Exercice 1. Familiarisation avec le jeu de donnees.

1. Quels sont les noms de variables?

[1] "Sex" "Bwt" "Hwt"

1. Combien de variables et dobservations contient le jeu de donnees cats?

[1] 144

1. A chez les 10 premieres observations du tableau. Quel est le sex et le poids du c ur du chat numero 6

Sex Bwt Hwt

1 F 2.0 7.0

2 F 2.0 7.4

3 F 2.0 9.5

4 F 2.1 7.2

5 F 2.1 7.3

6 F 2.1 7.6

7 F 2.1 8.1

8 F 2.1 8.2

9 F 2.1 8.3

10 F 2.1 8.5

Chat numéro 6 : sexe :Femelle ; poids : 2.1 kg Heartweight 7.6 g

> attach(cats)

> Bwt

[1] 2.0 2.0 2.0 2.1 2.1 2.1 2.1 2.1 2.1 2.1 2.1 2.1 2.2 2.2 2.2 2.2 2.2 2.2

[19] 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.4 2.4 2.4 2.4 2.5 2.5

[37] 2.6 2.6 2.6 2.7 2.7 2.7 2.9 2.9 2.9 3.0 3.0 2.0 2.0 2.1 2.2 2.2 2.2 2.2

[55] 2.2 2.2 2.2 2.2 2.3 2.4 2.4 2.4 2.4 2.4 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5

[73] 2.6 2.6 2.6 2.6 2.6 2.6 2.7 2.7 2.7 2.7 2.7 2.7 2.7 2.7 2.7 2.8 2.8 2.8

[91] 2.8 2.8 2.8 2.8 2.9 2.9 2.9 2.9 2.9 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0

[109] 3.1 3.1 3.1 3.1 3.1 3.1 3.2 3.2 3.2 3.2 3.2 3.2 3.3 3.3 3.3 3.3 3.3 3.4

[127] 3.4 3.4 3.4 3.4 3.5 3.5 3.5 3.5 3.5 3.6 3.6 3.6 3.6 3.7 3.8 3.8 3.9 3.9

>

Exercice 2. Caracterisation des series statistique

A laide du tableau precedent, calculez pour la variable Bwt la moyenne empirique letendu et

[1] 2.723611

la mediane, les quartiles dordre 1, 2 et 3

> quantile(Bwt,0.25)

25%

2.3

> quantile(Bwt, 0.50)

50%

2.7

> quantile(Bwt, 0.75)

75%

3.025

la variance

> var(Bwt)

[1] 0.2355225

et ecart type empiriques,

> sd(Bwt)

[1] 0.4853066

letendu interquartile.

> IQR(Bwt)

[1] 0.725

Commentez les resultats obtenus

The descriptive statistics of the **Bwt** variable show an empirical mean of **2.72**, close to the median (**2.7**), suggesting a relatively symmetric distribution. The interquartile range (**0.725**) and standard deviation (**0.4853**) indicate **moderate dispersion** around the mean. The first quartile (**2.3**) and third quartile (**3.025**) show that **50% of the values lie within this range**. The low variance (**0.2355**) confirms a certain homogeneity in the data, though graphical analysis (histogram or boxplot) would help detect any potential outliers.

> summary(cats)

Sex Bwt Hwt

Length:144 Min. :2.000 Min. : 6.30

Class :character 1st Qu.:2.300 1st Qu.: 8.95

Mode :character Median :2.700 Median :10.10

Mean :2.724 Mean :10.63

3rd Qu.:3.025 3rd Qu.:12.12

Max. :3.900 Max. :20.50

Comparing the descriptive statistics of **Bwt** from the previous analysis with the summary of the **cats** dataset, we observe that the values remain consistent. The **mean (2.724)** and **median (2.7)** are nearly identical to the previously computed values (**2.723611** and **2.7**, respectively). Similarly, the first quartile (**2.3**), third quartile (**3.025**), and interquartile range (**0.725**) are unchanged.

The **minimum (2.0) and maximum (3.9)** values confirm the total range of Bwt, which was not explicitly calculated before. The overall consistency between both sets of statistics suggests that our previous calculations were correct and aligned with the dataset’s summary. The additional **Hwt (Heart Weight)** variable provides more context but does not affect our analysis of Bwt.

Exercice 3. Bote a moustach

A graph with a line and a line

AI-generated content may be incorrect.

A graph of a diagram

AI-generated content may be incorrect.

Les **boxplots confirment les résultats du summary()**, tout en offrant une meilleure compréhension de la répartition et des éventuelles valeurs aberrantes.

Exercice 4. Histogramme de frequence

A graph of a bar graph

AI-generated content may be incorrect.

A graph with blue squares

AI-generated content may be incorrect.

A graph of a number of lines

AI-generated content may be incorrect.

A graph of a number of numbers

AI-generated content may be incorrect.

The histogram of cat weights (Bwt) shows different patterns depending on the number of bins used. With only 2 bins, the histogram is too coarse, making it difficult to observe the data distribution. At 20 bins, the distribution becomes clearer, revealing key trends in the data. Increasing to 200 bins introduces excessive detail, making the histogram appear noisy, while at 2000 bins, it resembles a scatter plot, losing its interpretability. A balance is needed, as too few bins obscure important patterns, while too many add unnecessary complexity. Generally, using around 10 to 30 bins provides a good trade-off between clarity and detail, with 20 bins being a reasonable choice for effectively visualizing the weight distribution.

> table(Sex)

Sex

F M

47 97

Bwt

2 2.1 2.2 2.3 2.4 2.5 2.6 2.7 2.8 2.9 3 3.1 3.2 3.3 3.4 3.5 3.6 3.7 3.8 3.9

5 10 14 13 9 10 9 12 7 8 11 6 6 5 5 5 4 1 2 2

>

> table(Hwt)

Hwt

6.3 6.5 7 7.1 7.2 7.3 7.4 7.6 7.7 7.9 8 8.1 8.2 8.3 8.4 8.5

1 2 1 1 2 3 1 2 1 5 1 1 1 2 1 3

8.6 8.7 8.8 9 9.1 9.3 9.4 9.5 9.6 9.7 9.8 9.9 10 10.1 10.2 10.4

1 4 3 4 4 2 3 2 6 2 2 2 2 8 3 2

10.5 10.6 10.7 10.8 10.9 11 11.1 11.2 11.3 11.4 11.5 11.6 11.7 11.8 11.9 12

1 4 1 1 2 3 1 2 1 1 3 2 1 2 1 3

12.1 12.2 12.3 12.4 12.5 12.7 12.8 12.9 13 13.3 13.5 13.6 13.8 14.1 14.3 14.4

1 2 1 2 2 3 1 1 3 3 2 1 1 1 1 2

14.8 14.9 15 15.4 15.6 15.7 16.8 17.2 20.5

2 1 1 1 1 1 1 1 1

>

2. Variable qualitativ

Barplot btw

A barcode diagram with numbers

AI-generated content may be incorrect.

Barplot(Hwt)

A graph of a graph

AI-generated content may be incorrect.

Pie(Hwt)

A circular object with numbers with Brighton Wheel in the background

AI-generated content may be incorrect.

> boxplot(Bwt-Hwt)

A graph with a line and a line

AI-generated content may be incorrect.

Exercice 6. \*

A chart with a red and blue box

AI-generated content may be incorrect.

Exercice 7

A diagram of a weight scale

AI-generated content may be incorrect.

A diagram of body weight vs heart weight

AI-generated content may be incorrect.

> # Display the results

> cov\_Bwt\_Hwt

[1] 0.9501127

> cor\_Bwt\_Hwt

[1] 0.8041274

A graph of body weight and heart weight

AI-generated content may be incorrect.

In this plot:

* **Males** will be shown in **blue**, and **females** will be shown in **pink**.
* This distinction will allow you to observe if there is any difference in the distribution of **Body Weight** and **Heart Weight** between males and females. For example, if males have consistently higher weights than females, you would observe that the male points cluster at higher values on both axes.