Computational Statistics

In order to to determine whether infection surveillance and control programs have reduced the rates of nosocomial (hospital-acquired) infection in United States hospitals (the primary objective of the study), it was recorded a data set consisting 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. The data presented here are for the 1975-1976 study period. The 12 variables are:

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

Perform a thorough statistical analysis of the data, in particular using MCMC methods to select a suitable generalized linear model and interpret the associated results concerning the infection risk study.

Reference: Kutner M, Nachtsheim C, Neter J, Li W (2005). Applied Linear Statistical Models, 5th Edition. New York: McGraw-Hill/Irwin.