IQR – INTER QUARTILE RANGE

What is outlier?

An outlier is the data point that differs significantly from the other observations in the dataset.

Types:

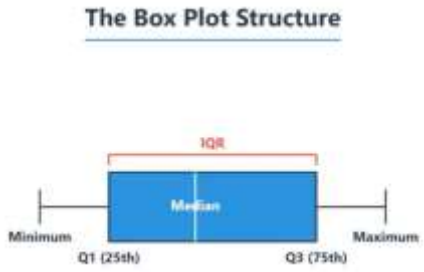
1. Lesser outlier
2. Greater outlier

What is IQR?

IQR = Q3 – Q1

IQR represents the middle 50% of your data

It is measure of statistical dispersion, representing the spread of data



Formula to find outliers:

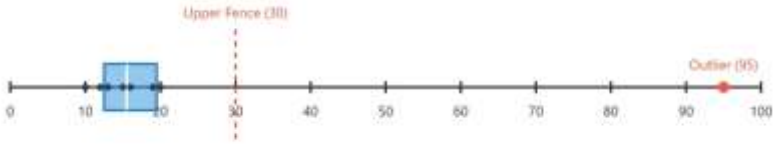
Lesser outlier should be less than -> Q1 – 1.5 * IQR

Greater outlier should be greater than -> Q3 + 1.5 * IQR

Why 1.5?

the threshold value of **1.5 in the IQR method** was determined by statisticians as a balanced approach for identifying outliers. When 1.5 is multiplied by the IQR, it helps capture most of the data points within the normal range, typically covering nearly all expected values in a normally distributed dataset. Values that fall outside this range are considered potential outliers.

Statisticians arrived at this value by comparing IQR-based boundaries with standard deviation-based methods and by testing different multipliers. When IQR is multiplied by 1, the rule becomes too strict and may classify normal values as outliers. When IQR is multiplied by 2 or 3, the rule becomes too lenient and may fail to detect actual outliers. Therefore, multiplying the IQR by 1.5 provides a balanced threshold, which is widely accepted as a standard rule for outlier detection.



Reference:

https://builtin.com/articles/1-5-iqr-rule#:~:text=The%20interquartile%20(IQR)%20method%20of,bound%20quartile%20is%20an%20outlier.