

Linear Statistical Models

Week-7: Graded Assignment

Subjective Assignment: (Manual-grading)

Max. Marks: 50

Note: *R* is required for this assignment.

1. An analyst collected the data (“forbes.txt”) to study the relationship between atmospheric pressure and the boiling point of the water. The variables recorded in the dataset are:

- *Temp* ($^{\circ}F$): Boiling point in degrees Fahrenheit
- *Pressure* (Inches Hg): Atmospheric pressure in inches of mercury
- $Lpres = 100 \times \log_{10}(Pressure)$

Based on the given information, answer the following questions.

- (a) Read the dataset as a data frame in *R*. [1 Mark]
- (b) Using the ‘*ggplot*’ in *R*, plot the scatter plot between the ‘*Temp*’ and ‘*Pressure*’. Based on the obtained plot comment on the relationship between two variables. [2 Marks]
- (c) Obtain the least square estimates using normal equations and *lm()* function in *R* for the model:

$$Lpres = \beta_0 + \beta_1 Temp + \epsilon$$

where, $\epsilon \sim N(0, \sigma^2)$.

Compare the least square estimates computed using the above-mentioned two methods. [5 Marks]

- (d) Using ‘*ggplot*’ in *R*, add a regression line of the fitted model in the scatter plot between ‘*Lpres*’ and ‘*Temp*’. [3 Marks]
- (e) Compute the Residual Sum of Squares (*RSS*) and Total sum of squares (*TSS*) for the above fitted model. [4 Marks]
- (f) Find Residual standard error (*RSE*) for the above fitted model. And, comment on the obtained value. [3 Marks]
- (g) Find R^2 for the above fitted model. And, interpret the value obtained. [3 Marks]

2. If the model of the above question is modified as follows:

$$Lpres = \beta_0 + \beta_1 u_1 + \epsilon$$

where, $u_1 = \frac{1}{Ktemp}$

And, $Ktemp$ represents the temperature in degrees Kelvin which is given by:

$$Ktemp = 255.37 + \frac{5}{9} \times Temp.$$

Based on the information given, answer the following questions:

- (a) Using the ‘ggplot’ in R, plot the scatter plot of ‘Lpres’ versus u_1 and comment on the relationship between two variables. [2 Marks]
 - (b) Fit the linear regression for the given model, and summarize your results. [4 Marks]
 - (c) Plot the fitted values obtained from model fitted in 1(c) versus the fitted values from the model 2(b). On the basis of it, comment on if a model can be preferred over the other. [4 Marks]
3. Consider a model as

$$Y = -1 + 5X + u$$

- .
- (a) Generate 80000 random values from the uniform distribution with parameters $a = 0$ and $b = 30$ for X . [1 Mark]
 - (b) Generate 80000 random values from the normal distribution with parameters $mean = 0$ and $sd = 20$ for u . [1 Mark]
 - (c) Using the above generated values, compute the values of Y . And, store the respective values of X and Y in a data frame. [2 Marks]
 - (d) Take a sample of size 100 from the created data frame. Using the sampled values of X and Y , fit a linear model, i.e. $Y = \beta_0 + \beta_1 X + u$. [3 Marks]
 - (e) Repeat the steps in (d) 8000 times, and store the values of the least square estimates of the coefficients in a matrix. [4 Marks]
 - (f) Plot the histograms for the values of $\hat{\beta}_0$ obtained in part (e). Also, find the distribution for $\hat{\beta}_0$ and plot the density curve of the identified distribution on the plotted histogram. [4 Marks]
 - (g) Plot the histograms for the values of $\hat{\beta}_1$ obtained in part (e). Also, find the distribution for $\hat{\beta}_1$ and plot the density curve of the identified distribution on the plotted histogram. [4 Marks]