Proj 1, STAT 6338

Name (Please Print):

Instructions: The total points for this project are 15 points.

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. (Reference: Special Issue, "The SENIC project", American Journal of Epidemiology 111 (1980), 465-653.)

Each line of the data set has an identification number and provides infomation on 11 other variables for a single hospital. The data presented here are for the 1975-76 study period. The 12 variables are:

Variable Number	Variable name	Description
1	Identification number	1-113
2	Length of stay	Average length of stay of all patients in hospital (in days)
3	Age	average age of patients (in years)
4	Infection risk	average estimated probability of acquiring infection in hospital (in percent)
5	routine culturing ratio	ratio of number of culture performed to number of patients without signs or symptoms of hospital-acquried infection, times 100
6	routine chest x-ray ratio	Ratio of number of X-rays performed to number of patients without signs or symptoms of pneumonia, times100
7	number of beds	Average number of beds in hospital during study period
8	Medical school affiliation	1=Yes 2=no
9	region	Geographic regions, where 1=NE, 2=NC, 3=s, 4=W
10	Average daily census	Average number of patients in hospital per day during study period
11	Number of nurses	Average number full-time equivalent registered and licensed prectical nurses during sutdy period
12	Available faciliteies and services	Percent of 35 petential facilites and services that are provided by the hospital

1. (a) Test whether or not the mean infection risk (variable 4) is the same in the four geographic regions (variable 9); use $\alpha = .05$. Assume that ANOVA model is applicable. State the alternatives, conclusion.

   (b) Obtain confidence intervals for all pairwise comparisons between four regions, use the Tukey

procedure and a 90 percent family confidence coefficient. Interpret your result and state your findings. Prepare a line plot of the estimated factor level means and underline all nonsignificant comparisons.

(c) For the same family confidence coefficient, try a different pairwise comparison procedure. Interpret your result and state your findings.

2. The effect of average age of patient (variable 3) on mean infection risk (variable 4) is to be studied. For purposes of this ANOVA study, average age is to be classified into four categories: under 50, 50-54.9, 55.0-59.9, 60.0 and over. Assume that ANOVA model is applicable. Test whether or not the mean infection risk differs for the four age groups. Control the α risk at .10. State the alternatives and conclusion.

3. Conduct a test of whether or not mean length of stay (variable 2) is the same in the four geographic regions. Then do the following questions.

(a) Obtain the residuals and prepare aligned residual dot plots by region. Are any serious departure from ANOVA model?

(b) Examine by means of the Brown-Forsythe test whether or not the geographic region error variances are equal.

(c) For each geographic region, calculate $\bar{Y}_i$ and s_i. Examine the three relations found in the table on page 791 and determine the ransformation that is the most appropriate one here. What do you conclude?

(d) Use the Box-Cox procedure to find an appropriate power transformation of Y. Evaluate SSE for the values of λ given in Table 18.6. Does $\lambda = -1$, a reciprocal transformation, appear to be reasonable, based on the Box-Cox procedure?

(e) Use the reciprocal transformation $Y' = 1/Y$ to obtain transformed response data. Fit ANOVA model to the transformed data and obtain the residuals.

(f) Examine by means of the Brown-Forsythe test whether or not the geographic region variances for the transformed response variable are equal. use $\alpha = .01$.