

DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: BIOS 507

SECTION NUMBER: 1

CREDIT HOURS: 4

SEMESTER: Spring 2023

COURSE TITLE: Applied Regression Analysis

CLASS HOURS: Monday and Wednesday 10-11:50AM

CLASS LOCATION: TBD

INSTRUCTOR NAME: Yi-An Ko

INSTRUCTOR CONTACT INFORMATION

EMAIL: yi-an.ko@emory.edu

PHONE: (404) 712-8128

COURSE DESCRIPTION

This is the first regression analysis course in applied statistics designed for BIOS 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, Poisson regression, and nominal and ordinal logistic regression (if time permits). Parameter interpretation and scientific interpretation of results will be emphasized throughout the course. Students are expected to use R (or SAS if preferred), when necessary, for homework assignments.

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.

EVALUATION

Homework (40%)

Homework assignments will be posted on Canvas and are to be turned in by 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.

Quizzes (10%)	There will be in-class quizzes.
Midterms (20%)	There will be one midterm project or in-class midterm exam.
Final Project (30%)	Students will present the analysis plan and turn in the written report covering a complete data analysis using techniques covered in the course. Data and details will be provided.

COURSE STRUCTURE

Course materials (e.g. lecture handouts, homework assignments and solutions, sample 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: R will be used for statistical computing and data analysis. R can be accessed via RSPH Desktop. R is freely available and can be downloaded from: <http://cran.r-project.org>

COURSE OUTLINE

Week	Topic
1	Simple linear regression (SLR) review; covariance review
2	Simple linear regression; linear regression assumptions
3	Matrix approach to SLR
4	Multiple linear regression (MLR)
5	Polynomial regression;
6	Lack of fit test; confounding and interaction
7	Review; Midterm project or exam
8	Regression diagnostics; weighted least squares
9	Spring break
10	Ridge regression; robust regression; model selection and validation
11	Introduction to logistic regression
12	Logistic regression model checking
13	Modern statistical practice; final project topics
14	Logistic regression model selection and prediction
15	Multinomial and ordinal logistic regression
16	Final project presentations