Statistical Learning, Exercise 4 : Linear regression with R

Question 1

Download from the ILIAS website the file Education4.txt containing a new version of the Education dataset (without outliers). The variable **Gender** is now coded as *male* or *female*. The goal is still to predict the variable **Wage**.
To do that, build two linear models, one for men and one for women. Visualize graphically the two linear relationships on the same graph. Explain what you see and draw some conclusions about the two linear relationships you have just built.

Question 2

Download from the ILIAS website the **Computer** dataset (ComputerData.txt). It contains various variables to predict the performances of a computer system (response time). Read the description of the dataset in the file ComputerDescription.pdf.
The performance of the system is indicated by the variable *PRP*. First, check the different variables available to predict the system's performances. Which variables do you think can or cannot be used to explain the system's performances. Why?

Question 3

You can only use one variable to explain the value of *PRP*. Which one to you choose and why? Does your model significantly explains something and how can you prove it? Can you give a confidence interval for the slope?

Question 4

Display graphically the linear model you obtained at question 3.

Question 5

Download from the ILIAS website the **Cars2** dataset (Cars2Data.txt). This dataset contains different variables used to predict car's performance indicated by the variable *mpg*.
Check the different variables you have to predict *mpg*. Which variables do you think can or cannot be used to explain performances? Why?

Question 6

You can only use one variable to explain the value of *mpg*. Which one do you choose and why? Does you model significantly explains something and how can you prove it? Can you give a confidence interval for the slope?

Question 7

Visualize graphically the (linear) relationship you have found.