Project-Math 20011

The final project is to apply the multiple linear regression model to a data set "iris" given in ggplot2. You need to consider the following steps to analyze the data.

1. Describe the data set. For example, what is the data set about? How many variables are there? What do they represent?
2. Consider the variable "Sepal.length" as the response. Graphically demonstrate the relationship between the response and other variables.
3. Establish the multiple regression model of the response and other variables. Estimate all the unknown parameters in the model, and establish the fitted multiple regression.
4. Use the ANOVA to test if all the slopes are zeros. Report the p-value and the corresponding conclusion.
5. Test if the slope related to the predictor "Petal.Width" is zero. Report the p-value and the corresponding conclusion.
6. Construct a 90% confidence interval for the slope related to the predictor "Petal.Length". Is zero included by the interval?

After analyzing the dataset, you need to check if all the assumptions are satisfied by the regression model.

1. Use both graphical and testing procedure to check the assumption "linearity" and "constant variance".
2. Use the QQ-plot and a testing procedure to check the normality assumption.
3. Use Cook's distance to check if the data consist of any influential outliers.

You need to submit a detailed report in pdf for all the results you obtain. Also attach the R code in the end of your report.