Problem Set #4 – Hospital Infection Rates

There is often an increased risk for patients acquiring an infection during their stay in a hospital. The medical community would certainly like to reduce the rate of infection in their hospitals, and so they are interested in identifying factors that contribute to infection rates. Data from n = 113 hospitals in the United States is collected and the variables are y = infection rate (no. per 100 patients), x1 = average patient age, x2 = average length of patient stay (no. of days), x3 = measure of how many x-rays are given in the hospital. For each of the following questions, assume that your audience are nurses with moderate statistical training. Please attach your clearly commented code (R or SAS) to the back of your answers as an appendix.

1. In your own words, summarize the overarching problem and any specific questions that need to be answered using the hospital data. Specifically, what can be gained by analyzing the *relationship* between the covariates and infection rates, and what can be gained by accurate *predictions* of infection rates? Discuss how statistical modeling will be able to answer the posed questions.
2. Perform appropriate exploratory analysis to assess if a multiple linear regression (MLR) model is suitable to analyze the hospital infection data. Justify your answer using any necessary graphics and relevant summary statistics that would suggest an MLR model would be successful at achieving the goals of the study. Provide a brief discussion on why an MLR model is or is not appropriate.
3. Regardless of your answer to #2, write out (in mathematical form with greek letters) a MLR model that would help answer the questions you stated in #1. Clearly state any assumptions you are using in your model. Provide an interpretation of at least 1 slope coefficient included in your model.
4. Fit your model in #3 to the hospital data and summarize the results by displaying estimated coefficients in a table (do NOT just provide a screen shot of the R or SAS output). Interpret at least 1 of the coefficients in the context of the problem. Identify which variables, if any, have a significant impact on predicting infection rates, and which, if any, do not.
5. Justify your model assumptions using appropriate graphics or summary statistics. Provide discussion of your assessment of the model fit and validity of assumptions on the level of your target audience.

1. A nurse wishes to make a prediction for the average number of infections per 100 patients at her hospital. She collects the following data: length of stay = 8 days, avg. age of patient = 50, and no. of x-rays = 82. Use your model in #4 to make this prediction, and include an appropriate margin of error.
2. The medical community would like to know how accurate the predictions are using this MLR model. Carry out a cross-validation analysis using 10,000 iterations. Display appropriate numerical summaries and interpret these summaries in the context of this problem (at the level of your target audience).
3. Given the above results, write a short paragraph summarizing the findings of this study. Specifically, what if any benefit is there for hospitals and the medical community? Try to answer any questions you raised in (1).