Regression Analysis - ISYE 6414
Instructor: Dr. Nicoleta Serban

Course description: An introduction to commonly used linear regression models along with implementation of the models within data examples using the R statistical software.

Course prerequisites: A sound familiarity with undergraduate or graduate statistics and probability but also basic programming proficiency, linear algebra and basic calculus.

Honor Code All course participants (myself, teaching assistants, and learners) are expected and required to abide by the letter and the spirit of the edX honor code (https://www.edx.org/edx-terms-service) and the Georgia Tech honor code (www.honor.gatech.edu). Keep in mind:
• Ethical behavior is extremely important in all facets of life.
• You are responsible for completing your own work.
• Any learner found in violation of the edX Honor Code will be subject to any/all of the actions listed in the edX Honor Code.

Communication Please contact your instructor, teaching assistants, and fellow learners via the edX course web site discussion forums.
• In the online course, there is a discussion forum called General Course Questions. You can ask questions about the course or content in this forum.
• Participation in the discussion forum is recommended but it is not mandatory. Often, discussions with fellow learners are the sources of key pieces of learning.
• The discussion will be supervised by teaching assistants under instructor's guidance.

Netiquette *Netiquette* refers to etiquette that is used when communicating on the Internet. Review the Core Rules of Netiquette.
• When you are communicating via email, discussion forums or synchronously (real-time), please use correct spelling, punctuation and grammar consistent with the academic environment and scholarship.
• *In Georgia Techs MS in Analytics program, we expect all participants (learners, faculty, teaching assistants, staff) to interact respectfully. Learners who do not adhere to this guideline may be removed from the course.*

Recommended Textbooks: Two primary textbooks are highly recommended:

    1. Kleinbaum, D. G., Kupper, L.L., Nizam, A., Muller, K.E. (2013) *Applied Regression Analysis and Other Multivariable Methods*, Cengage Learning; 5 edition.
    2. P. McCullagh, J.A. Nelder (1989), *Generalized Linear Models*, Chapman & Hall.

What students will learn in this course? By the end of the this class, students will learn the basics of regression analysis such as linear regression, generalized linear regression and model selection. Students will be given fundamental grounding in the use of some widely used tools, but much of the energy of the course is focus on individual investigation and learning. Active participation in the class is very important. This class is more about the opportunity for individual discovery than it is about mastering a fixed set of techniques.

What activities will the course involve students in to help them practice and demonstrate their learning?

Homework Assignments: There will be five assignments including both concept and data example problems. Conclusions and interpretation of results are very important in statistical modeling. These assignments are intended to help you prepare for the midterm exams and final exam. You are also allowed (and encouraged) to ask questions, although you should try to think about the problems before asking.

Midterm: There will be one midterm exam with problems reviewing the material (lectures and assignments) provided in this course throughout the the first three units. The exam is close notes (including homework) and books for the first part and open notes for the R data analysis portion of the exam. The midterm is designed to help students grasp standard regression analysis methodology which will further facilitate a deeper understanding in the application context.

Final Data Analysis Exam: The general goal of the final exam is to provide you with experience in applying regression analysis methodology to real data. This final exam is a guided data analysis using the tools you have learnt throughout the course and it will serve as a means for students to demonstrate what they understand and can do with the content of the course. In grading, we will primarily look for a sensible approach to the problem, and clearly-made connections between your analyses and the substantive questions. You can use any computing equipment and any written source material.

How will students be evaluated? The course will be letter graded. The grade for the course will be based on one midterm, one final exam, and assignments during the semester - Midterm: 30%, Final: 40%, Assignments: 30%.

Course Topics and Schedule: The table below contains a course topic outline and study pace.

Weeks	Unit	Lectures	Assessments	Estimated time (hours)
1-3	Units 1& 2	Lectures 1.1-1.3, 2.1-2.2	Homework 1	30 hours
4-5	Unit 3	Lectures 3.1-3.2	Homework 2	20 hours
6-7	Unit 3	Lectures 3.3-3.4	Homework 3	20 hours
8			Midterm	10 hours
9-11	Unit 4	Lectures 4.1-4.5	Homework 4	30 hours
12-14	Unit 5 & Unit 6	Lectures 5.1-5.3	Homework 5	30 hours
15			Final Exam	10 hours

Technology/Software Requirements:
- Internet connection (DSL, LAN, or cable connection desirable)
- R statistical software (free download; see cran.r-project.org)
- Adobe Acrobat PDF reader (free download; see https://get.adobe.com/reader/)