Percentiles

A Measure of Relative Standing

A percentile is a measure that indicates the relative position of a value within a dataset. It tells you the percentage of data points that fall below a given value.

Understanding Percentiles

- The p-th percentile means that p% of the data falls below that value.
- Example: If you score in the 90th percentile on a test, it means you scored higher than 90% of the other test-takers.

Percentile Formula

To find the percentile rank of a value in an ordered dataset, use:

Formula: $$\text{Percentile Rank} = \frac{\text{Number of values below } X}{\text{Total number of values}} \times 100$$

where X is the value for which you want the percentile.

Examples of Percentiles

Example 1: Test Scores

Suppose you took a test and scored 85. If 40 out of 50 students scored below you:

$$\text{Percentile Rank} = \frac{40}{50} \times 100 = 80$$

You are in the 80th percentile, meaning you performed better than 80% of students.

Example 2: Heights

If a child's height is in the 30th percentile, it means the child is taller than 30% of other children of the same age but shorter than 70%.

Percentiles vs. Percentages

- Percentile = Relative rank compared to others.
- Percentage = The proportion of correct responses or total amount.

Example:

- If you score 85% on a test → You got 85 out of 100 questions right (percentage).
- If you are in the 85th percentile → You did better than 85% of other test-takers (percentile).

Range, Quartiles, and Interquartile Range

These are measures of data dispersion, helping us understand how spread out the values in a dataset are.

Range:

The range is the difference between the largest and smallest values in a data set.

Formula: $\text{Range} = \text{Maximum Value} - \text{Minimum Value}$

Quartile:

Quartiles divide a dataset into four equal parts, where each part contains 25% of the data.

Quartile Definitions:

- Q_1 (Lower Quartile): The 25th percentile, or the middle of the lower half.
- Q_2 (Second Quartile / Median): The 50th percentile, or the middle of the entire dataset.
- Q_3 (Upper Quartile): The 75th percentile, or the middle of the upper half.

Interquartile Range (IQR):

The Interquartile Range (IQR) measures the middle 50% of the data by removing the influence of extreme values.

Formula: $\text{IQR} = Q3 - Q1$

Outliers & The IQR Rule

Outliers are extreme values in a dataset. The 1.5 × IQR Rule helps detect them.

Formula:

$$\text{Less than } Q1 - 1.5 \times \text{IQR}$$
$$\text{Greater than } Q3 + 1.5 \times \text{IQR}$$

Any values in one of these sections of the data are considered outliers.