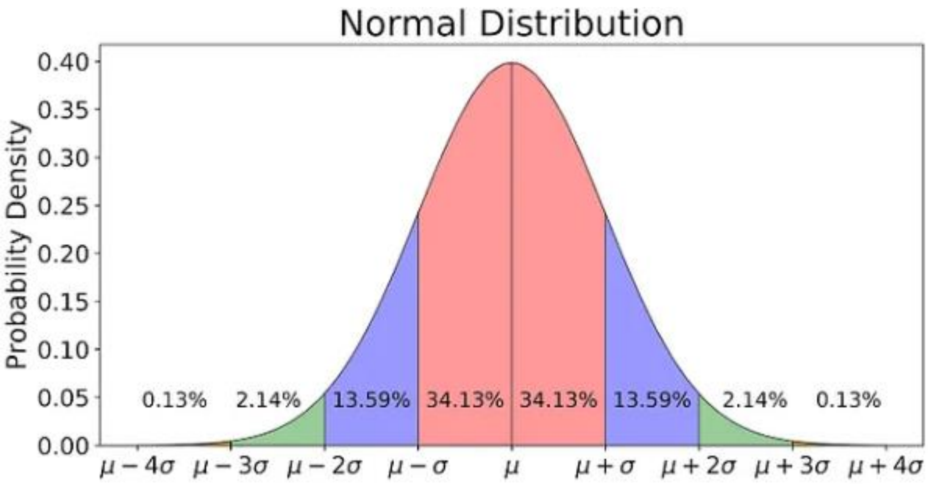
Why do we use 1.5 scale to find outliers



According to the given figure,

About 68% of the whole data lies within one standard deviation (**<σ**) of the mean (μ), taking both sides into account, the pink region in the figure.

About 95% of the whole data lies within two standard deviations (**2σ**) of the mean (μ), taking both sides into account, the pink+blue region in the figure.

About 99% of the whole data lies within three standard deviations (**<3σ**) of the mean (μ), taking both sides into account, the pink+blue+green region in the figure.

And the rest 0.20% of the whole data lies outside three standard deviations (**>3σ**) of the mean (μ), taking both sides into account, the little red region in the figure. **And this part of the data is considered as outliers**.

Calculation of IQR based on σ :

Consider Scale = 1

Lower Bound = Q1 - 1 * IQR

= Q1 - 1 * (Q3 - Q1)

= -0.675σ - 1 * (0.675 - [-0.675]) σ

= -0.675σ - 1 * 1.35σ

= **-2.025σ**

Upper Bound = Q3 + 1 * IQR

= Q3 + 1 * (Q3 - Q1)

= 0.675σ + 1 * (0.675 - [-0.675])σ

= 0.675σ + 1 * 1.35σ

= **2.025σ**

Conclusion:

If the scale is taken as 1, then according to IQR any data lies beyond 2.025 is outlier, but as we know upto 3standard deviation, the data is useful.

Hence we cannot consider scale as 1.

Consider Scale = 2

Lower Bound = Q1 - 2 * IQR

= Q1 - 2 * (Q3 - Q1)

= -0.675σ - 2 * (0.675 - [-0.675]) σ

= -0.675σ - 2 * 1.35σ

= **-3.375σ**

Upper Bound = Q3 + 2 * IQR

= Q3 + 2 * (Q3 - Q1)

= 0.675σ + 2 * (0.675 - [-0.675])σ

= 0.675σ + 2 * 1.35σ

= **3.375σ**

Conclusion

If the scale is 2, then according to IQR method any data lies beyond 3.375 is outlier, but as we know those data are useful. we cannot consider scale as 2.

Consider Scale = 1.5

Lower Bound $= Q1 - 1.5 * IQR$

$= Q1 - 1.5 * (Q3 - Q1)$

$= -0.675\sigma - 1.5 * (0.675 - [-0.675])\ \sigma$

$= -0.675\sigma - 1.5 * 1.35\sigma$

$= \mathbf{-2.7\sigma}$

Upper Bound $= Q3 + 1.5 * IQR$

$= Q3 + 1.5 * (Q3 - Q1)$

$= 0.675\sigma + 1.5 * (0.675 - [-0.675])\sigma$

$= 0.675\sigma + 1.5 * 1.35\sigma$

$= \mathbf{2.7\sigma}$

When we considering scale as 1.5, according to IQR data lies beyond 2.7 is outlier, this shows the decision range closet to Guassian distribution considers for outlier detection.

Hence we conclude the scale 1.5 will be optimum for detecting outliers.