Nonparametric Data Analysis - ISYE 6404

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

Instructor Assistant:  
Carlos Valencia  
Office: 414 Main building, ISyE  
E-mail: carlos.valencia@gatech.edu  
Office Hours: Tuesday 4-5pm

Class Schedule  
1:35pm - 2:55pm Monday and Wednesday (Boggs B6)

Class Web Address  
https://t-square.gatech.edu

Most of our class material will be submitted via t-square 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 probability.

Textbook:  
The course material will be based on the following textbook:  

Other recommended books:  

What students will learn in this course?  
By the end of the this class, students will learn the basics of nonparametric data analysis such as rank based methods and categorical data analysis, but we will also cover more advanced topics including nonparametric regression. 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 and team discoveries than it is about mastering a fixed set of techniques.

Midterms 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 throughout the full semester. The final will be divided into two parts, one part on the paper and the second on the computer. 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 nonparametric statistics 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: September 26th in class
- Midterm 2: November 14th in class
- Final: December 10th 2:50pm - 5:40pm

Assignments: Assignments will include mainly statistical derivation exercises but some assignment will also include 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. 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.

How will students be evaluated?

The course will be letter graded. The grade for the course will be based on two midterms, a final exam, and assignments during the semester - Midterm 1 20%, Midterm 2 20%, Final exam: 40%, Assignments: 20%.

Objectives: A tentative list of specific topics in this course is as follows:

Part 1: Review
1. Probability and Distribution Theory
2. Common Distributions
3. Large Sample Theory

Part 2: Rank Based Methods
1. Order statistics
2. Goodness of Fit
3. Rank Test
4. Design of Experiments

Part 3: Inference
1. The CDF
2. Bootstrap and Jacknife

Part 4: Smoothing
1. The BiasVariance Tradeoff
2. Density Estimation
Part 5: Nonparametric Regression

1. Kernel Regression
2. Local Polynomials
3. Penalized Regression, Regularization and Splines
4. Smoothing Using Orthogonal Functions
5. Wavelet-based Smoothing