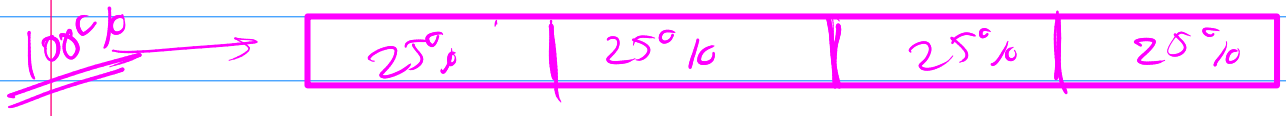
Quantile & percentile

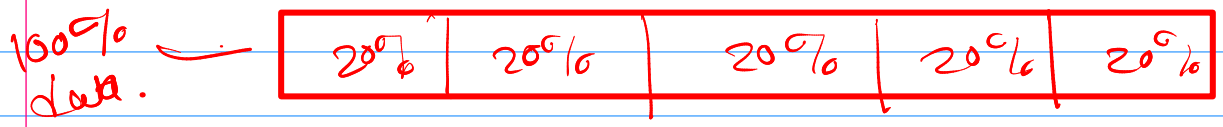
Quantile :→ Its a statistical measure used to divide a set of numerical data into equal-size group.

They can also be used to identify the outlier.

① **Quartile :→** We divide the data into four equal part.



② **Quintile :→** We divide the data into five equal part.



③ **Decile :→** We divide the data into ten equal part.

④ **Percentile :→** We divide the data into 100 equal part.

Note :→ ① The data should be sorted from low to high.

② When you calculate the Quantile then the result is the Location not a data point.

percentile formula

$$PL = \frac{P}{100}(N+1)$$ ✓

N → no. of population data.

P → Percentile Rank.

PL → Data Location.

data = [78, 82, 84, 88, 91, 93, 94, 96, 98, 99]

1 2 3 4 5 6 7 8 9 10

Calculate 75th percentile in the class.

N = 10

P = 75

$$PL = \frac{P}{100}(N+1)$$

$$= \frac{75}{100}(10+1) = \frac{3}{4}(11) = \frac{33}{4} = 8.25$$

Location

8.25 Location $= 8^{th}$ Location $+ 0.25(9^{th} - 8^{th})$

$= 96 + 0.25(98 - 96)$

$= 96 + 0.25(2)$

$= 96 + 0.5$

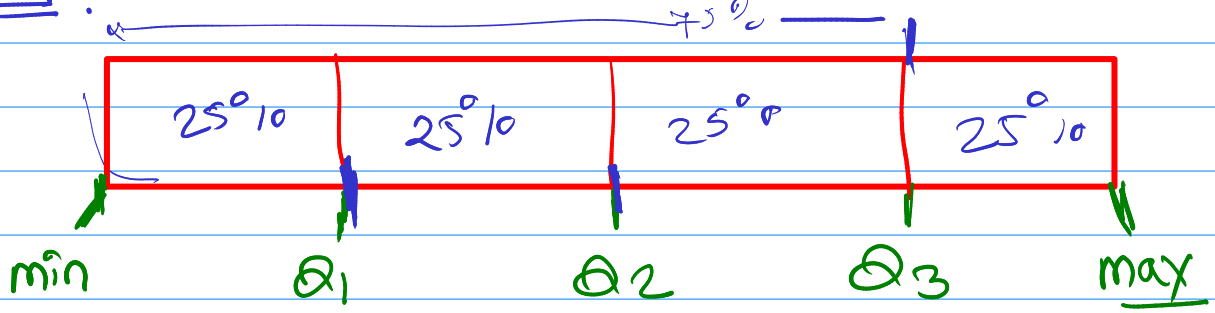
Rank $= 96.5$

96.5^{th}

To detect the outlier in the dataset.

Five Number Summary

We Calculate five numbers in the dataset by dividing the data into four equal parts.



① Minimum Value :→ The minimum Value after dividing the data into four equal part.

Lower Boundry

Note :→ This minimum Value not present/ available in the dataset.

② Q1 (First Quartile) :→ The value that seperate the lowest 25% of data.

③ Q2 (second Quartile) :- The Value that seperate the 50% of data.

- It may be median.

④ Q3 (Third Quartile) :- The value that seperate 75% of data.

⑤ max : :→ The largest Value of dataset.

Upper Boundry

Note :→ The value not present in dataset.

IQR ⇒ InterQuartile Range ⇒

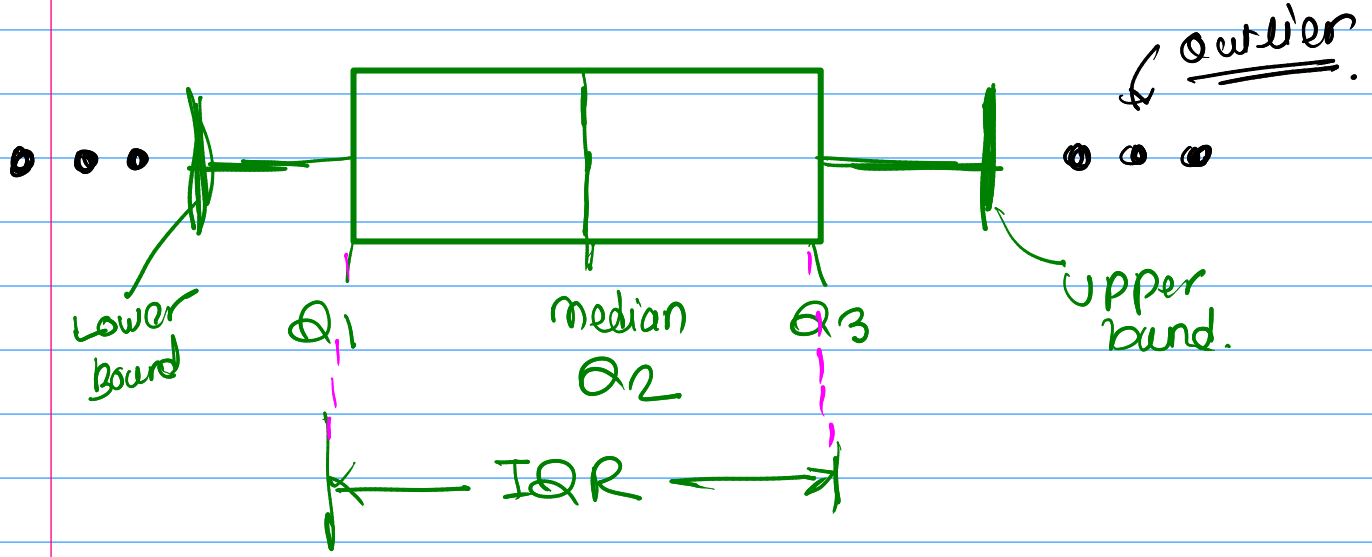
The IQR is based on five number Summary and Calculated as difference between Third Quartile (Q3) & First Quartile (Q1).

$$IQR = Q3 - Q1$$ — Formula

$$\text{Lower Bound (min)} = (Q1 - 1.5 \times IQR)$$

$$\text{Upper\_bound (max)} = Q3 + 1.5 \times IQR$$

Draw = Box plot



data = [6, 213, 241, 260, 281, 290, 314, 321, 350, 1500]
(1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

$Q_1 = \frac{P}{100}(N+1) = \frac{25}{100}(10+1) = \frac{1}{4}(11) = \frac{11}{4} = 2.75$ Location

25%

$= 213 + 0.75(241 - 213)$

$2^{nd} + 0.75(3^{rd} - 2^{nd})$

$Q_1 = 234$

$Q_2 = \frac{P}{100}(N+1) = \frac{50}{100}(11) = \frac{2}{4}(11) = \frac{22}{4} = 5.5$

50%

$= \frac{5^{th} + 6^{th}}{2} = \frac{281 + 290}{2} = \frac{285.5}{Q_2}$

$Q_3 = \frac{P}{100}(N+1) = \frac{75}{100}(11)$ - - $Q_3 = 328.25$

75%

$IQR = Q_3 - Q_1$

$= 328.25 - 234$

$IQR = 94.$

Lower $= Q_1 - 1.5 \times IQR = \boxed{93}$

Upper $= Q3 + 1.5 \times IQR = \boxed{469}$

Box plot