Initiation to R Software

Pierre Michel

Master AMSE 1st year, 2019

Problem Set IV

Exercice 1

a) Import the `data.frame` *iris* from the package `datasets`. The obtained `data.frame` contains the columns "Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width" and "Species". How many rows are there?

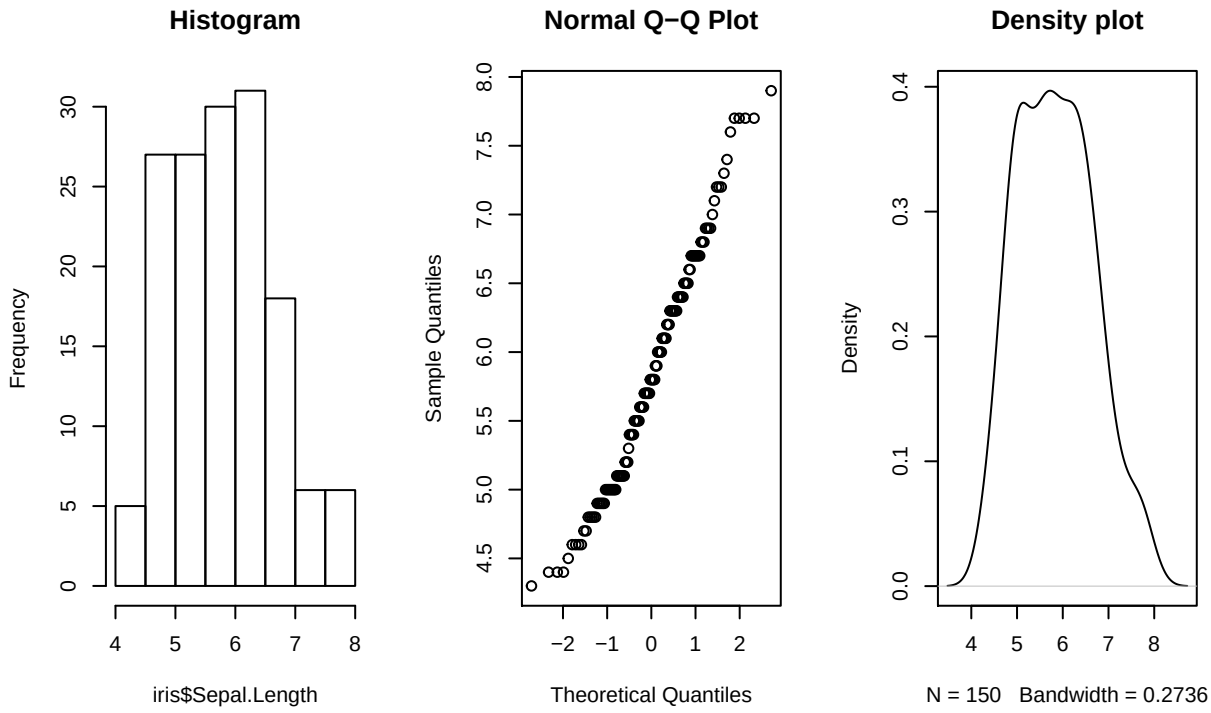
b) We focus on the distribution of the variable "Sepal.Length". Compute its mean (denoted m) and standard deviation (denoted s).

c) Study and comment the following graphs:

- histogram (`hist()`)
- tree (`stem()`)
- graph "Sepal.Length" versus normal distribution (`qqnorm()`)
- density (`plot(density())`)

Below is what you should get.

```
##
##   The decimal point is 1 digit(s) to the left of the |
##
##   42 | 0
##   44 | 0000
##   46 | 000000
##   48 | 00000000000
##   50 | 0000000000000000000
##   52 | 00000
##   54 | 0000000000000
##   56 | 00000000000000
##   58 | 0000000000
##   60 | 000000000000
##   62 | 0000000000000
##   64 | 000000000000
##   66 | 0000000000
##   68 | 0000000
##   70 | 00
##   72 | 0000
##   74 | 0
##   76 | 00000
##   78 | 0
```



d) Generate 8 samples of same size as "Sepal.Length", following a normal distribution $\mathcal{N}(m, s)$. Plot in the same graphical window: 9 histograms, 8 obtained through sample simulation, and the one of "Sepal.Length". Comment.

e) Do the same for the other graphs (`plot(density())` and `qqnorm()`).

Exercice 2

a) Plot in a same graphical window 9 qq-plots, the first 3 are obtained with samples drawn from a normal distribution $\mathcal{N}(0, 1)$ of size 10, the next 3 of size 100, the last 3 of size 1000. Comment.

b) Do the same thing for densities, with 9 samples drawn from a χ^2 distribution; the first 3 with samples of size 100, the next 3 of size 1000, the last 3 of size 10000. Comment.

Exercice 3

Import the dataset *airquality* from the package `cluster`. The `data.frame` obtained contains the columns "Ozone", "Solar.R", "Wind", "Temp", "Month", "Day". Plot the graph of "Ozone" in function of "Solar.R" (use `plot()`), then plot the same graph for the three levels of temperature ("Temp"), use `coplot()`. Comment.