University of Cincinnati
College of Medicine
Department of Environmental Health
Division of Biostatistics and Epidemiology
Division of Public Health Sciences

Course:	BE 7024 & PH 7024	
Semester:	Spring, 2019	
Title:	Computational Statistics	
Credits:	3	
Instructor:	Roman Jandarov	
Office:	146 Kettering Lab	
E-mail:	jandarrn@ucmail.uc.edu	
Telephone:	513-558-7975	
Hours:	Tuesday	11:30am – 12:50pm
	Thursday	11:30am – 12:50pm
Venue:	221 Kettering Lab	
Office Hours:	Tuesday	1:00pm – 2:00pm

Course Description: We devote the first two weeks to learning R and its use at an advanced level. We will then use R for data manipulation, visualization, simulation, advanced statistical methods, and data mining techniques.

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

Reference Books:

1. Eric A Suess and Bruce E Trumbo – Introduction to Probability Simulation and Gibbs Sampling with R, Springer, New York, 2010.
2. Christian P Robert and George Casella – Introducing Monte Carlo Methods with R, Springer, New York, 2010.
3. Giovanni Seni and John Elder – Ensemble Methods in Data Mining, Morgan and Claypool Publishers, 2010.
4. Yanchang Zhao – R and Data Mining, Elsevier, New York, 2013.
5. Hadley Wickham – Advanced R, CRC Press, New York, 2015.

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

Course Objectives: 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 numbers; design and perform simple Monte Carlo experiments; and learn to use methods such as the Bootstrap and cross-validation. They will be introduced to tools 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 numbers
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 demand. 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. **40 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 **20 points**

   TOTAL POINTS 100 points

Grades:

94 points and above	=A
83 – 93 points	=B
70 – 82 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 home work (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.**