Homework 4 – Due Feb 24th @11pm

1. 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 (papercompany.txt) can be found in Homework 6 link in Bruinlearn. Use R to answer the following questions.

(a) Explore the relationship among the variables :

- Generate a correlation coefficient matrix. Try *cor(data),* in R.
- Generate a scatter plot matrix. Try *pairs(data)* in R.
- What do you feel about the relationship among the variables? Comment.

(b) Fit a multiple regression using cost as the response variable, and the other four variables as explanatory variables. Write down the regression equation.

(c) Conduct the F-test for the overall fit of the regression. What conclusions can you draw?

(d) What proportion of the variation in cost has been explained by the regression?

(e) Test each of the individual regression coefficients. Interpret the result for each test. Which of the variables are significant?

(f) Perform a partial F-test to determine whether the variables associated with overhead and labor hours can be removed from the model. Comment on the results of the test.

2. Exercise 5.4.3