Statistical Terms

Concept	Definition	Formula	Strengths	Weaknesses
Mean	**Average**: The sum of all values divided by the number of values.	Add up all the numbers then divide by the amount of numbers you added.	✅ Represents the overall data well in symmetric distributions. ✅ Uses all data points.	❌ Sensitive to outliers, which can distort the average. ❌ Not always representative in skewed distributions.
Median	**Middle Value**: The middle value in an ordered dataset.	If n is odd: middle value. If n is even: average of two middle values.	✅ Resistant to outliers, making it reliable for skewed distributions. ✅ Works well with ordinal data.	❌ Does not use all data values, so may lose information.
Mode	**Most Frequent Value**: The most frequently occurring value(s).	No formula; count occurrences.	✅ Useful for categorical data. ✅ Can highlight most common trends.	❌ Some datasets have no mode or multiple modes, making interpretation tricky.
Range	The difference between the highest and lowest values.	$\text{Range} = \text{Max} - \text{Min}$	✅ Easy to calculate and understand. ✅ Gives a quick sense of spread.	❌ Extremely sensitive to outliers. ❌ Does not consider data distribution.

Statistical Terms

Concept	Definition	Formula	Strengths	Weaknesses
Lower Quartile (Q1)	The 25th percentile (**median** of the lower half).	No direct formula; sort and find Q1.	✅ Helps in understanding data distribution. ✅ Used in boxplots and IQR calculations.	❌ Requires sorting the dataset first.
Upper Quartile (Q3)	The 75th percentile (**median** of the upper half).	No direct formula; sort and find Q3.	✅ Helps detect skewness and spread. ✅ Used in IQR and outlier detection.	❌ Requires sorting the dataset first.
Interquartile Range (IQR)	The range of the middle 50% of data.	$\text{IQR} = \text{Q3} - \text{Q1}$	✅ Resistant to outliers, unlike the range. ✅ Provides a better measure of spread.	❌ Does not consider extreme values, which may be important in some cases.
Percentile	The value below which a given percentage of data falls.	$\text{Rank} = \frac{\text{Number of values below } X}{\text{Total number of values}} \times 100$	✅ Used in standardized testing and benchmarking. ✅ Helps in comparing relative positions of data points.	❌ Can be misleading if dataset is small or unevenly distributed.
Outliers & IQR Rule	Identifies extreme values using the 1.5 × IQR rule.	$\text{Less than } \text{Q1} - 1.5 \times \text{IQR}$ $\text{Greater than } \text{Q3} + 1.5 \times \text{IQR}$	✅ Helps detect unusual data points. ✅ Used in data cleaning and anomaly detection.	❌ Some outliers may be valid data points, not errors. ❌ The 1.5×IQR rule is arbitrary and may not fit all datasets.