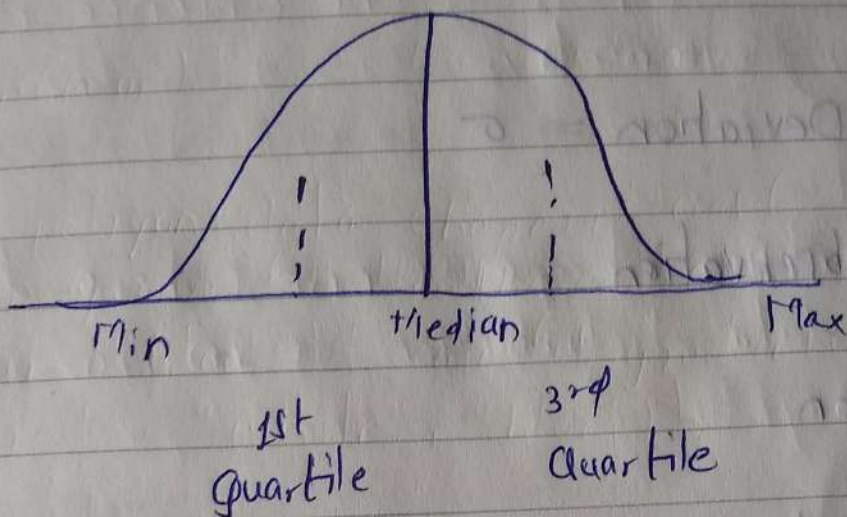# 4. Descriptive statistics

What is statistics ??

→ The collection, classification, analysis and interpretation of numeric data (Descriptive / Summary statistics)

→ The use of probability theory to impose order on aggregates of data (inferential statistics)

→ In general, statistics summarizes information about data in meaningful, relevant way.

Range

Min   Median   Max

1st   3rd
Quartile   Quartile

Range = Max - Min

Interquartile Range = 3rd quartile (Q_3) - 1st Quartile (Q_1)

Variance $= \sigma^2 = \sum (x - \mu)^2 / N$

Standard Deviation $= \sigma$

$X =$ Observation

$\mu =$ Mean

$N =$ No of observations

CHEBYSHEV's Inequality

→ Atleast 75% of all data points will lie within 2 std deviations of the mean

→ Atleast 89% of all data points will lie within 3 std deviations of the mean

Different Types of Summary Statistics

1) Measures of Central Tendency

   Mean, median, Mode

2) Measures of Variations
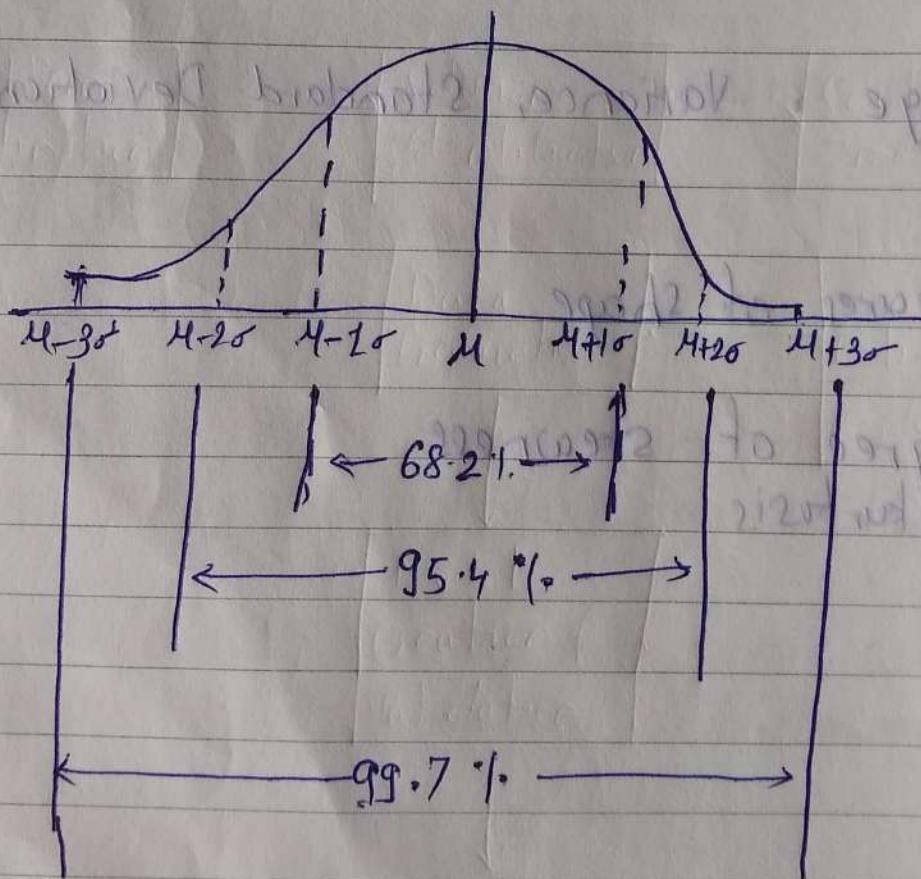
   Range, Variance, Standard Deviation

3) Measures of shape

   degree of skewness
   kurtosis

Skewness:-

Skewness is the absense of symmetry

+ A symetric shape is one where left side of data distribution is mirror image of the right

$$M-3\sigma \quad M-2\sigma \quad M-1\sigma \quad M \quad M+1\sigma \quad M+2\sigma \quad M+3\sigma$$

← 68·2% →

← 95·4% →

← 99·7% →

→ 68·2% of Data lies 1 Standard deviation (SD) away from the mean, 95·4% between 2 SD & 99·7% of data lies between 3 SD.

skewness $= \dfrac{\sum_{i=1}^{N} (Y_i - \bar{Y})^3}{(N-1) s^3}$

where Y_i = value of ith variable
$\bar{Y}$ = sample mean, s = sample SD

* Kurtosis :
   Sharpness of the peak of the distribution.

→ High kurtosis distribution has sharp peak and long tails

→ Low kurtosis distribution has flat peak and thin tails

kurtosis $= \dfrac{\sum_{i=1}^{N} (Y_i - \bar{Y})^4}{(N-1) s^4}$

→ $s \Rightarrow$ std deviation