

HOMework #10

Read:

- Chapter 11 in ‘Regression Analysis by Example’.

R Assignment:

Solve the following questions using R. Hand in your R code and output file for Questions 1-2, as well as answers to all questions.

1. A hospital surgical unit was interested in predicting survival time in patients undergoing a particular type of liver operation. From a random sample of 54 patients, information on the patient's survival time, blood clotting score, prognostic index, enzyme function test score and liver function test score were extracted. The data can be found on the web page:

<http://www.stat.columbia.edu/~martin/W2024/Data/Surgical.txt>

- (a) Fit a multiple regression model using the natural logarithm of survival time as the response variable and the other four variables as explanatory variables.
 - (b) Conduct an F-test for the overall fit of the regression model in (a). Comment on the results.
 - (c) Test each of the individual regression coefficients. Do the results indicate that any of the explanatory variables can be removed from the model?
 - (d) Perform variable selection by finding the subset model that minimizes the BIC criteria. State the ‘best’ model.
 - (e) Using the model from part (d) make appropriate diagnostic plots to determine whether the model assumptions are valid. Comment on the plots.
2. A paper company is interested in making its operations more efficient. They collect data on the total manufacturing cost per month (in dollars), the total production of paper per month (in tons), the total number of machine hours per month, the total variable overhead cost per month (in thousands of dollars) and the total number of labor hours each month. The data can be found at

www.stat.columbia.edu/~martin/W2024/Data/papercompany.txt

Suppose we are interested in fitting a regression model using cost as the response variable and some subset of the other variables as explanatory variables.

- (a) Perform variable selection by finding the subset model that minimizes the BIC criteria. State the ‘best’ model.
- (b) Perform variable selection using forward selection. State the ‘best’ model.
- (c) Perform variable selection using backward selection. State the ‘best’ model.

- (d) The results in (a) - (c) indicate that two variables can be removed. Which ones? Perform a partial F-test to determine whether the two variables can be removed from the model.
- (e) Using the model from part (d) make appropriate diagnostic plots to determine whether the model assumptions are valid. Comment on the plots.