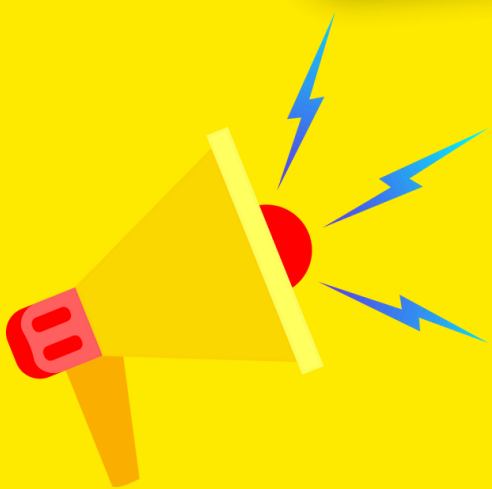
DID YOU KNOW WHY?

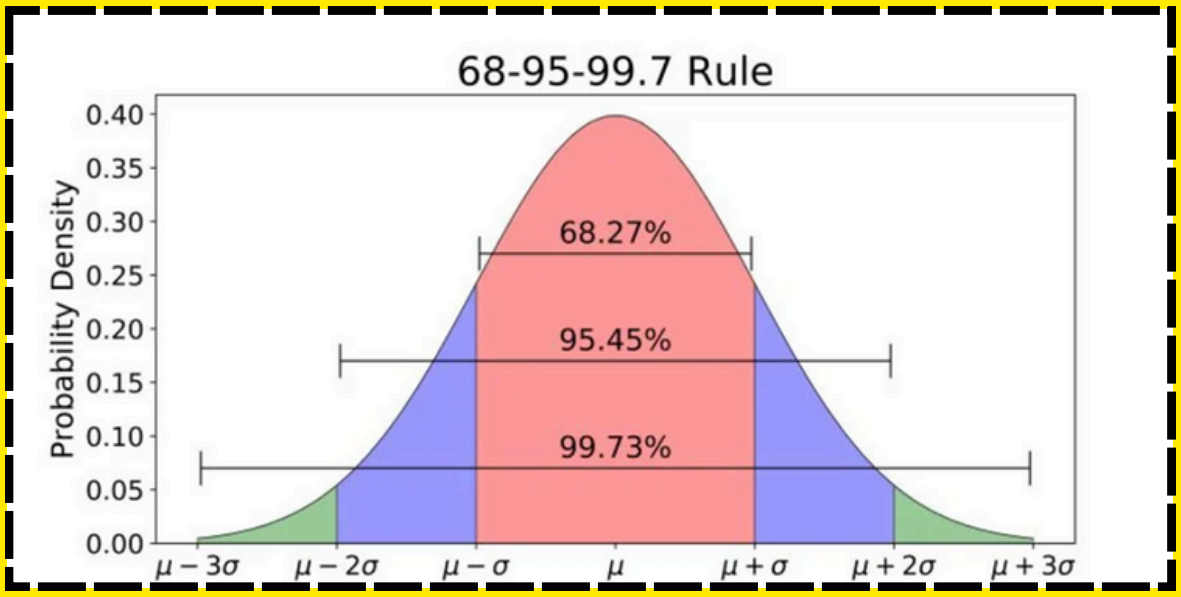
1.5 Rule is used with IQR to find outliers



Statistical Reasoning

The empirical rule(1.5 rule) is derived from the Gaussian Distribution(refer below image)

- **68%** of data lies within one Standard Deviation(<σ) of the mean (μ).
- **95.4%** of data lies within two Standard Deviation (<2σ) of the mean (μ).
- **99.7%** of data lies within three Standard Deviation (<3σ) of the mean (μ).
- The rest of the data, which lies outside the three standard deviations of the mean (μ), is considered outliers.



Different scales with IQR

scale = 1

with IQR

Lower bound
= Q1 - 1 * IQR
= Q1 - 1 * (Q3-Q1)
= **-2.025σ**
Upper bound
= Q3 + 1 * IQR
= Q3 + 1 * (Q3 - Q1)
= **2.025σ**

❌

scale= 1.5

Lower Bound
= Q1 - 1.5 * IQR
= Q1 - 1.5 * (Q3 - Q1)
= **-2.7σ**
Upper Bound
= Q3 + 1.5 * IQR
= Q3 + 1.5 * (Q3 - Q1)
= **2.7σ**

✅

scale = 2

Lower Bound
= Q1 - 2 * IQR
= Q1 - 2 * (Q3 - Q1)
= -3.375σ
Upper Bound
= Q3 + 2 * IQR
= Q3 + 2 * (Q3 - Q1)
= 3.375σ

❌

Whereas Q1= 0.675σ, Q3 =-0.675σ

Scale = 1:

- When scale is taken as **1,** any data that lies beyond **2.025σ** from the mean(**μ**) on either side shall be considered as an outlier
- However, we cannot use scale=1, as this makes the decision range **so small** (compared to 3σ) that it considers some data points as outliers, which is incorrect.

Scale = 2:

- When scale is taken as **2,** any data that lies beyond **3.375σ** from the mean (**μ**), on either side, shall be considered an outlier.
- However, we cannot use scale=2, as this makes the decision range **too large** (compared to 3σ) to consider some outliers as data points, which is undesirable.

Scale = 1.5:

- When scale is taken as **1.5**, any data that lies beyond **2.7σ** from the mean **(μ)**, on either side, shall be considered an outlier**.**
- Hence, we are using **scale=1.5** as this makes the decision range closest to 3σ
- ***It is neither as small as 1 nor as large as 2, so it's ideal to use 1.5 with IQR to detect outliers***

LIKE

&

FOLLOW

for more