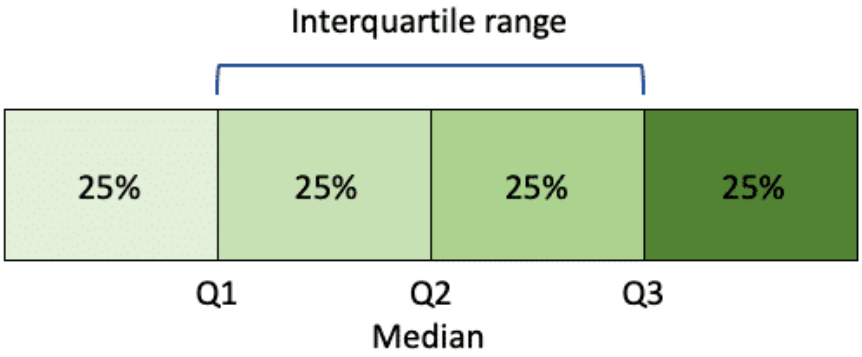
INTERQUARTILE RANGE (IQR)

The **InterQuartile Range** tells you the spread of the middle half of your distribution.

Quartiles segment any distribution that's ordered from low to high into four equal parts. The interquartile range (IQR) contains the second and third quartiles, or the middle half of your data set



Calculate the interquartile range

The interquartile range is found by subtracting the Q1 value from the Q3 value:

$$IQR = Q3 - Q1$$

- IQR = interquartile range
- Q3 = 3rd quartile or 75th percentile
- Q1 = 1st quartile or 25th percentile

Althougth There is only one formula that I have discussed already,

But there are different methods of identifying the quartiles. And everytime you use a different methods you will get the different values of interquartile Range.

The most commonly used one is median

Exclusive method vs inclusive method

The **Exclusive Method** did not include the median when identifying Q1 and Q3

But , in **Inclusive Method** did include the median in identifying the Quartiles

The procedure for finding the median is different depending on whether your data set is odd- or even-numbered.

- When you have an **odd number** of data points, the median is the value in the middle of your data set. You can choose between the inclusive and exclusive method.

- With an **even number** of data points, there are two values in the middle, so the median is their mean. It's more common to use the exclusive method in this case.

Steps for the exclusive method

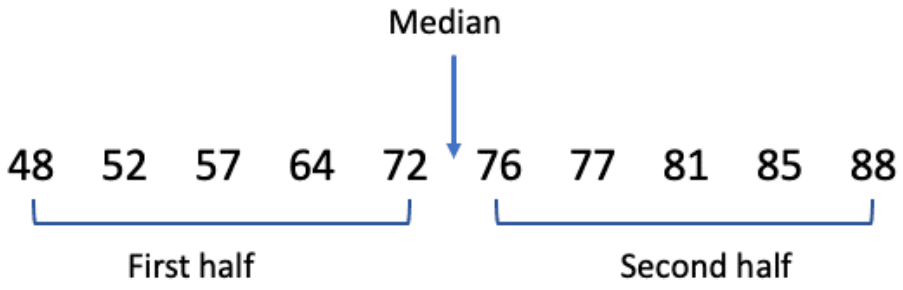
Even-numbered data set:-

Step 1: Order your values from low to high.

48 52 57 64 72 76 77 81 85 88

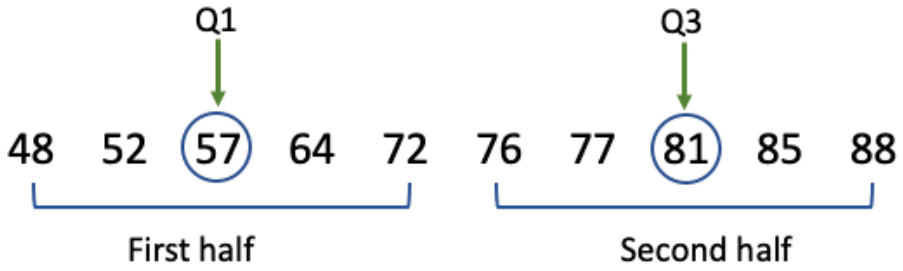
Step 2: Locate the median, and then separate the values below it from the values above it.

With an even-numbered data set, the median is the mean of the two values in the middle, so you simply divide your data set into two halves.



Step 3: Find Q1 and Q3.

Q1 is the median of the first half and Q3 is the median of the second half. Since each of these halves have an odd number of values, there is only one value in the middle of each half



Step 4: Calculate the interquartile range..

$$IQR = Q3 - Q1$$
$$IQR = 81 - 57 = 24$$

Hence same goes for the odd numbered dataset but in odd numbered data set median will be the median not the mean

But now the question is how to use it in program.

1. We will find out the Q1 and Q3 (by df["column_name"].quantile(0.25) and same goes by using for 75% as quantile as (0.75))
2. Now IQR = Q3-Q1
3. Now we will take lower object (lower = q1 - (1.5 * iqr))→ this is the formula
4. And upper object as(upper = q3 + (1.5 * iqr))→ same
5. Now all the values lower then our object name lower will be considered as outlier and all the object upper then our variable name upper will be considered as outlier
6. So now we have two option after this
    - Capping
    - Trimming

Capping:-

Think like we are having the data points 0,5,6,7,8,12

In this dataset 0 and 12 are considered as outlier so we will make 0 and 12 as 5 and 8 respectively as our lower and upper limit will be 5 and 8

Trimming:-

In this we will remove the outliers like the same dataset you consider

As we have 6 datapoints we have but after trimming it we will have 4 datapoints only

So trimming is not recommended everytime you should use

KEEP LEARNING KEEP GROWING

REGARDS

RAJ KULHADE