Question 3

> The following data represent measurements (in cm) of shell lengths collected from a deposit in Spain

---

1. Compute the mean, variance, and median of the shell lengths.

Let the number of sample be:

$$n = 33$$

Let the data in ascending order be:

$$x_1 = 1.23,\ \ x_2 = 2.77,\ ...,\ x_i,\ ...,\ x_n = 8.86, \quad i \in \{1, 2, ..., n\}$$

Then we have the mean:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{15\,984}{3\,300} \approx 4.844$$

Then we have the variance:

$$\sigma_x^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}^2 \approx 3.342$$

We have the index of the median:

$$i_m = \frac{n+1}{2} = 17$$

Then we have the median:

$$m = x_{i_m} = 4.4$$

Answer

- $\bar{x} = \frac{15\,984}{3\,300} \approx 4.844$

- $\sigma_x^2 \approx 3.342$
- $m = 4.4$

2. Determine the interquartile range (IQR) of the data.

Let the second quartile be:

$$Q_2 = m = 4.4$$

As well as the index of the second quartile:

$$i_{Q_2} = i_m = 17$$

Then we have the index of the first quartile (Q_1) and the index of the third quartile (Q_3):

$$\begin{aligned} i_{Q_1} &= \frac{i_m + 1}{2} &= 9 \\ i_{Q_3} &= \frac{i_m + n}{2} &= 25 \end{aligned}$$

Then we have the first quartile and the third quartile:

$$\begin{aligned} Q_1 &= x_{i_{Q_1}} &= 3.58 \\ Q_3 &= x_{i_{Q_3}} &= 5.24 \end{aligned}$$

Then we have the IQR:

$$\mathrm{IQR} = Q_3 - Q_1 = 1.66$$

Answer

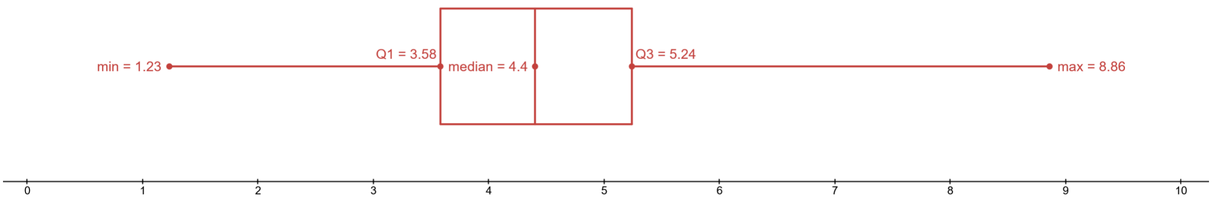
- $\mathrm{IQR} = 1.66$

3. Construct a boxplot to visualize the distribution of shell lengths.

Let the minimum ($\min$) and the maximum ($\max$) be:

$$\min = x_1 = 1.23$$
$$\max = x_n = 8.86$$

Then we have the boxplot:



Answer