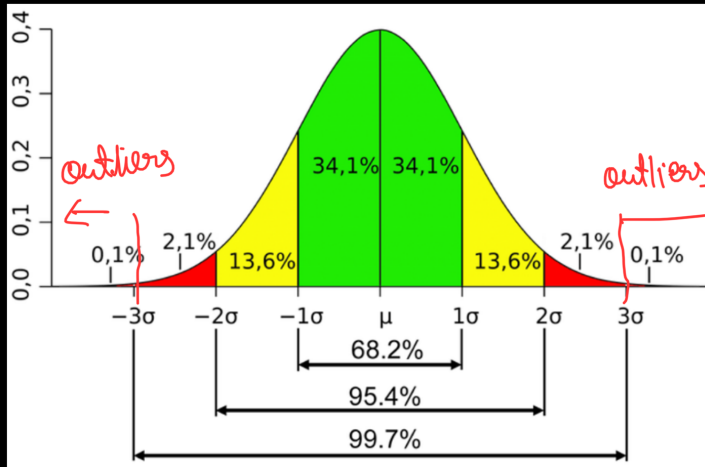


# Outliers Detection and Removal using Z-Score Method

Assumption: Column should be normally distributed.



$$\left\{ \begin{matrix} \mu + \sigma \\ \mu - \sigma \end{matrix} \right\} 68.1\%$$

$$\left\{ \begin{matrix} \mu + 2\sigma \\ \mu - 2\sigma \end{matrix} \right\} 95.4\%$$

$$\left\{ \begin{matrix} \mu + 3\sigma \\ \mu - 3\sigma \end{matrix} \right\} 99.7\%$$

Z:

Z score : 
$$Z_i = \frac{x_i - \mu}{\sigma}$$

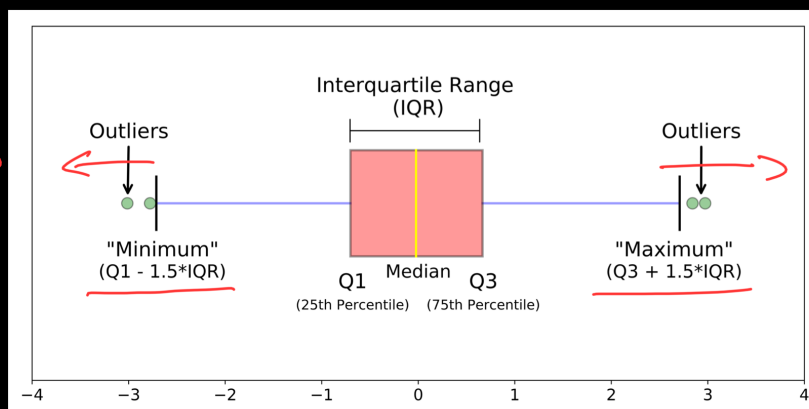
Outlier treatment

Trimming

Capping

→ replace outliers with some upper limit and lower limit.

# Outliers Detection and Removal using the IQR Method



→ Box Plot

Outliers

Outliers

IQR Proximity Rule

# Outliers Detection using the Percentile Method

| Winsorization technique

In this method we choose our upper limit and lower limit in form of percentile such as

upper limit : 95 percentile

lower limit : 5 percentile then

both limit should always be symmetric means if you are ignoring 5% in upper then in lower you will take 5%.

any value which is less than lower limit and greater than upper limit is outliers.

Generally people take :  
→ 99% (upper limit)  
→ 1% (lower limit)

## Outliers treatment

Trimming

Capping

(In this case capping is also known as winsorization)

