment 2. Data analysis project
Apply regression modeling techniques for continuous, binary, categorical, and count data.	1. Bi-weekly homework assignment 2. Midterm exams 3. Data analysis project
Communicate the results of statistical analyses to a broad audience.	1. Data analysis project

COURSE STRUCTURE

Course materials (e.g. lecture handouts, homework and solutions, sample SAS or R code) will be posted on Canvas.

Primary textbook: Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005). Applied Linear Statistical Models, 5th edition. WCB McGraw-Hill/Irwin, Boston. Older versions of KNNL are acceptable.

Software: SAS and R will be used for statistical computing and data analysis. Both SAS and R can be accessed via RSPH Desktop. R is freely available and can be downloaded from: https://cran.r-project.org/

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

Tentative Schedule

Week	Topic
1	Simple linear regression (SLR) review; covariance review; matrix overview
2	Matrix approach to SLR; intro to R
3	Multiple linear regression (MLR)
4	Multiple linear regression (MLR)
5	Confounding and interaction
6	Regression diagnostics
7	Weighted least squares; ridge regression; robust regression
8	Review/catch-up; Midterm I
-	SPRING BREAK
9	Model selection and validation
10	Introduction to maximum likelihood estimation
11	Logistic regression
12	Logistic regression; Review/catch-up
13	Midterm II; nominal and ordinal logistic regression
14	Poisson regression and log-linear models
15	Review/wrap-up
16	Final project due