stk310 PRACTICAL ASSIGNMENT A5

A food company collected data on the sales of and the price for its new cereal, **paul-se-pap**, at 16 randomly selected stores in Pretoria. They also obtained the price of **Three Bears Bran**, a popular cereal from another food company, at these 16 stores. The dataset to be used in this assignment is given in the file cereal.csv on clickUP. Specifically consider the following variables for $i = 1, 2, \ldots, 16$:

- Y_i ➡ number of boxes of **paul-se-pap** sold
- X_{2i} ➡ price in Rand for a box of **paul-se-pap**
- X_{3i} ➡ price in Rand for a box of **Three Bears Bran**
- X_{4i} ➡ number of boxes of **Three Bears Bran** sold

Where applicable, use appropriate procedures in SAS as well as functions in R to answer the following questions.

Question 1

Considering all three available explanatory variables, find the best possible linear regression model based upon the adjusted coefficient of determination.

Question 2

Consider only the number of boxes of **paul-se-pap** sold, the price in Rand for a box of **paul-se-pap** and the price in Rand for a box of **Three Bears Bran**.

(a) Calculate and interpret the partial correlation coefficient between the demand for **paul-se-pap** and the price of **paul-se-pap**.

(b) Calculate and interpret the partial correlation coefficient between the demand for **paul-se-pap** and the price of **Three Bears Bran**.

Question 3

Ignoring the number of boxes of **Three Bears Bran** sold, suppose the demand function for **paul-se-pap** is

$$Y_i = \alpha X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i}.$$

The corresponding linear regression model is

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i,$$

where $\beta_1 = \ln\alpha$.

(a) Fit the linear regression model to the dataset using ordinary least squares regression and interpret $\hat{\beta}_2$ and $\hat{\beta}_3$.

(b) Use the Shapiro-Wilk test to verify whether the assumption that u_i follows a normal distribution is true.

(c) Test whether each of the regression coefficients is statistically significant.

(d) Test whether the overall regression model is statistically significant.

(e) Test whether the two partial regression coefficients are equal.

(f) Use a 99% confidence interval to test whether the mean demand for **paul-se-pap** will decrease by 3% if there is an increase of 1% in the price for a box of **paul-se-pap**, while the price of **Three Bears Bran** remains unchanged.

(g) Assuming that the price of **paul-se-pap** and the price of **Three Bears Bran** are both R36 per box, estimate the mean number of boxes of **paul-se-pap** sold and calculate a 99% confidence interval for the mean number of boxes of **paul-se-pap** sold.

Question 4

Consider the linear regression model from **Question 3**.

Use the matrix approach to regression analysis to answer the questions below.

(a) Fit the linear regression model to the dataset using ordinary least squares regression.

(b) Calculate and give the values of $\hat{\sigma}^2$, R^2 and $\bar{R}^2$.

(c) Test whether each of the regression coefficients is statistically significant.

(d) Test whether the overall regression model is statistically significant.