Regression Analysis - ISYE 6414

Instructor: Dr. Nicoleta Serban
Office: 438 Groseclose, ISyE
E-mail: nserban@isye.gatech.edu
Office Hours: Tuesday 4-5pm

Instructor Assistant:
E-mail: 
Office: , Main building, ISyE
Office Hours: Wednesday 4-5pm

Class Schedule
12:05pm - 1:25pm Tuesday and Thursday (IC 105)

Class Web Address https://t-square.gatech.edu
Most of our class material will be submitted via class email including
• Course syllabus
• Lecture and computer notes
• Homework assignments and solutions
• Your course grades
• Questions and replies for homework assignments
• Important announcements

Honor Code For any questions involving Academic Honor Code issues, please consult me, my
teaching assistants, or www.honor.gatech.edu.

Course prerequisites: A sound familiarity with undergraduate or graduate statistics and prob-
ability.

Textbook: The course material will be based on a set of lecture notes being prepared by the
instructor, but two primary textbooks are highly recommended:

and Statistics.

Other recommended books:

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, model selection and logistic regression, but we will also cover more advanced topics
including generalized linear regression and nonparametric regression. Students will be given funda-
mental 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 and team discoveries than it is about mastering
a fixed set of techniques.

What activities will the course involve students in to help them practice and demon-
strate their learning?
Midterms: There will be two midterm exams with problems reviewing the material (lectures and assignments) provided in this course throughout the full semester. The exams are close notes (including homeworks) and books but a two (one-sided) pages with formulas will be allowed. The midterms are designed to help students grasp standard regression analysis methodology which will further facilitate a deeper understanding in the application context. All students will take the midterm in the same day at the same time.

Dates:
Midterm 1: October 11th
Midterm 2: November 16th

Assignments: Assignments will include both theoretical and computer problems; the latter problems will ask you carry out analysis of data sets and simulations using computer software. Keep in mind that you should not hand in raw computer output. Conclusions and interpretation of results are more important than good printouts. These assignments are intended to help you prepare for the midterm exams and final project. You are allowed (and encouraged) to work together with other students on homework, as long as you write up and turn in your own solutions. You are also allowed (and encouraged) to ask me questions, although you should try to think about the problems before asking. Late Homework will not be accepted.

Project: This project is a requirement you must fulfill in order to pass this course. The general goal of the project is to provide you with experience in applying regression analysis methodology to real data. For this project, you and your team must find a data set on your own. The data cannot be a data set found in a textbook or been analyzed in detail and results published. This project will serve as a means for students to demonstrate what they understand and can do with the content of the course. There will be an oral presentation of the project (∼ 15 minutes). In grading, I 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 computing resources in the school, any written source material you can find, in or out of the school. However, replicating results which have been already published without referencing to the source of publication is subject to plagiarism. Plagiarizing is defined by Websters as “to steal and pass off (the ideas or words of another) as one’s own : use (another’s production) without crediting the source.” Be sure to document carefully your project work.

Deadline to submit an abstract of the project: October 23rd, 2012.
Deadline to submit the project work (report): December 7th, 2012.
Class presentation dates: December 5th and December 7th, 2012.

How will students be evaluated?

The course will be letter graded. The grade for the course will be based on two midterms, a final project, and assignments during the semester - Midterm 1 25%, Midterm 2 25%, Project, presentation and class participation: 35%, Assignments: 15%.

The final project grade is 40% the presentation grade and 60% the report grade. The presentations and the reports will be graded by students in this class. Each project team will assign grades for all the presentations including its own presentation. The average over all the team grades will be the presentation grade. Each team will be in charge of reading and grading two or more reports. The report grade will be a weighted average between students’ grade and the instructor grade. Please refer to grading guidelines documents.
**Objectives:** A tentative list of specific topics in this course is as follows:

**Part 1: Simple Linear Regression**
1. Introduction
2. Estimation
3. Inference

**Part 2: Multiple Linear Regression**
1. Review of Linear Algebra
2. Estimation
3. Inference
4. Diagnostics
5. Misc Topics

**Part 3: Variance-Bias Decomposition and Variable Selection**
1. Bias-Variance Decomposition
2. Model Selection Criteria
3. Regularized Model Selection
4. Variable Selection vs. Hypothesis Testing

**Part 4: Logistic Regression**
1. Model Estimation
2. Logistic Regression With Replication

**Part 5: Generalized Linear Regression**
1. Models
2. Overdispersion in GLM models
3. Model Validation

**Part 6: Nonparametric Regression**
1. Smoothing
2. Kernel Regression
3. Local Polynomials
4. Penalized Regression, Regularization and Splines
5. Smoothing Using Orthogonal Functions