

# UNM Statistics Qualifying Exam Take-Home

## August 2015

**Due 4:00pm Aug 11 Tuesday, 2015. Return to Ana Parra Lombard in the Math/Stat Dept Office SMLC 395.**

*Directions:* The answer to each problem should be presented as a summary. It should be word processed and double spaced, and be identified by your UNM ID number (do not include your name). **A suggested length of the report to each problem is no longer than 3 pages.** Create brief, well-organized appendixes for each problem.

In your data analyses, RAW AND UNINTERPRETED COMPUTER OUTPUT IS UNACCEPTABLE. You should have a caption by every figure and table that describes it and tells the reader briefly what you see. Organize the sections to tell the story you uncovered, not the circuitous path you may have taken to get there. Remember that even that best data analysis is worthless if your reader cannot understand it.

You may **not** consult any other person when working on this exam or discuss your exam with anyone else regardless of whether or not the person is taking the exam. You may use your course notes as well as any available books or web resources for the exam. Questions pertaining to clarification about these questions can be directed to Yan Lu, luyan@math.unm.edu.

- (50 pts) The data for this problem is available at “<http://www.math.unm.edu/~luyan/senic.txt>”. The primary objective of the Study on the Efficacy of Nosocomial Infection Control (SENIC project) was to determine whether infection surveillance and control programs have reduced the rates of nosocomial (hospital-acquired) infection in United States hospitals. This data set consists of a random sample of 113 hospitals selected from the original 338 hospitals surveyed.

Each line of the data set has an identification number and provides information on 11 other variables for a single hospital. Following is the description of the variables:

Variable Number	Variable name	Description
1	<i>Identification number:</i>	1-113
2	<i>Length of stay:</i>	average length of stay of all patients in hospital (in days)
3	<i>Age:</i>	average age of patients (in years)
4	<i>Infection risk:</i>	average estimated probability of acquiring infection in hospital
5	<i>Routine culturing ratio:</i>	ratio of number of cultures performed to number of patients without signs or symptoms of hospital-acquired infection, times 100
6	<i>Routine chest X-ray ratio:</i>	Ratio of number of X-rays performed to number of patients without signs of symptoms of pneumonia, times 100
7	<i>Number of beds:</i>	average number of beds in hospital during study period
8	<i>Medical school affiliation:</i>	1=Yes, 2=No
9	<i>Region:</i>	geographic region, where: 1=NE, 2=NC, 3=S, 4=W
10	<i>Average daily census:</i>	average number of patients in hospital per day during study period
11	<i>Number of nurses:</i>	average number of full-time equivalent registered and licensed practical nurses during study period (number full time plus one half the number part time)
12	<i>Available facilities and services:</i>	percent of 35 potential facilities and services that are provided by the hospital

Regress infection risk ( $Y$ : variable 4) against the other predictor variables **except medical school affiliation** (variable 8). Build an appropriate regression model to predict the infection risk using a subset of the variables. Make sure you tell the reader what you found and how you found it. Make sure you include important results in your report. Write a succinct, coherent, and complete summary of your analysis.

2. (50 pts) Refer to “senic” data in problem 1. The effects of region (factor A: variable 9), average age of patients (factor B: variable 3), and medical school affiliation (factor C: variable 8) on mean length of hospital stay (variable 2) are to be studied. For purpose of this ANOVA study, average age is to be classified into two categories:  $\leq 53.9$  and  $\geq 54$ .

Build a model that best describes the relationship between the mean length of hospital stay and the three factors . Make sure that you carefully assess all assumptions and write a succinct, coherent, and complete summary of your analysis.