Course: 26 BE 7080 & 26 PH 7080 Spring, 2014
Course No: 009260
Title: Analysis of Internet Health Data
Credits: 3
Instructor: MB Rao
Office: 247 Kettering Lab
E-mail: marepalli.rao@uc.edu
Telephone: 513 558 3602
Hours: 2:00-3:50 Monday and Wednesday
Venue: Kehoe Auditorium
Office Hours: Monday 1:00 – 2:00
Wednesday 1:00 – 2:00
By appointment

Course Description: Sampling techniques; Simple random sampling; Stratified random sampling; Cluster sampling; Systematic sampling; Estimates and standard errors; Internet Health Data sets; NIS (National Inpatient Sample); KID (Kid’s Inpatient Database); SID (State Inpatient Database); AARP Diet and Health Cohort Study; Physician’s Health Study; Women’s Health Study; Women’s anti-oxidant cardiovascular study; Nurses’ Health study; others; Projects from the studies

Text Book: None
My notes are self-sufficient.

Reference Books:

Prerequisite: 26 BE 7022 – Introduction to Biostatistics or equivalent

Course Objectives: Teach some sampling techniques and data analyses that will be helpful to analyze internet data. Give training in using the software R for data analyses. Introduce a variety of internet health data sets. Let each student pick up
one data set or a subset of a health data set. Guide the student in analyzing the data. Help the student to produce a research paper for publication.

**Purpose:** The internet is a treasure trove of health data. The students are introduced to a number of data sets. Let each student pick up one dataset and set about analyzing the data after formulating a number of research questions. Teach the relevant statistical methods. The students are expected to produce a research paper for possible publication.

**INSTRUCTIONAL METHODS:**
1. Lectures.
2. Group projects worked out in the class.

**GRADING:**
1. Five homework sheets on methodology 30 points
2. Five interim presentations on the progress of research 20 points
3. Research paper 30 points
4. Final presentation 20 points

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

**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

After presenting a series of lectures on data analysis, the students are introduced to a number of internet health data sets. The main focus in this class will be on National Inpatient Sample (NIS) and Kid’s Inpatient Database. In NIS, admission-discharge data are available for 24 years starting from 1988 to 2011. The size of data for each year is approximately 8,000,000 rows and 300 columns. One of the columns is DRG code. The code from 001 to 999 codifies the reason why the patient was admitted into the hospital in the first place. The admission data was collected following the stratified-systematic-cluster sampling procedure. Students will be taught how to obtain national estimates from the data on hand. Subsets of the data based on selections of DRG code will be assigned to students on mutual agreement. The instructor has published a paper following this route. The instructor has NIS and KID in his possession in Ascii format.

<table>
<thead>
<tr>
<th>Topic</th>
<th># Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sampling Techniques: Simple; Stratified; Cluster; Systematic</td>
<td>3</td>
</tr>
<tr>
<td>2. Estimates and Standard Errors of Population Mean; Variance; Proportion; Size; Total and standard errors</td>
<td>3</td>
</tr>
<tr>
<td>3. Introduction to R</td>
<td>3</td>
</tr>
<tr>
<td>4. Graphics with R</td>
<td>3</td>
</tr>
<tr>
<td>5. Analysis of Variance + Regression</td>
<td>3</td>
</tr>
<tr>
<td>6. Introduction to Internet Health Data Sets</td>
<td>2</td>
</tr>
<tr>
<td>7. NIS and KID</td>
<td>1</td>
</tr>
<tr>
<td>8. Introduction to DRG code + ICD-9 + ICD-10 codes</td>
<td>1</td>
</tr>
<tr>
<td>9. Assigning data sets</td>
<td>2</td>
</tr>
<tr>
<td>10. In-class working on the projects + interim presentations</td>
<td>18</td>
</tr>
<tr>
<td>11. Final presentations</td>
<td>3</td>
</tr>
</tbody>
</table>