Descriptive Statistics

Study of 5-Number Summary

&

Box Plot

What is the 5–Number Summary of a sample?

The **5–Point Summery** consists of the

- Minimum Value
- First Quartile Q_1
- Median $\widetilde{x}$
- Third Quartile Q_3
- Maximum Value

What are the **Quartiles** of a sample?

Quartiles are values associated with locations which divide the sample into four groups in such a way that about 25% of data belongs to each group.

What is the **MEDIAN** of a sample?

MEDIAN of a sorted sample is the number that separates the bottom 50% of the data from the top 50% of the data. The symbol often used for median is $\widetilde{x}$ and it is pronounced x–tilde.

What is the **First Quartile** of a sample?

First Quartile of a sorted sample is the number that separates the bottom 25% of the data from the top 75% of the data. The symbol often used for First Quartile is Q_1.

What is the **Third Quartile** of a sample?

Third Quartile of a sorted sample is the number that separates the bottom 75% of the data from the top 25% of the data. The symbol often used for Third Quartile is Q_3.

What about the **Second Quartile** of a sample?

Second Quartile of a sorted sample is the same as the **Median** of the sorted sample

---

How can we use **5–Number Summary** of a sample?

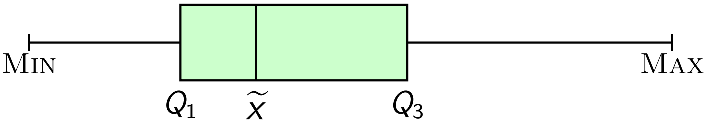
We can use the **5–Number Summary** to construct another statistical graph of a sample called **Box Plot** .

How do we construct a **Box Plot** of a sample?

We use the **5–Number Summary** to create a **BOX PLOT**.

Here is a general shape of a **BOX PLOT**, we should always scale this graph accordingly.

A Sample Boxplot



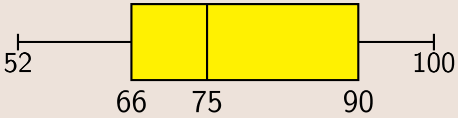
Example:

A sample of 35 exams has the 5–number summary of 52, 66, 75, 90, and 100 . Draw and label its box plot.

Solution:

We are given the **5–Number Summary** to create a **BOX PLOT**.

So we draw the general shape of a **BOX PLOT**, but we should always scale the graph accordingly.

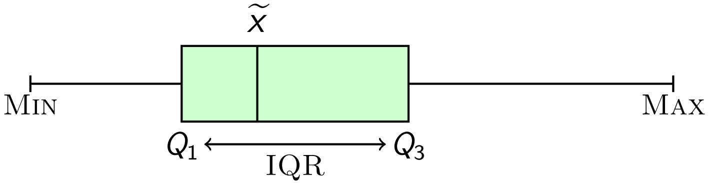


What is the **Interquartile Range** of a sample?

Interquartile Range (IQR) is the difference of Q_3 and Q_1, so

$$\text{IQR} = Q_3 - Q_1$$

A Sample Boxplot



What is the **Upper Fence** of a sample?

Upper Fence of a sorted sample is defined to be the number that can be calculated by using $Q_3 + 1.5 \bullet \text{IQR}$.

What is the **Lower Fence** of a sample?

Lower Fence of a sorted sample is defined to be the number that can be calculated by using $Q_1 - 1.5 \bullet \text{IQR}$.

Why do we need these **Fences** of a sample?

We can use these **Fences** to identify any **Outliers** .

What are the **Outliers** of a sample?

Outliers of a sorted sample are defined to be those data elements that are greater than the **Upper Fence** or smaller than the **Lower Fence** .

Example:

A sample of 35 exams has the five–point summary of $25, 66, 75, 90,$ and 100 .
Compute its **Interquartile Range**, **Upper**, and **Lower** fences.
Identify any **Outliers**.

Solution:

We are given the **5–Point Summary** , so we can use our definitions to answer these questions.

- $IQR = Q_3 - Q_1 = 90 - 66 = 24$.
- $Upper\ Fence = Q_3 + 1.5 \times IQR = 90 + 1.5 \times 24 = 126$.

Solution Continued:

- *Lower Fence* $= Q_1 - 1.5 \times IQR = 66 - 1.5 \times 24 = 30$.
- This sample has at least one outlier since the minimum value 25 is lower than the lower fence 30.

How are **Percentiles** and **Quartiles** related?

Quartiles are special cases of **Percentiles** for any sorted sample.

- $P_{25} = Q_1$
- $P_{50} = Q_2 = \widetilde{x} =$ Median
- $P_{75} = Q_3$

Example:

Consider these sorted exam scores below

58	59	60	61	63	65	67	72	75	78
79	80	82	84	85	85	86	87	88	89
90	91	92	94	95	95	96	97	99	100

Find

- Find the median.
- Find Q_1.
- Find Q_3.
- Find its IQR.
- Find its upper and lower fences.
- Draw its Box Plot.

Solution:

We use the fact that the median $= P_{50}, Q_1 = P_{25},$ and $Q_3 = P_{75}$ with $n = 30$.

- For the median $\rightarrow L = \dfrac{k}{100} \cdot n = \dfrac{50}{100} \cdot 30 = 15$

  Since the value of L is a whole number.

  Median $= \dfrac{\text{15th value} + \text{16th value}}{2} = \dfrac{85 + 85}{2} = 85.$

- For $Q_1 \rightarrow L = \dfrac{k}{100} \cdot n = \dfrac{25}{100} \cdot 30 = 7.5$

  Since the value of L is a decimal number, we round it up.

  $Q_1 = \text{8th value} = 72.$

Solution Continued:

- For $Q_3 \rightarrow L = \frac{k}{100} \cdot n = \frac{75}{100} \cdot 30 = 22.5$

  Since the value of L is a decimal number, we round it up.

  $Q_3 = 23\text{rd value} = 92.$

- $IQR = Q_3 - Q_1 = 92 - 72 = 20$.
- $Upper\ Fence = Q_3 + 1.5 \times IQR = 92 + 1.5 \times 20 = 122$.
- $Lower\ Fence = Q_1 - 1.5 \times IQR = 72 - 1.5 \times 20 = 42$.
- Here is the Box Plot.