

SNS COLLEGE OF TECHNOLOGY

**An Autonomous Institution
Coimbatore-35**



DEPARTMENT OF ARTIFICIAL INTELLIGENCE & DATA SCIENCE

23ADT202 – FUNDAMENTALS OF DATA SCIENCE AND ANALYTICS

II YEAR IV SEM

UNIT II – Interquartile Range (IQR)

EMPATHY:

Interquartile Range (IQR)

Quartiles are three values in an ordered data set that divide the data into four equal parts, with each part containing one-fourth of the data.

Q1 is defined as the middle number between the smallest number and the median of the data set.

Q2 is the median of the data.

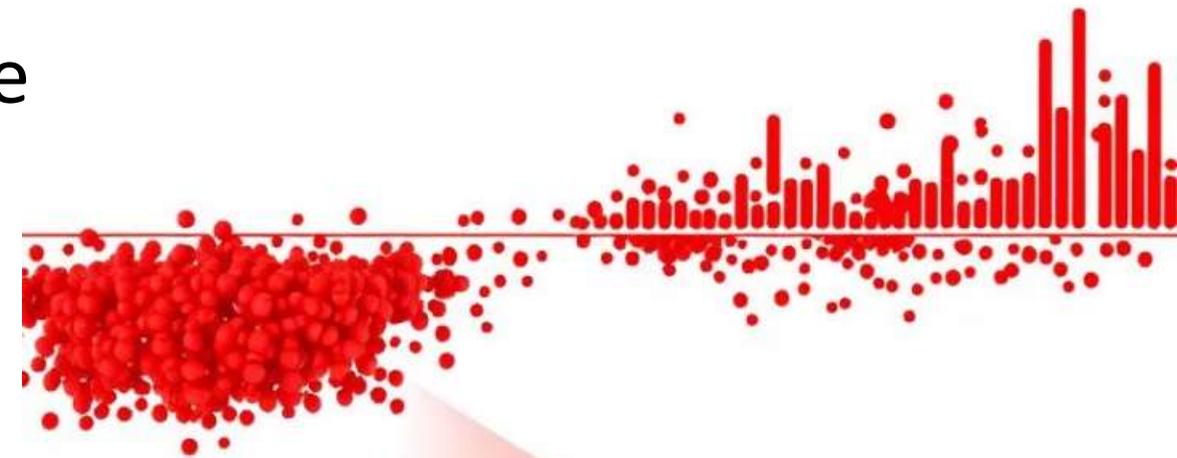
Q3 is the middle value between the median and the highest value of the data set.

Consider the data set:

10, 20, 30, 40, 50, 60, 70, 80 (already arranged in order)

- **Q1 (First Quartile):** Middle of the first half → **25**
- **Q2 (Second Quartile / Median):** Middle of the whole data → **45**
- **Q3 (Third Quartile):** Middle of the second half → **65**

These three quartiles divide the data into **four equal parts**, and each part contains **two values**.



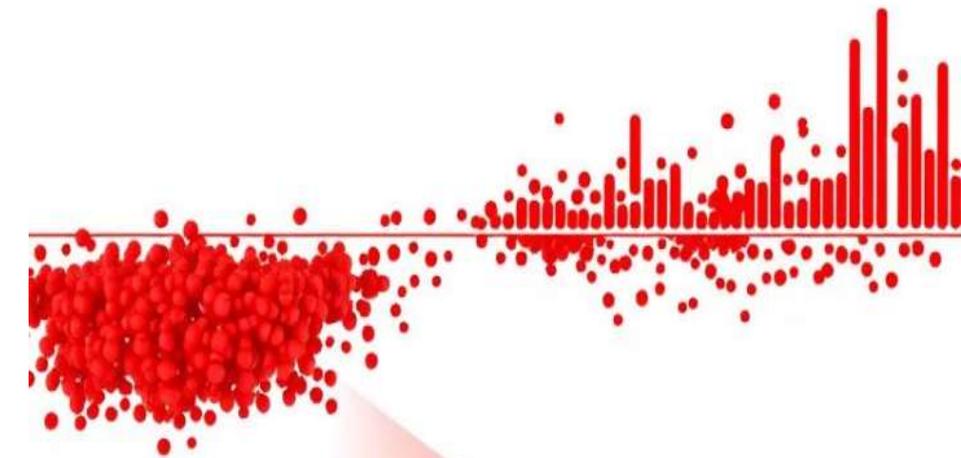
DEFINE:

- The interquartile range IQR tells us the range where the bulk of the values lie.
- The interquartile range is calculated by subtracting the first quartile from the third quartile.

$$\text{IQR} = Q3 - Q1$$
$$65 - 25 = 40$$

Uses

1. Unlike range, IQR tells where the majority of data lies and is thus preferred over range.
2. IQR can be used to identify outliers in a data set.
3. Gives the central tendency of the data.



Examples:

Input : 1, 19, 7, 6, 5, 9, 12, 27, 18, 2, 15

Output : 13

The data set after being sorted is

1, 2, 5, 6, 7, 9, 12, 15, 18, 19, 27

As mentioned above Q2 is the median of the data.

Hence $Q2 = 9$

Q1 is the median of lower half, taking Q2 as pivot.

So $Q1 = 5$

Q3 is the median of upper half talking Q2 as pivot.

So $Q3 = 18$

Therefore IQR for given data= $Q3-Q1=18-5=13$

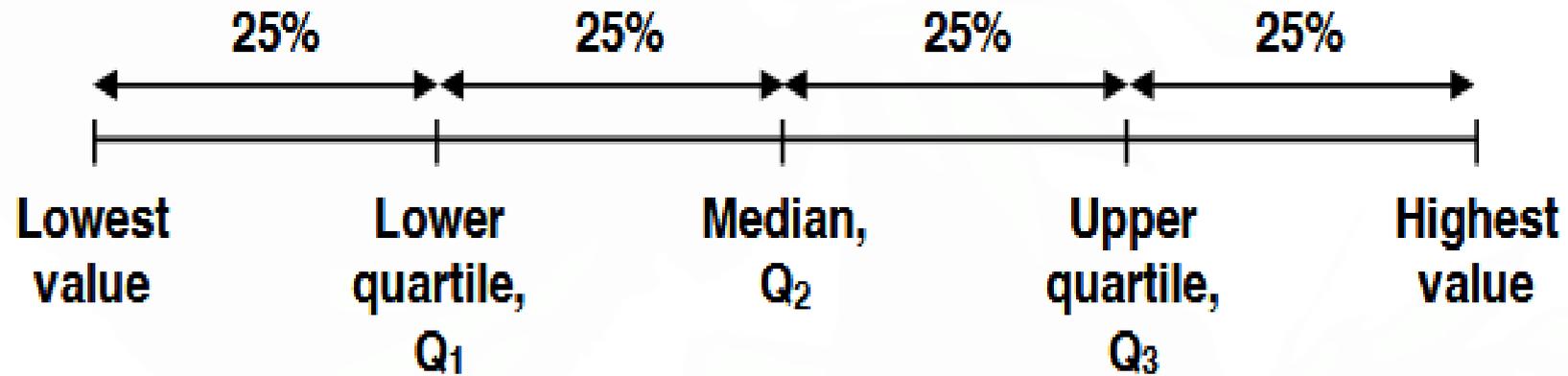
Input : 1, 3, 4, 5, 5, 6, 7, 11

Output : 13

IDEATE:

Quartiles

The upper and lower quartiles are two useful statistical values which, along with the median, help to split the data into quarters.



25% of the data are between the lowest value and the lower quartile (Q₁).

25% of the data are between the lower quartile (Q₁) and the median (Q₂) etc.

Important: 50% of the data are between the lower quartile (Q₁) and the upper quartile (Q₃).

How to find the upper and lower quartiles

$$\text{Lower quartile, } Q_1 = \frac{1}{4}(n + 1)\text{th value}$$

$$\text{Upper quartile, } Q_3 = \frac{3}{4}(n + 1)\text{th value}$$

TESTING:

- The lower quartile, Q1, is the value that is one-quarter of the way along
- The upper quartile, Q3, is the value that is three-quarters of the way along

The data must be in ascending order

Interquartile range

The *interquartile range, IQR*, is the range of the *middle 50%* of the data set.

$$\text{Interquartile range} = \text{Upper quartile} - \text{Lower quartile} = Q_3 - Q_1$$

The *interquartile range* is a useful measure when there are *extreme values* at either or both the ends.

PROTOTYPING:

PROBLEMS

1 For the data values 7, 9, 9, 13, 13, 17, 18, 20, 27, 28, 81 find:

- (a) the median
- (b) the upper and lower quartiles
- (c) the interquartile range

2. Find the median for the following data:

- (a) 7, 2, 9, 4, 8, 1, 5
- (b) 12, 11, 19, 15, 17, 14

3 Calculate the IQR for the values: 80 70 34 21 21 56 75 89 84 20 17 45 87

