DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: 509 **SECTION NUMBER: 001**

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

COURSE TITLE: Applied Linear Models

INSTRUCTOR NAME: Robert Lyles

INSTRUCTOR CONTACT INFORMATION

EMAIL: rlyles@emory.edu

PHONE: 404-727-1310

SCHOOL ADDRESS OR MAILBOX LOCATION: GCR # 360

OFFICE HOURS Fridays, 2-3pm

Teaching Assistant(s): Samuel Aiyedipe (office hour TBD)

COURSE DESCRIPTION

BIOS 509 is a required core course for the department's MSPH degrees, and this or a prior equivalent course is also required for the degree of PhD. It is taken during the spring semester of the student's first year in the program, with college-level calculus, linear algebra, and the equivalent of BIOS 506/508 as pre-requisites. The course covers statistical methodology for the analysis of continuous outcome data, primarily from cross-sectional studies and designed experiments. We introduce the key matrix-based methods for estimation and inference based on the multiple linear regression model. Subsequently, topics include secondary hypothesis testing and restrictions, regression diagnostics, model selection, confounding and interaction, analysis of variance and covariance, and an introduction to random effects modeling. Students will also be introduced to logistic regression modeling for binary outcome data.

BIOSTATISTICS CONCENTRATION COMPETENCIES:

1. Analyze continuous data using linear regression models and discrete data using generalized linear models.

2. Explain fundamental concepts of probability and inference used in statistical methodology

EVALUATION

Unless otherwise noted, evaluation will be based upon the following:

Homework:	30%
In-class midterm:	25%
Individual project report:	20%
Final:	25%

Attendance and in-class attentiveness are also considerations in the evaluation process.

Possible letter grades are A, A-, B+, B, B-, C, F. Approximate overall course average score cut-points (possibly subject to change) are as follows:

A:	93-100
A-:	90-93
B+:	87-90
B:	83-87
B-:	75-83
C:	65-75
F:	< 65

COURSE STRUCTURE

The course will consist of a series of in-class lectures to which attendance is expected and during which questions are welcome. The tentative schedule of topics and recommended readings for these class sessions is provided below (see Course Calendar/Outline). The major assessments during the course will be as follows:

a) Homework: Approximately 5-6 take-home assignments will be given, usually due within 10-14 days. Students are welcome to discuss aspects of these problems collaboratively, but are expected to complete <u>their own written or typed reports</u> summarizing their work. *

b) Midterm exam: The midterm exam will be in class, covering approximately the first ½ of the lecture topics. Students will be allotted up to 3 hours to complete the exam, and are permitted to use their text book and in-class notes.

c) Individual project report: This will consist of a 5-7 page write-up of a statistical analysis. Students will receive a set of data to analyze with a description of objectives and a suggested report structure.

d) Final exam: The final exam will be in class and will be cumulative, but focused primarily on the second ½ of the lecture topics. Students will be allotted up to 3 hours to complete the exam, and are permitted to use their text book and in-class notes.

BIOS Concentration Competencies assessed	Representative Assignment
Analyze continuous data using linear regression models and discrete data using generalized linear models.	Homework assignments will encourage students to master matrix-driven calculations and results pertaining to the linear model, as well as the analysis of data that takes advantage of available statistical software. The goals of these exercises focus on estimation and inferential problems using simple and multiple linear regression models, mixed and random effects ANOVA, and logistic regression models. An in-class midterm and final exam will be used to evaluate conceptual understanding. Students will work in pairs to analyze a set of data and produce a written report that will be evaluated with regard to the strategy and conduct of the analysis as well as the quality of written communication.

COURSE POLICIES AND RELATED INFORMATION

Time and location: TBD

Prerequisites: Coursework in statistics up to and including BIOS 506 or equivalent. Experience with statistical software packages (especially SAS) is helpful. 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.

For whom intended: The course is intended for second-semester graduate students in Biostatistics.

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.

Also recommended: Kleinbaum, D.G., Kupper, L.L., Muller, K.E., and Nizam, A. (1998). *Applied Regression Analysis and Multivariable Methods*, 3rd, 4th or 5th edition, Duxbury Press.

Homework: Assignments will be provided in class or via email, and are to be turned in at the start of class on the due date unless otherwise noted. Students are permitted to discuss homework assignments with others, but each student is expected to complete exercises conscientiously and turn in his or her own written work.

Exams and class project: For these, all work must be solely that of the individual student (honor code applies).

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 CALENDAR / OUTLINE

1/14 Intro.; linear regression “primer”; review of covariance concepts

1/16 Simple linear regression (SLR) review and algebraic details (KNNL, Ch. 1-2)

1/21 **MLK Day (No class)**

1/23 Matrices and matrix operations; Linear combinations of random variables (KNNL 5.1-5.9)

1/28 Matrix approach to SLR (KNNL 5.9-5.13) / SAS PROC IML Intro.

1/30 Multiple linear regression (MLR): Concepts, least squares (KNNL 6.1-6.3)

2/4 MLR: Residuals, inference, prediction, ANOVA table (KNNL 6.4-6.7)

2/6 MLR: Further topics; multiple and partial F tests; partial correlation (KNNL 7.1-7.4; 12.11)

2/11 Simulation studies to evaluate statistical properties of MLR (class notes)

2/13 Dummy variables; coding schemes; comparing regression lines (KNNL 8.3-8.7)

2/18 Testing secondary hypotheses in MLR (class notes)

2/20 Restrictions in MLR; Piecewise linear models (class notes)

2/25 Polynomial regression; Lack-of-fit test (KNNL 8.1; 3.7)

2/27 Catch-up; Special topic: Intro to maximum likelihood and nonlinear models (class notes)

3/4 **Midterm Exam (covers topics listed up to 2/25)**

3/6 Project “primer” and scientific reports; Intro to regression diagnostics (KNNL 6.8; 10.1-10.6)

3/11-3/15 **SPRING BREAK!**

3/18 Finish regression diagnostics and multicollinearity (KNNL 10.1-10.6; 7.6)

3/20 Weighted least squares (KNNL 11.1)

3/25 All possible regressions, forward/backward model selection (KNNL 9.1-9.5)

3/27 Confounding and interaction in MLR (class notes; KNNL 8.2)

4/1 One-way ANOVA: Hypothesis testing; regression approach (KNNL 15.1-15.3; 16.4-16.8)

4/3 One-way ANOVA: Multiple comparisons (KNNL 17.4-17.7)

4/8 Intro to two-way studies; Two-way ANOVA with equal sample sizes (KNNL 19.1-19.6)

4/10 Two-way ANOVA: Regression approach; interaction (KNNL 19.7-19.9)

4/15 Intro to random effects; Two-way ANOVA with unequal cell counts (KNNL 23.1-23.3)

4/17 Random effects in ANOVA, cont'd. (KNNL 25.1-25.4)

4/22 Logistic regression "primer"; introduction and basic concepts (class notes)

4/24 Logistic regression example(s) (class notes)

4/29 Catch-up; notes provided as intro to the general linear mixed model (**PROJECT PAPERS DUE)**

FINAL EXAM DATE AND TIME is TBA