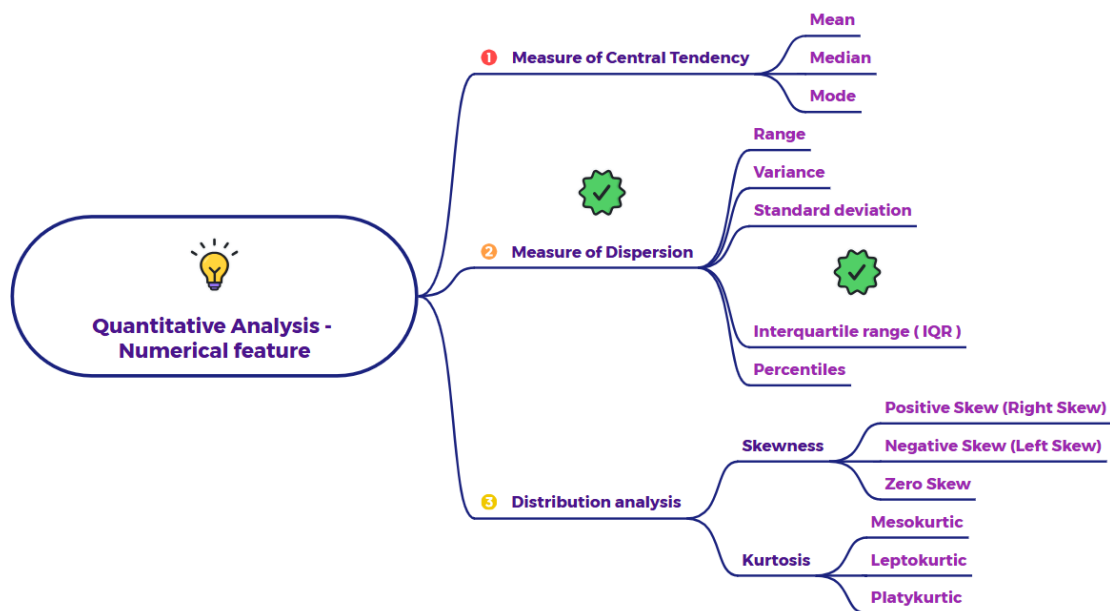# Explain measure of dispersion – Inter Quartile Range(IQR)

Concept:

The IQR is the difference between the third quartile (Q3) and the first quartile (Q1). Quartiles divide a dataset, ordered from smallest to largest, into four equal parts.

- **First Quartile (Q1):** The value below which 25% of the data falls.
- **Second Quartile (Q2):** This is the median, the value below which 50% of the data falls.
- **Third Quartile (Q3):** The value below which 75% of the data falls.
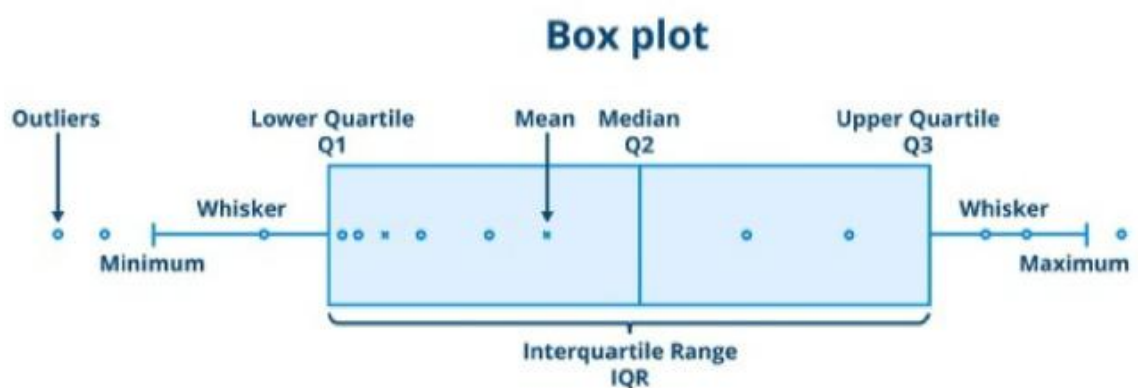
The IQR essentially tells us the range of values that the middle half of the data occupies.

Formula:

$IQR = Q3 - Q1$

Detailed Examples:

Let's use our math quiz scores again to illustrate the IQR.

Scores: 12, 15, 18, 16, 15, 10, 15, 19, 14, 15

Step 1: Order the Data

First, we arrange the scores in ascending order:

10, 12, 14, 15, 15, 15, 15, 16, 18, 19

Step 2: Find the Quartiles

With 10 data points, the quartiles are found as follows:

- **Q2 (Median):** As we found earlier, the median is the average of the 5th and 6th values: $\frac{15+15}{2} = 15$.

- **Q1:** The first quartile is the median of the lower half of the data (values below Q2). The lower half is: 10, 12, 14, 15, 15. The median of this lower half is the middle value, which is 14. So, Q1=14.

- **Q3:** The third quartile is the median of the upper half of the data (values above Q2). The upper half is: 15, 16, 18, 19. Since there are four values, the median is the average of the two middle values: $\frac{16+18}{2} = 17$. So, Q3=17.

Step 3: Calculate the IQR

$IQR = Q3 - Q1 = 17 - 14 = 3$

Interpretation:

The interquartile range for the math quiz scores is 3. This means that the middle 50% of the students' scores fall within a range of 3 points.

Example 2: Impact of an Outlier

Let's include the outlier score of 2 in our dataset:

2, 10, 12, 14, 15, 15, 15, 15, 16, 18, 19

Now we have 11 data points.

- **Q2 (Median):** The middle value is the 6th value, which is 15.

- **Q1:** The lower half (below the median) is: 2, 10, 12, 14, 15. The median of this is 12. So, Q1=12.

- **Q3:** The upper half (above the median) is: 15, 15, 16, 18, 19. The median of this is 16. So, Q3=16.

Step 3: Calculate the IQR

IQR=Q3–Q1=16–12=4

Interpretation:

With the outlier included, the IQR is 4. Notice that while the range increased dramatically (from 9 to 17) due to the outlier, the IQR only increased slightly (from 3 to 4). This highlights the robustness of the IQR to extreme values because it focuses on the central portion of the data.

Example 3: Comparing Spread with IQR

Consider two datasets of monthly salaries (in thousands of dollars):

- **Company A:** 4, 5, 5, 6, 6, 7, 7, 8

- **Company B:** 3, 4, 5, 6, 7, 8, 9, 15

Company A:

- Ordered data: 4, 5, 5, 6, 6, 7, 7, 8

- Q1 (median of 4, 5, 5, 6) = 25+5=5

- Q3 (median of 6, 7, 7, 8) = 27+7=7

- IQR-A=7–5=2

Company B:

- Ordered data: 3, 4, 5, 6, 7, 8, 9, 15

- Q1 (median of 3, 4, 5, 6) = 24+5=4.5

- Q3 (median of 7, 8, 9, 15) = 28+9=8.5

- IQR-B=8.5–4.5=4

Interpretation:

The IQR for Company A (2) is smaller than the IQR for Company B (4). This indicates that the middle 50% of salaries in Company A are more tightly clustered than the middle 50% of salaries in Company B, even though Company B has a much larger overall range due to the high outlier salary of 15.

Strengths of the IQR:

- **Robust to Outliers:** The IQR is not affected by extreme high or low values because it focuses on the central 50% of the data.

- **Good for Skewed Data:** When the data distribution is skewed, the IQR can be a more representative measure of spread than the standard deviation, which can be influenced by the long tail.

- **Easy to Understand:** The concept of the middle 50% range is relatively intuitive.

Limitations of the IQR:

- **Ignores Extremes:** While robustness to outliers is a strength, it also means that the IQR doesn't provide information about the spread of the top and bottom 25% of the data.

- **Less Information than Standard Deviation:** For normally distributed data without significant outliers, the standard deviation provides a more comprehensive measure of the overall spread as it considers all data points.

In summary, the Interquartile Range (IQR) is a valuable measure of dispersion that describes the spread of the middle 50% of the data. Its resistance to outliers makes it particularly useful for datasets with extreme values or skewed distributions, providing a more stable measure of central spread compared to the range and sometimes the standard deviation.