Day 5

Wednesday, July 24, 2024 11:11 PM

Percentiles and quartiles are statistical measures used to describe the distribution and spread of a data set.

Percentiles
Percentiles divide a data set into 100 equal parts. They are used to understand the relative standing of a value within the data set. For example, the 25th percentile (also known as the first quartile, or Q1) is the value below which 25% of the data falls. Similarly, the 50th percentile (the median, or Q2) is the value below which 50% of the data falls, and the 75th percentile (Q3) is the value below which 75% of the data falls.

In general:
- The n-th percentile of a data set is the value below which n% of the data points fall.
- Percentiles are especially useful in large data sets and for understanding the distribution and outliers.

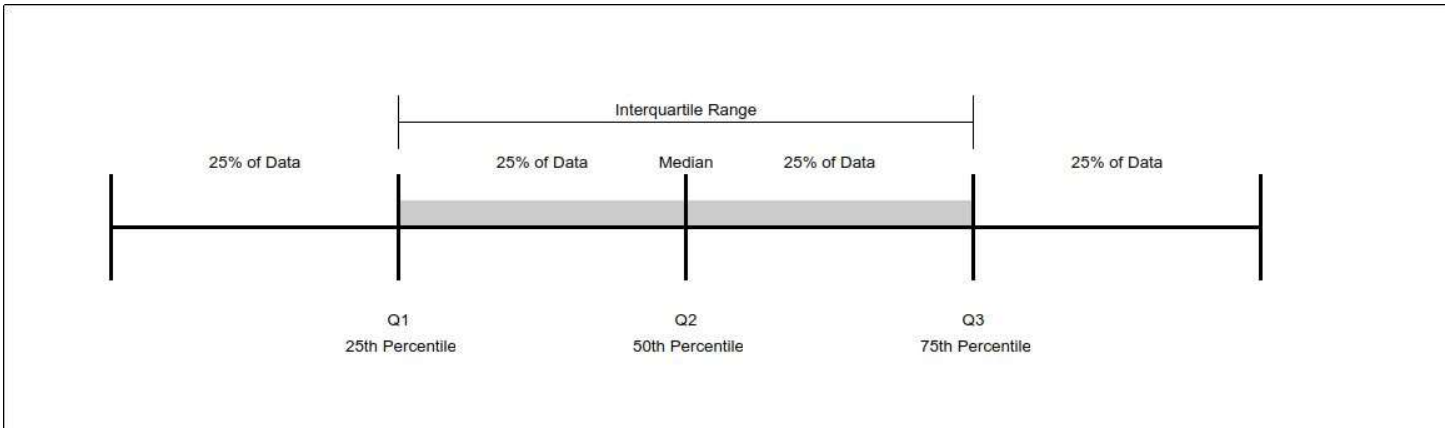
Quartiles
Quartiles are specific percentiles that divide a data set into four equal parts. They are commonly used to understand the spread and skewness of data. The three main quartiles are:
1. First Quartile (Q1): The 25th percentile. It separates the lowest 25% of the data from the rest.
2. Second Quartile (Q2): The 50th percentile or the median. It divides the data into two equal halves.
3. Third Quartile (Q3): The 75th percentile. It separates the lowest 75% of the data from the top 25%.

The Interquartile Range (IQR), defined as Q3 - Q1, measures the middle 50% of the data and is a useful indicator of the spread of the central portion of a data set.

Example: Suppose you have a data set of exam scores: 55, 62, 70, 72, 75, 80, 82, 85, 90, 95.

To calculate quartiles:
1. Q1 (25th percentile) is the median of the first half: Q1 = 70.
2. Q2 (50th percentile or median) is the overall median: Q2 = 76.5.
3. Q3 (75th percentile) is the median of the second half: Q3 = 85.

These values help summarize and interpret the data, providing insights into the distribution and central tendency.



A 5-number summary is a descriptive statistic that provides a quick overview of a data set. It consists of the following five values:

1. Minimum: The smallest value in the data set.
2. First Quartile (Q1): The value below which 25% of the data falls. This is also known as the 25th percentile.
3. Median (Q2): The middle value of the data set, dividing it into two equal parts. This is also known as the 50th percentile.
4. Third Quartile (Q3): The value below which 75% of the data falls. This is also known as the 75th percentile.
5. Maximum: The largest value in the data set.

These five values provide a concise summary of the data's distribution, indicating its center, spread, and any potential outliers.

Example

Consider the following data set: 5, 7, 8, 12, 14, 15, 21, 22, 25, 30.

1. Minimum: 5
2. First Quartile (Q1): The median of the first half of the data (excluding the overall median) is 8.
3. Median (Q2): The middle value of the sorted data is 14.5.
4. Third Quartile (Q3): The median of the second half of the data (excluding the overall median) is 22.
5. Maximum: 30

So, the 5-number summary is: 5, 8, 14.5, 22, 30.

This summary helps quickly understand the range, central tendency, and variability in the data set.