EDA-IQR Method

The Interquartile Range (IQR) method is a type of exploratory data analysis (EDA) technique used to summarize and understand the distribution of a dataset. Here's how it works:

What is IQR?

IQR is the difference between the third quartile (Q3) and first quartile (Q1) in a dataset. It measures the spread or dispersion of the middle 50% of the data.

How to calculate IQR:

1. Arrange your data in ascending order.
2. Find the median (Q2). This is the value that separates the lower half from the upper half of the data.
3. Find Q1, which is the median of the lower half of the data (excluding the median).
4. Find Q3, which is the median of the upper half of the data (excluding the median).
5. Calculate IQR = Q3 - Q1

Interpretation:

- A small IQR indicates a tight or symmetric distribution.
- A large IQR indicates a wide or skewed distribution.

Example:

Suppose we have a dataset of exam scores:

Score
60

To calculate IQR, we need to find Q1 and Q3:

- Find the median (Q2): 80
- Divide the data into two halves:
  - Lower half: 60, 70, 80
  - Upper half: 85, 90, 95, 98
- Find Q1: median of lower half = 75
- Find Q3: median of upper half = 92.5

Now, calculate IQR:

IQR = Q3 - Q1 = 92.5 - 75 = 17.5

Interpretation:

Since the dataset is relatively small and symmetric, we expect a moderate IQR value. The actual IQR value of 17.5 indicates that the middle 50% of scores are spread out from approximately 78 (Q1) to 96 (Q3).

In summary, the IQR method provides a useful way to understand the distribution of your data and identify potential issues with skewness or outliers.

Curated by Brajesh Kumar