First Statistics Homework

Semester 2025.2

Students: Arthur Lopes Pamplona, 570357 / Ana Livia Cordeiro, 567183,
Lecturer: Michela Mulas

Problem 1: Industrial Pollutant Gas

The daily emissions of a pollutant gas from an industrial plant were recorded 80 times, using a specific unit of measurement. The data obtained are given in Table 2.

Table 1: Pollutant gas emission data.

15.8	22.7	26.8	19.1	18.5	14.4	8.3	25.9	26.4	9.8	21.9	10.5
17.3	6.2	18.0	22.9	24.6	19.4	12.3	15.9	20.1	17.0	22.3	27.5
23.9	17.5	11.0	20.4	16.2	20.8	20.9	21.4	18.0	24.3	11.8	17.9
18.7	12.8	15.5	19.2	13.9	28.6	19.4	21.6	13.5	24.6	20.0	24.1
9.0	17.6	25.7	20.1	13.2	23.7	10.7	19.0	14.5	18.1	31.8	28.5
22.7	15.2	23.0	29.6	11.2	14.7	20.5	26.6	13.3	18.1	24.8	26.1
7.7	22.5	19.3	19.4	16.7	23.5	18.4					

- 1. Calculate the central tendency measures (mean, median, and mode) and the dispersion measures (range, variance, standard deviation, and coefficient of variation) for the dataset in Table 1. Interpret the results].

—-

- 2. Create a histogram and a boxplot for the emission data. Do the data appear to be symmetrically distributed? Are there any outliers?

Arthur: On this item the knowledge for graphics build is very important, the quantitative data shown in the problem can be viewed and interpreted by these graphics much easier than the table given.
In order to create a histogram its necessary to organize the data and divide into sections, creating the x section of the graph, and in these sections sum their frequencies to draw the y section of the graph.

For the data given, a division of sections of 3 is appropriate, the minimum value is 6.2 and the max value is 31.8, so the range of the graph is between 6 and 33.

Table 2: Frequencies of each interval

$[6, 9)$	3
$[9, 12)$	7
$[12, 15)$	9
$[15, 18)$	11
$[18, 21)$	21
$[21, 24)$	12
$[24, 27)$	11
$[27, 30)$	4
$[30, 33]$	1

Now creating the histogram:

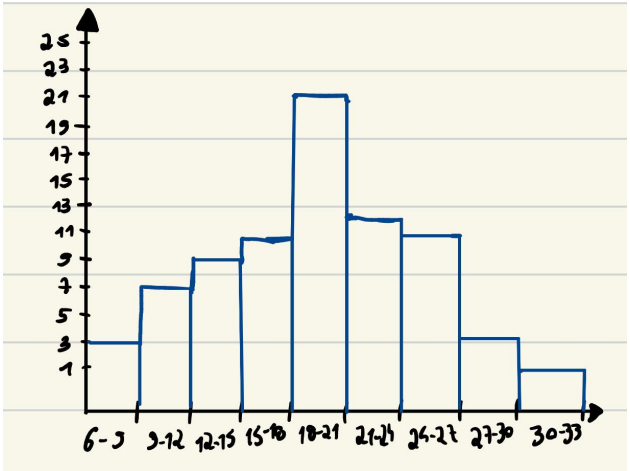

Figure 1: Histogram drew on tablet

R Histogram:

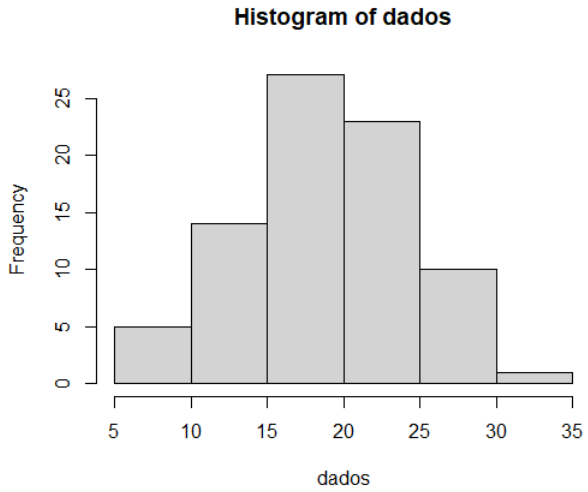

Figure 2: R's histogram

Comparing to R, my approach gives more detail about the data.

The next step is to create the boxplot of the emission data, to create the boxplot its necessary to divide the data in quartiles:

The first quartile is calculated as shown:

Q1 = wi + (N/4 - Ni-1)/li . wi = inferior limit of the quartile class N/4 = total divided by 4 Ni-1 = acumulated frequency before the quartile class li = frequency density

Creating the table of acumulative frequencies:

Class	**Frequency**	**Accumulated frequency**
$[6, 9)$	3	3
$[9, 12)$	7	10
$[12, 15)$	9	19
$[15, 18)$	11	30
$[18, 21)$	21	51
$[21, 24)$	12	63
$[24, 27)$	11	74
$[27, 30)$	4	78
$[30, 33]$	1	79

Table 3: Adding the accumulated frequency

- Calculating the first quartile:

N/4 is 79/4 = 19.75, falling into the [15,18) class, with wi = 15 and Ni-1 = 19
The frequency density of the class is 11/3.

$$\textbf{Q1} = 15 + (19.75\text{-}19)/(11/3) = 15.21.$$

The median is: Q2 = wi + (N/2 - Ni-1)/li;

N/2 = 79/2 = 39.5, fitting into the (18,21] class, with wi=18 and Ni-1 = 30, the frequency density is 21/3 = 7. So the median is:

$$\mathbf{Q2} = 18 + (39.5\text{-}30)/7 = 19.36$$

The third quartile is: Q3 = wi + (3N/4 - Ni-1)/li.

3N/4 = 3.79/4 = 59.25, falling into the (18,21] class, with wi=18 and Ni-1 = 30
The frequency density is 21/3 = 7.

$$\mathbf{Q3} = 18 + (59.25\text{-}30)/7 = 22.17$$

Now the **inferior limit** is: Q1 - 1.5x(Q3-Q1) = 15.21 - 1.5x(22.17-15.21) = 4.77
The **superior limit** is: Q3 + 1.5(Q3-Q1) = 22.17 + 1.5x(22.17-15.21) = 32.61

Drawing the boxplot:

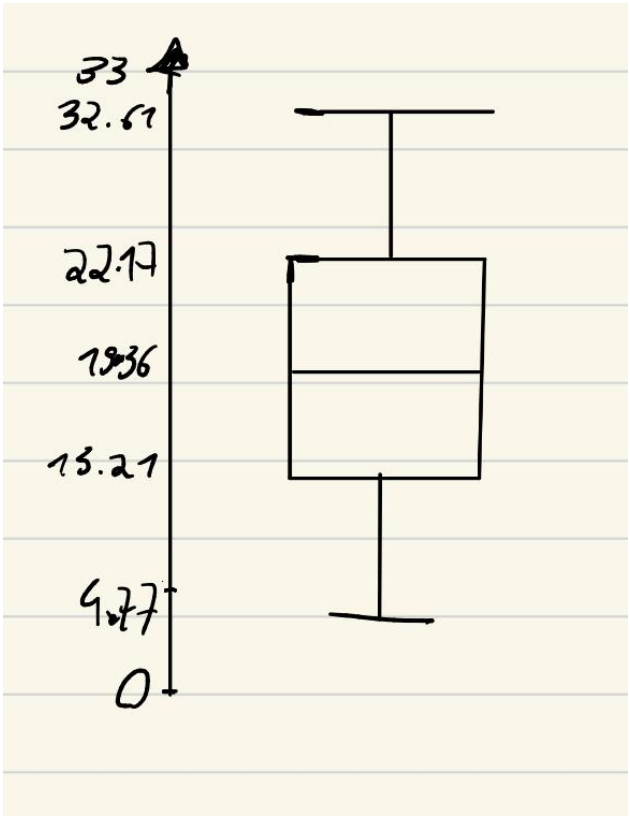

Figure 3: Boxplot drew on tablet

Comparing with R: The boxplots are similar in both approaches.

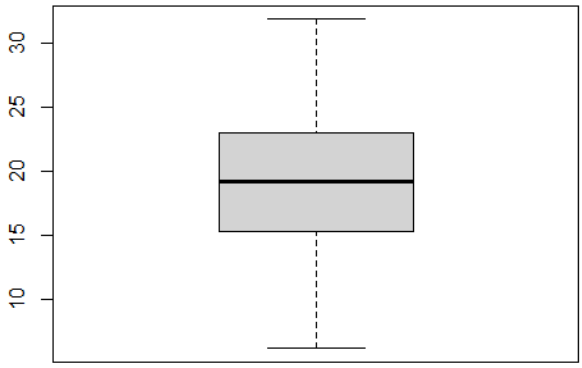

Figure 4: R's boxplot

At last, the data doesnt show any outliers, following a Gauss curve pattern.

Prompts: https://chatgpt.com/share/68ea73c1-19b4-800d-a8c2-3a6597f6407e

- 3.

- 4. Suppose the maximum acceptable daily emission limit is 25 units. What proportion of days exceeded this limit? Would the overall behavior of the emissions comply with this regulatory standard?

Arthur: Analyzing the given data, its noticeable that in 79 emissions data, 11 exceed the 25 limit, which is $(11/79).100\% = 13.92\%$ of the total.
The overall behavior of the emissions definitively comply with the limit of 25, since only 10% of the analised emissions are exceeding the limit.