

The **Interquartile Range (IQR)** is a measure of statistical dispersion that represents the range between the first quartile (Q1) and the third quartile (Q3) in a dataset. It's calculated as:

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

One common method for identifying outliers in a dataset is the use of the "1.5*IQR rule." According to this rule, data points that lie more than 1.5 times the IQR above the third quartile or below the first quartile are considered outliers.

Here's why the factor of 1.5 is commonly used:

1. **Robustness:** The 1.5 multiplier is chosen to balance sensitivity and robustness. It's a relatively conservative threshold that identifies outliers without being overly sensitive to slight variations in the data. This factor has been empirically found to work well in many situations across different types of datasets.
2. **Historical Convention:** The use of 1.5 as the multiplier for identifying outliers is a convention that has been widely adopted in statistical practice and data analysis. While there isn't a strict mathematical rationale behind why 1.5 is chosen over other values, it has become a standard due to its effectiveness in practice.
3. **Theoretical Considerations:** While the 1.5 multiplier is commonly used, it's not a fixed rule applicable to all situations. In some cases, a different multiplier might be more appropriate depending on the specific characteristics of the data or the requirements of the analysis. However, 1.5 is a good starting point that strikes a balance between sensitivity and robustness.

Overall, the 1.5*IQR rule provides a simple and practical guideline for identifying outliers in a dataset, although it's important to note that it's not a definitive rule and should be used in conjunction with other exploratory data analysis techniques and domain knowledge.