Paper 475.330 **Assignment 3** **Due: Thursday, 6 May 1999**

The data for this assignment come from the "Study on the Efficacy of Nosocomial Control" or **SENIC** Project. Data was collected from hospitals in the United States to determine whether infection surveillance and control programs have reduced the rates of nosocomial (hospital-acquired) infection. The data you are asked to analyse consists of a random sample of 113 hospitals from the 338 hospitals surveyed for the SENIC project.

A data frame called senic.df has been created in Splus on the advanced lab server. This data frame contains the following variables:

stay: the average length of stay of all patients in the hospital (in days)

risk: the average estimated probability of acquiring infection in the hospital (in percent)

culture: the ratio of number of cultures performed to number of patients without symptoms of hospital-acquired infections, times 100

xray: the ratio of number of X-rays performed to number of patients without symptoms of pneumonia, times 100

region: a factor indicating geographic region: 1 = Northeast, 2 = North-central, 3 = South, and 4 = West

nurses: the average number of full-time nurses employed at the hospital during the study period

services: the percent of 35 potential facilities and services that are provided by the hospital

Find a suitable model for risk as a function of the other variables in the dataset. Use your model to discuss how risk is related to these other variables. Compare the fitted regression surfaces for the four different regions.

As usual, your assignment should consist of two main components. The first part should be a report that is suitable to be read by a hospital administrator. It should explain your results in non-technical terms. The second part will be a statistical appendix where you explain how you analysed the data and came to your conclusions.