Course: 26 BE 7024 & 26 PH 7024
Semester: Fall, 2013
Course No.: 910276
Title: Computational Statistics
Credits: 3
Instructor: MB Rao
Office: Kettering Lab
E-mail: marepalli.rao@uc.edu
Telephone: 513 558 3602
Hours: Tuesday 10:00 – 11:20
Thursday 10:00 – 11:20
Venue: 221 Kettering Lab
Office Hours: Tuesday 11:30 – 12:30
By appointment

Course Description: If you want to analyze data, you need a computing software to crunch the numbers in the data. There are scores of statistical computing software available, almost all of them costing money. In this class we focus on software: SAS and R. The software R is free. We devote the first three weeks to SAS. The rest of the time is given to R. In SAS, we describe the use of some basic PROC procedures. We use R for simulations and advanced statistical methods.

Text Book: None
My notes are self-sufficient.

Reference Books:

**Prerequisite:** BE 7022/PH 7022: Introduction to Biostatistics or its equivalent

**Course Objectives:** Exemplify the role of software, especially SAS and R, in analyzing data, big and small. Train the students in using software for all data analysis needs. Relate the classroom experience to surrounding life and work.

**Purpose:** In physical, biological, and medical sciences, vast amounts of data are generated in response to scientific investigations. To pursue a successful career as a data analyst, one needs to be adept in using good software. The basic goals are to introduce SAS and R, two most popular software in commercial and research worlds, and train the students to acquire reasonable expertise is using these. A number of modern statistical methods will be used as fodder to gain a good degree of mastery of the software.

**OBJECTIVES:**
At the conclusion of the course, the student will be able to:
1. use software for data analyses;
2. learn how to simulate univariate and multivariate distributions;
3. indulge in Monte Carlo Simulations to tame seemingly intractable statistical problems;
4. delve in the world of bagging and boosting to reduce misclassification errors in pattern recognition problems;
5. pursue some text mining algorithms;
6. tackle social networks problems.

**INSTRUCTIONAL METHODS:**
1. Lectures.
2. Presentation of statistical stories relevant to biological and agricultural sciences.
3. Working on problems in the class.

**GRADING:**
**EVALUATIONS AND EXAMINATIONS**
1. Eleven Homework assignments will be given. Homework will be distributed on Thursdays.
Homework is due the following Thursday. 30 points

2. Mid-term Exam: October 24, 2013 30 points
3. Final Exam: December 05, 2013 30 points
4. Project – Presentation in the last two weeks of the semester 10 points

TOTAL POINTS 100 points

GRADES:
90 points and above =A
80 – 89 points =B
70 – 79 points =C
60 – 69 points =D
Below 60 points =F

GROUND RULES FOR EXAMS AND QUIZZES:
1. Exams are closed book, closed notes.
2. You may bring a formula sheet to the exam.
3. Calculators are allowed. However, you are not permitted to share calculators.

HOMEWORK GRADING POLICY: All homework is due on the date stipulated on the homework sheet. Submission a day late results in a loss of 20% of the points allocated for the homework. Submission two days late results in a loss of 40% of the points. After that the homework will not be accepted. These rules are designed to protect the homework grader. However, you can drop one homework (supposedly the one with the lowest score) for the final grade.

NOTES AND HOMEWORK: They will be posted on the blackboard.

LEARNING DISABLED STUDENTS: Any student with disabilities or other special needs, who need special accommodations in this course, are invited to share these concerns with the instructor as soon as possible.

APPROVED ACADEMIC HONESTY STATEMENT: All work in this course must be completed in a manner consistent with the University of Cincinnati Policy. See Page 28 of the Department of Environmental Health Graduate Student Guidelines Handbook.

Important note: The students are expected to have SAS installed in their laptops. They can bring their laptops to the class. We will download R in one of the classes.
TENTATIVE COURSE OUTLINE

1. Entering data into SAS; Importing data from external sources 3 hours
2. Preparing data in SAS for analysis 3 hours
3. PROC MEANS; PROC UNIVARIATE; PROC FREQ 3 hours
4. PROC CORR; PROC REG; PROC ANOVA 3 hours
5. Introduction to R 1 hour
6. Simulating a distribution 3 hours
7. Monte Carlo Optimization 2 hours
8. Gibbs Sampler in Bayesian Analysis 3 hours
9. Monte Carlo Markov Chain Algorithms 2 hours
10. Boot Strapping 2 hours
11. Ensemble methods – Bagging and Boosting 2 hours
12. Classification trees and Random forests 2 hours
13. Text Mining 2 hours
14. Social Network Analysis 3 hours
15. Reduced Rank Regression 3 hours
16. Support Vector Machines 3 hours
17. Neural Networks 3 hours
18. Meta-Analysis 3 hours

*The schedule of lectures is only a rough guide. Every effort will be made to maintain this schedule.