INTER QUARTILE RANGE(IQR)

* **EXPLANATION:**

The **Interquartile Range (IQR)** is a statistical measure used to assess the spread or variability of the middle 50% of a dataset. It is calculated as the difference between the 75th percentile (Q3) and the 25th percentile (Q1). Here’s the breakdown of IQR for the given data

IQR=Q3−Q1

The **1.5 Rule** (also known as the **1.5 \* IQR rule**) is a commonly used method for identifying **outliers** in a dataset. The rule helps determine which data points fall outside the "normal" range, based on the interquartile range (IQR).

**Formula for 1.5 Rule:**

1. **Lower Bound**: Lower Bound=Q1−1.5×IQR\text{Lower Bound} = Q1 - 1.5 \times \text{IQR}Lower Bound=Q1−1.5×IQR
2. **Upper Bound**: Upper Bound=Q3+1.5×IQR\text{Upper Bound} = Q3 + 1.5 \times \text{IQR}Upper Bound=Q3+1.5×IQR

Where:

* **Q1** is the 25th percentile (the first quartile),
* **Q3** is the 75th percentile (the third quartile),
* **IQR** is the **Interquartile Range**, which is calculated as Q3−Q1Q3 - Q1Q3−Q1.
* **How it works:**
* Any data point **below the Lower Bound** or **above the Upper Bound** is considered an **outlier**.

**Example from given dataset:**

Let’s take the variable **ssc\_p (Secondary School Percentage)** as an example to apply the 1.5 rule:

1. **Q1 (25th percentile)** = 54.5
2. **Q3 (75th percentile)** = 161.5
3. **IQR** = 107.0 (as calculated earlier)

**Step 1: Calculate the Lower and Upper Bounds.**

* **Lower Bound**: 54.5−(1.5×107.0)=54.5−160.5=−106.054.5 - (1.5 \times 107.0) = 54.5 - 160.5 = -106.054.5−(1.5×107.0)=54.5−160.5=−106.0
* **Upper Bound**: 161.5+(1.5×107.0)=161.5+160.5=322.0161.5 + (1.5 \times 107.0) = 161.5 + 160.5 = 322.0161.5+(1.5×107.0)=161.5+160.5=322.0

**Step 2: Identify Outliers.**

* Any data point **less than -106.0** or **greater than 322.0** is an outlier.

Given that the data for **ssc\_p** ranges from 1 to 215, there are **no outliers** for this variable based on the 1.5 rule (because all values fall between -106.0 and 322.0).

* **Summary of the 1.5 Rule:**
* It’s used to flag potential outliers by calculating the lower and upper bounds based on **IQR**.
* Data points falling outside these bounds are considered outliers.
* The **1.5 multiplier** is a standard threshold, but it can be adjusted for stricter or more lenient outlier detection.

This rule helps in understanding where the "normal" data points lie and where extreme or unusual values occur.

* **Overall Analysis Report:**



* **General Conclusion from IQR:**
* **Higher IQRs** (like for **ssc\_p** and **salary**) indicate **greater variability** in those fields, showing that there’s a larger range between the 25th and 75th percentiles.
* **Lower IQRs** (like for **hsc\_p**, **degree\_p**, and **etest\_p**) show **less variation** and a more clustered distribution around the median.
* **In Summary:**

The IQR analysis suggests that certain variables like **ssc\_p** and **salary** have a broader spread, indicating more diversity in those areas. In contrast, **hsc\_p**, **degree\_p**, and **etest\_p** are more consistent, with less spread between the 25th and 75th percentiles.