

# Outliers

How to calculate outlier

## Z-score method

Suppose you have data set and it's mean is 'u' and standard deviation is 'sigma'

Convert this data set into Standard normal dataset which has mean = 0 and 'sigma' as 1

$$Z = \frac{(x - u)}{\text{sigma}}$$

If and dataset falls:

    Z <= 3, Not and outlier  
else  
    Z > 3, Outlier

## IQR method

Calculate the 25th %ile and 75th %ile of your data set

IQR = 75th %ile - 25th %ile

Q3

Q1

Lower Bound = Q1 - 1.5 \* IQR

Upper Bound = Q3 + 1.5 \* IQR

Anything that lie outside Lower bound and Upper bound is an outlier

Q1, Q3 = np.percentile( dataset, [25, 75] )