Statistical Learning, Exercise 2 : Plot() and t-test (for the mean)

Question 1

Download the car seats data set from the ILIAS website (Carseats.txt). Using the plot() function, show *graphically* the possible relationship between the dependent variable **Sales** and the independent variable **Price**. Can you change the title of the plot as well as the labels appearing in the two axes? It seems that there is an outlier, can you highlight it in red?

Question 2

Download the education data set from the ILIAS website (Education.txt). As the variable **Country** is a factor (categorical), select the wage and education values corresponding to each of the two possible country values. Show graphically the possible relationship between the independent variable **Education** and the dependent variable **Wage** but only when considering the observations corresponding to the *USA* and to the *Canada*. Can you plot on the same graphics the information for the US and for the Canada but with different colors? Can you add a legend with the corresponding colors?

Question 3

Download from the ILIAS website the Mean20 data set (filename: Mean20.txt). This data set is composed by a single variable (*time*), the time delay in minutes between two calls in an info-center.

Compute the mean, the median, the standard deviation, the minimum and maximum value of the variable *time*. Is there any outliers and possible invalid samples? Do you need to preprocess this list?

Question 4

We suppose that the mean delay between two calls is 7.05 minutes. Can you test this hypothesis with the data available? What is your conclusion? Do you see a difference when considering the original values and the preprocessed values?

Question 5

For John, the average delay is greater than 7.05. Thus the only credible alternative hypothesis must take account of this fact. How can you test John's hypothesis?