

Statistical Modeling and Regression Analysis - ISYE 6414

Instructor: Dr. Nicoleta Serban

Course description: An introduction to commonly used linear regression models along with implementation of the models with 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 Codes and Student Conduct All course participants (instructor, teaching assistants, staff 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 policies of the Georgia Tech Academic Honor Code, and the Student Conduct expectations (<http://www.policylibrary.gatech.edu/student-life>). Keep in mind:

- Ethical behavior and personal integrity are extremely important in all facets of life.
- Learners are responsible for completing their own original work.
- Any course participant found in violation of the edX Honor Code, the Georgia Tech Academic Honor Code and/or the Georgia Tech Student Conduct expectations will be subject to any/all consequences listed.

Communication Course updates will be sent through the piazza platform on Canvas. Please contact your instructor, teaching assistants, and fellow learners via piazza.

- The course will host a class discussion forum using piazza. Feel free to ask questions and respond to other students' questions to the best of your knowledge. 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.
- We will send announcements about course changes through piazza. Check the piazza forum at least twice weekly for course schedule and/or assignment updates. You are responsible for staying abreast of these changes. Avoid missing deadlines and assignment rework by keeping updated through piazza.
- Communicate with instructors, teaching assistants and fellow learners using your name as listed in the student roster. E-communication can be less constructive or less thoughtful than in-person communication. When someone does not introduce himself or herself, it is easier to be less respectful. To avoid sensitive situations, we are asking everyone to post in piazza with their name. Posts made by Anonymous profiles will not be responded to.
- Please search piazza for your question prior to posting a new one as it may already be answered.
- Instructors and teaching assistants may not be able to address all piazza communications, so we encourage fellow learners to respond to posts. If there is a delay in instructor or teaching assistant response, please be patient and know there may not be a response if we are in a week with heavy volume and/or if the question has already been addressed in a different post.
- Overall, the discussions will be supervised and monitored 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 Tech's MS in Analytics program, we expect all participants (learners, faculty, teaching*

assistants, staff) to interact respectfully. Learners who do not adhere to these expectations 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 focused 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 allowed (and encouraged) to ask questions about the assignments, although you should try to think through potential solutions before asking.

Midterm Exams and Final Exam: There will be two midterm exams and a final exam with problems reviewing the material (lectures and assignments) provided in this course. The exams are designed to help students grasp standard regression analysis methodology which will further facilitate a deeper understanding in the application context. The exams are divided into two parts: a closed-book part consisting of True/False and Multiple Choice questions and an open-book part including an R data analysis. Please see Open-Book Policy below.

Open-Book Policy. Certain parts of the exams are open-book. What we mean by open-book is that student's are allowed to refer to course class material, including lectures and homework assignments, any material provided in the course, and the R help menu. Open-book includes any notes you may have stored on your computer on the course topics. Open-book in this course does not include access to the internet, or communication by any method. Use of the internet and/or communication with anyone during the exam will be subject to the Georgia Tech honor code and conduct policies/actions (<http://www.policylibrary.gatech.edu/student-life>).

Grading Policy. In grading the R data analysis problems in the exams, we will primarily look for a sensible approach to the problem, and clearly-made connections between your analyses and the substantive questions.

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 1: 25%, Midterm 2: 25%, Final: 35%, Assignments: 15%.

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

Weeks	Module	Topics	Assessments
1-3	1 & 2	1.1-1.4 & 2.1-2.2	Homework 1
4-5	3 (Part I)	3.1-3.2	Homework 2
6	1-3	Review	Midterm Exam I
7-8	3 (Part II)	3.3-3.5	Homework 3
9-11	4	4.1-4.5	Homework 4
12	3-4	Review	Midterm Exam II
13-14	5	5.1-5.4	Homework 5
15	1-5	Review	Final Exam

Course 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/>)