Course: BE 7024 & PH 7024  
Semester: Fall, 2016  
Title: Computational Statistics  
Credits: 3  
Instructor: Roman Jandarov  
Office: 162 Kettering Lab  
E-mail: jandarrn@ucmail.uc.edu  
Telephone: 513-558-7975  
Hours: Tuesday 10:00 – 11:20  
Thursday 10:00 – 11:20  
Venue: 221 Kettering Lab  
Office Hours: Tuesday 11:30 – 12:30

Course Description: If you want to analyze data, you need 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 the R software. The R software is free. We devote the first three weeks to R learning their usage at an advanced level. We will then use R for simulations, advanced statistical methods, and data mining techniques.

Text Book: None  
Our notes and class materials are self-sufficient.

Reference Books:  

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

Course Objectives: Exemplify the role of software 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.

At the conclusion of the course, the student will be able to manipulate data; learn modern data visualization techniques; be able to use approaches to generating random variables; design and
perform simple Monte Carlo experiments; and learn to use methods such as the Bootstrap and cross-validation. They will be introduced to technologies that are useful for statistical computing. By creating customized graphical and numerical summaries students will also be able to discuss the results obtained from their analyses. The tentative outline and topics of the course is as follows:

Introduction to R
Data manipulation in R
Visualization of data
Generating random variables
Monte Carlo simulation
The Bootstrap
Permutation methods
Cross-validation

Additional topics:

Dynamic and reproducible reports with R Markdown, smoothing and density estimation, classification trees and random forests, text mining, cluster analysis, social networks analysis

**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 goal is to introduce R, one of the most popular software in commercial and research worlds, and train the students to acquire reasonable expertise in using it. A number of modern statistical methods will be used as fodder to gain a good degree of mastery of the software.

**Instructional Methods:**
1. Lectures.
2. Online resources and tutorials.
3. Working on problems in the class.

**Grading:**
**Evaluations And Examinations**
1. Homework assignments will be given.
   Homework will be distributed on Thursdays.
   Homework is due the following Thursday.
   50 points
2. Mid-term Exam: TBA
   20 points
3. Final Exam: TBA
   20 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

**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.

**Exams:** Both exams are open book/internet access, but absolutely no communicating with each other.

**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.

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