SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 107
Accredited by NAAC-UGC with 'A' Grade
Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – 23ADT202 Fundamental of Data science and Analytics

II Year / IV Semester

Unit 2 – Descriptive Analytics

Interquartile range

Introduction to IQR

The Interquartile Range (IQR) measures the spread of the middle 50% of a dataset. It helps identify the variability without being affected by extreme values. IQR is crucial for assessing consistency and identifying potential outliers. It is the difference between the third (Q3) and first quartile (Q1) of a dataset.

Why IQR is Important

Resistant to Outliers: IQR is not influenced by extreme data points.
Describes Data Spread: It highlights variability within the central portion of the data.
Outlier Detection: Provides a basis for identifying data points outside expected ranges.
IQR offers a clearer view of data trends, avoiding the skew of outliers.

Steps for Calculating IQR

1. Find the first quartile (Q1) – the median of the lower half.
2. Find the third quartile (Q3) – the median of the upper half.
3. Calculate the IQR: IQR = Q3 - Q1.

This process focuses on the middle 50% of the dataset.

Example of IQR Calculation

Given Dataset: [10, 15, 20, 25, 30, 35, 40]
Q1 = 15 (Median of the lower half)
Q3 = 35 (Median of the upper half)
IQR = 35 - 15 = 20.

Identifying Outliers Using IQR

Outliers are determined by:

- Lower Bound = Q1 - 1.5 * IQR
- Upper Bound = Q3 + 1.5 * IQR

Outliers are data points outside these bounds.

Example: Dataset [10, 15, 20, 25, 30, 35, 40, 100] shows 100 as an outlier.

Outlier Detection Example

Dataset: [10, 15, 20, 25, 30, 35, 40, 100]

Q1 = 15, Q3 = 35, IQR = 20.

Lower Bound = 15 - (1.5 * 20) = -15.

Upper Bound = 35 + (1.5 * 20) = 65.

100 is an outlier since it exceeds the upper bound of 65.

Visualizing IQR with Boxplots

Boxplots are an effective way to visualize IQR, medians, and outliers.
The box represents Q1 to Q3, with the line inside showing the median.
Whiskers extend to the bounds of the data within 1.5 * IQR.
Outliers are displayed as points outside the whiskers

Applications of IQR

1. **Outlier Detection**: Helps identify anomalies like unusually high expenses.
2. **Data Quality Assessment**: Used in evaluating consistent measurements in manufacturing.
3. **Comparative Studies**: Compares variability across different datasets, such as test scores.

IQR is vital for both understanding data and ensuring quality.

Exercise

1. **Dataset**: [5, 7, 8, 12, 13, 14, 18, 21, 23, 30]
Calculate IQR and identify any outliers.
2.Dataset: [15, 18, 19, 20, 22, 25, 30, 35, 100]
Plot a boxplot and indicate the lower and upper bounds.
These exercises will help reinforce your understanding of IQR.

THANK YOU