EMORY
ROLLINS SCHOOL OF PUBLIC HEALTH

DEPARTMENT: Biostatistics and Bioinformatics

COURSE NUMBER: 526 **SECTION NUMBER:** 1

CREDIT HOURS: 3 **SEMESTER:** Fall 2020

COURSE TITLE: Modern Regression Analysis

CLASS HOURS AND LOCATION:
MW 1:30 – 2:50 PM in GCR 721/729

INSTRUCTOR NAME: Benjamin (Ben) Risk

INSTRUCTOR CONTACT INFORMATION

EMAIL: benjamin.risk@emory.edu

PHONE: 404-712-5081

SCHOOL ADDRESS OR MAILBOX LOCATION: Grace Crum Rollins 332

OFFICE HOURS Thursdays 3-4, GCR 332

COURSE DESCRIPTION

This course introduces students to modern regression techniques commonly used in analyzing public health data. Specific topics include: (1) parametric and non-parametric methods for modeling non-linear relationships (e.g., splines and generalized additive models); (2) methods for modeling longitudinal and multi-level data that account for within group correlation (e.g., mixed-effect models, generalized estimating equations); (3) Bayesian methods; and (4) shrinkage methods and bias-variance tradeoffs.

This course draws motivating examples from environmental and social epidemiology, health services research, clinical studies, and behavioral sciences. The course provides a survey of advanced regression approaches with a focus on data analysis and interpretation. Students will gain an understanding of methods that will facilitate future independent and collaborative research for modern research problems. Students will gain practical experience using the R language for statistical computing.

Target audience: This is a required course for MSPH and MPH students in BIOS in the fall of their second year. It is also an elective course for PhD BIOS students and other graduate students interested in advanced modeling techniques.

Prerequisites: Prerequisites are a course in linear algebra, BIOS 506, and BIOS 507.

MPH/MSPH FOUNDATIONAL COMPETENCIES:

- Select quantitative and qualitative data collection methods appropriate for a given public health context
- Analyze quantitative and qualitative 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:

- Communicate the results of statistical analyses to a broad audience
- Apply analytic methods to address specific research questions in the particular application area of interest
- Use standard statistical software for both data management and data analysis

COURSE LEARNING OBJECTIVES

See course description and competencies.

EVALUATION

Homework (40%) Seven problem sets will be assigned. The problem sets provide experience managing and analyzing data in R. R is required. Students are encouraged to use R markdown to create their write-ups.

Students should submit homework questions to the Canvas discussion board, which will be answered by the TAs, instructor, and other students. Students are allowed to discuss homework problems with each other. However, the final write-ups must be independent work. The homework assignments will also be graded for clarity.

Homeworks are to be submitted electronically in pdf form through Canvas and are due prior to the start of class on the due date. Without prior approval, no late homework will be accepted. 20% of the grade will be deducted for late homework. No homeworks will be accepted after solutions are posted (usually 3 days after the due

date). Questions about graded homework should be addressed to the TAs within 1 week of receiving the graded assignment.

Participation (5%) Students are expected to attend all lectures. Students are encouraged to ask questions and participate in discussions. The use of cellphones and inappropriate use of computers will lead to a lower participation score. It is expected that most students will receive full participation points.

Group Presentation (10%)

Students will be divided into groups of approximately five and asked to present an academic article employing modern regression methods. Presentations will be 15 minutes and include an introduction to the motivating problem accessible to a general audience, data description, the statistical approach, results, and discussion. Students will also give the other members in their group a participation score from 1 (no participation) to 5 (full participation). The group score will then be adjusted for students with less than full participation at the instructor's discretion.

The groups will be created by the TAs and instructor.

Each group must email their selected article to the instructor for approval before 10/31. Articles may be selected from either the scientific literature or statistical literature. Possible journals include Statistics in Medicine, Annals of Applied Statistics, Biometrics, Epidemiology, and NeuroImage. Articles in PNAS, Nature, and Science will typically require an examination of the Supplemental Materials, which tend to include details of the statistical analysis.

The presentations will be on 12/3 and 12/5, and the group should also submit their slides on Canvas.

Midterm (20%) A take-home independent data analysis project will be assigned. Evaluation is based on a written report which will consist of sections on motivation/goals, data description, analytic approach, analysis results, and discussion. The report will include a lay abstract that communicates your findings to a broad audience without statistical training.

Final Exam (25%) A take-home independent data analysis project will be assigned in which students will analyze the patterns and factors affecting the prevalence of a disease. Evaluation is based on a written report which will consist of sections on motivation/goals, data description, analytic approach, results, and discussion. The report will include a

discussion of the public healthy and policy implications of the findings.

Grading: A (93-100), A- (90-92), B+ (87-89), B (83-86), B- (80-82), C (60-79), F (<59).

COURSE STRUCTURE

The course is organized into learning modules, where each module is devoted to an analytic method.

MPH/MSPH Foundational Competencies assessed	Representative Assignment
Select quantitative and qualitative data collection methods appropriate for a given public health context	1. Final Exam
Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate	1. Midterm Exam
Interpret results of data analysis for public health research, policy, or practice	1. Final Exam

MPH/MSPH Concentration Competencies assessed	Representative Assignment
Communicate the results of statistical analyses to a broad audience	1. Midterm Exam 2. Group Presentations
Apply analytic methods to address specific research questions in the particular application area of interest	1. Final Exam
Use standard statistical software for both data management and data analysis	1. Homework 1 2. All exams and homework address this competency

COURSE POLICIES

Attendance:

Students are expected to attend all lectures.

Course participation:

Students are encouraged to ask the instructor questions during the lectures. Students are encouraged to participate in discussions as prompted during lecture.

Assignment submission and policies:

As noted above: For homeworks, students should submit questions to the Canvas discussion board, which will be answered by the TAs, instructor, and other students. Students are allowed to discuss homework problems with each other. However, the final write-ups must be independent work. The homework assignments will also be graded for clarity. Homeworks are to be submitted electronically in pdf form through Canvas and are due prior to the start of class on the due date. Without prior approval, no late homework will be accepted. 20% of the grade will be deducted for late homework. No homeworks will be accepted after solutions are posted (usually 3 days after the due date). Questions about graded homework should be addressed to the TAs within 1 week of receiving the graded assignment.

The write-ups *must reflect the student's own work*. No two students should have the same code or same sentences. Students violating this policy will receive a 0 on the homework and be faced with possible disciplinary action by the Student Honor and Conduct Code Council. See honor code below.

For exams, students are *not* allowed to discuss the questions with each other. Questions can only be asked via the Canvas discussion board where all other students can view responses, and the questions will be answered by the TAs and course instructor. Unlike homeworks, students are not allowed to answer other students' questions on the discussion board. The write-ups *must reflect the student's own work*. No two students should have the same code or same sentences. Students violating this policy will receive a 0 on the exam and be faced with possible disciplinary action by the Student Honor and Conduct Code Council. See honor code below. Students must submit any regrade requests within two weeks of the exam due date. If a student asks for a regrade, the entire exam will be regraded, which could result in a lower grade.

Assignment submission:

PDF copies of assignments and exams are to be submitted electronically via Canvas.

Technology in the classroom:

Students are not allowed to use cell phones during lectures. Students are not allowed to use computers for anything not directly related to the lecture. Inappropriate use of technology is disrespectful.

Textbook:

There is no required textbook. Recommended textbooks include:

Ruppert, D., Wand, M.P. & Carroll, R.J. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning, 2nd Edition*. New York, New York: Springer. https://web.stanford.edu/~hastie/Papers/ESLII.pdf

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York, New York: Springer. http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambrige: Cambridge University Press. Google for electronic copy.

A new textbook is scheduled for release in late 2018 / early 2019:

Harezlak, J., D. Ruppert, and M.P. Wand. (To Appear). *Semiparametric regression with R.* New York, New York: Springer.

RSPH POLICIES

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.

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

The following is a tentative list of topics and timeline. The schedule will be periodically updated and available on Canvas.

Date	Topic	Due
08/29/18	Linear regression review (Module 1 notes)	
09/03/18	**No class (Labor Day)**	
09/04/18	Optional R review: 4-6 pm GCR 721/729 (M0)	
09/05/18	Basis functions, splines (M2)	
09/10/18	Parametric splines, model selection (M2)	
09/12/18	Penalized splines (M3)	HW 1
09/17/18	GAMs, bivariate splines (M4)	
09/19/18	Normal random intercept model (M5)	
09/24/18	Normal random intercept model (M5)	
09/26/18	Normal random slope model (M6)	HW 2
10/01/18	Generalized linear model review (M7)	
10/03/18	**No Class (Fall break)**	
10/08/18	Generalized linear mixed model (M8)	HW 3
10/10/18	Introduction to Bayesian inference (M9)	
10/15/18	Introduction to Bayesian inference (M9)	Midterm Due
10/17/18	Bayesian linear regression (M10)	
10/22/18	Bayesian hierarchical model (M11)	HW 4
10/24/18	Higher-level GLMM (M12)	

10/29/18	Higher-level GLMM (M12)	
10/31/18	Generalized estimating equation (M13)	Choose paper
11/05/18	Generalized estimating equation (M13)	HW 5
11/07/18	SVD, PCA, and PCA regression (M14)	
11/12/18	Sparse regression and shrinkage (M15)	
11/14/18	Sparse regression and shrinkage (M15)	HW 6
11/19/18	Local regression and functional PCA (M16)	
11/21/18	**No class (Thanksgiving break)**	
11/26/18	Power analysis (M17)	HW 7
11/28/18	**No class (Ben out of town)**	
12/03/18	Group presentations I	
12/05/18	Group presentations II	
12/10/18	Additional Topics	Final Due

COURSE OUTLINE

See above. Note that the readings and objectives for each module are listed in the second slide of each module.