**LAB 3: MULTIPLE LINEAR REGRESSION[[1]](#footnote-1)**

**Problem 1.** A hospital administrator wished to study the relation between patient satisfaction (Y) and patient’s age (, in years), severity of illness (, an index), and anxiety level (, an index). The administrator randomly selected 46 patients and collected the data presented below, where larger values of Y, , and are, respectively, associated with more satisfaction, increased severity of illness, and more anxiety.

Table

Description automatically generated

a. Prepare a dot plot for each of the predictor variables. Are any noteworthy features revealed by these plots?

b. Obtain the scatter plot matrix and the correlation matrix. Interpret these and state your principal findings.

c. Calculate coefficients in regression model for three predictor variables and state the estimated regression function. How is interpreted here?

d. Obtain the residuals and prepare a box plot of the residuals. Do there appear to be any outliers?

e. Plot the residuals against , each of the predictor variables, and each two-factor interaction term on separate graphs. Also prepare a normal probability plot. Interpret your plots and summarize your findings.

f. Test whether there is a regression relation; use α = .10. State the alternatives, decision rule, and conclusion. What does your test imply about ? What is the P-value of the test?

g. Conduct the Breusch-Pagan test for constancy of the error variance, assuming log ; use . State the alternatives, decision rule, and conclusion.

**Problem 2**. Refer to the SENIC data set in Appendix C.1. Two models have been proposed for **predicting the average length of patient stay in a hospital** ().

- Model I utilizes as predictor variables age (), infection risk (), and available facilities and services ().

- Model II uses as predictor variables number of beds (), infection risk (), and available facilities and services ().

a. Prepare a dot plot for each of the predictor variables. What information do these plots provide?

b. Obtain the scatter plot matrix and the correlation matrix for each proposed model. Interpret these and state your principal findings.

c. For each of the two proposed models, fit first-order regression model (6.5) with three predictor variables using R’s function.

d. Calculate R-squared, adjusted R-squared for each model. What do they indicate here? Is one model clearly preferable in terms of this measure?

e. For each model, obtain the residuals and plot them against , each of the three predictor variables, and each of the two-factor interaction terms. Also prepare a normal probability plot of the residuals for each of the two fitted models. Interpret your plots and state your findings. Is one model clearly more appropriate than the other?

**-- END --**

1. Reference: Chapter 6, Kutner’s book. [↑](#footnote-ref-1)