Modèle linéaire et généralisations – TP2

Data

Cepages dataset

We consider the dataset "CepagesB.csv" of 36 observations and 10 variables. We want to explain $Y = pH$ with respect to the others variables.

- pH : pH of the wine.
- Origine : factor which admits $J = 2$ levels : Bordeaux and Bourgogne.
- Couleur : factor which admits $J = 2$ levels : Blanc and Rouge.
- Alcool : It is the alcohol content of the wine.
- Malique : Malic acid that reflects greenness / biting wine (green apple).
- Tartrique : Tartaric acid that reflects hardness / structure of the wine (the acid most present in the grapes).
- Citrique : Citric acid that reflects freshness of the wine (lemony taste).
- Acetique : Acetic acid is a natural organic acid, the main constituent of the volatile acidity of a wine.
- Lactique : Lactic acid is an organic acid that plays a role in various biochemical processes.
- AcTot : Total acidity.

Question 0

Download the file "CepageB.csv".

Question 1

Check the nature of the dataset and make the necessary changes. Check the types of covariates and get rid of useless variables.

Question 2

Display the Table of counts according to Colour and Origin.

Question 3

Display the table of the average pH according to Couleur modalities and average pH. Display the table of the average pH according to Couleur modalities and Origine modalities.

Question 4

Plot the regression lines of ph over AcTot for the different modalities of Couleur. Display boxplots for pH based on color modalities. Comment. (You can spend time on this question to obtain beautiful plots with ggplot2)

Question 5

Plot the regression lines of ph over AcTot for the different modalities of Origine. Display boxplots for pH based on color modalities. Comment.

Question 6

Fit an ANOVA model explaining the pH as a function of the color. Write the mathematical model and give an interpretation of the coefficients. Comment the plots.