Course: 26 BE 7023 & 26 PH 7023  
Semester: Fall, 2014  
Course No.: 209719  
Title: Advanced Biostatistics  
Credits: 3  
Instructor: MB Rao  
Office: 247 Kettering Lab  
E-mail: marelalli.rao@uc.edu  
Telephone: 513 558 3602  
Hours: Wednesday 4:00 – 6:30  
Venue: 221 Kettering Lab  
Office Hours: Wednesday 3:00 – 4:00  
By appointment  
Grader: Lixia Zhang

**Course Description:** All articles published in the New England Journal Medicine during 2004-05 were surveyed for statistical content. Simple linear regression was used in 6%, Multiple regression in 51%, Repeated measures in 12%, and Survival methods in 61% of the articles. Multiple regression and Cox regression are the most dominant methodologies used. This class will pay homage to regressions in all their glory. The list presented below is a proof to what extent we are committed to regression in this class. This is purely an applied class with examples culled from a variety of sources. The software R will be used for all computational needs.

**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 regression in all its diversity in data analysis. Channel the students into recognizing the regression problem when one presents itself. 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. A substantial level of activity is devoted developing cause and effect models using the data collected. Regression is the predominant methodology used in the development of models. The concept of regression is pursued in all its manifestations. The computing software R will be used to assist us in our quest.

**OBJECTIVES:**
At the conclusion of the course, the student will be able to:
1. comprehend the role of regression, in general, in developing cause and effect models;
2. learn the mechanics of multiple regression;
3. indulge in graphical presentations;
4. grasp when to go after model selection;
5. delve in the world of Box-Cox transformations;
6. build robust regression models;
7. pursue nonparametric inference through bootstrap and cross-validation;
8. tackle lasso and least angle regression;
9. analyze quantitative responses with excessive number of zeros;
10.explore multivariate multiple regression;
11.penetrate the realm of logistic regression;
12.evaluate usefulness of reduced rank regression.

**A SPECIAL FEATURE:** R, free statistical software, will be introduced to the students. Successful completion of homework needs the statistical software.

**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. 50 points
2. Mid-term Exam: October 23, 2014 20 points
3. Final Exam: December 04, 2014 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

**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.

TENTATIVE COURSE OUTLINE

1. History of Regression + Introduction to R
2. Simple Linear Regression + Diagnostics
3. Multiple linear Regression
4. Model Selection + Akaike Information Criterion (AIC)
5. Box-Cox Transformations
6. Influential Observations + Outliers + Robust Regression
7. Non-parametric Regression – LOWESS
8. Multivariate Multiple Regression
9. Lasso
10. Least Angle Regression
11. Responses with Excessive zeros
12. Poisson Regression
13. Counts with Excessive zeros – zero-inflated models
14. Logistic Regression
15. Cox Regression
16. Quantile Regression
17. Classification and Regression Trees
18. Boolean Regression

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