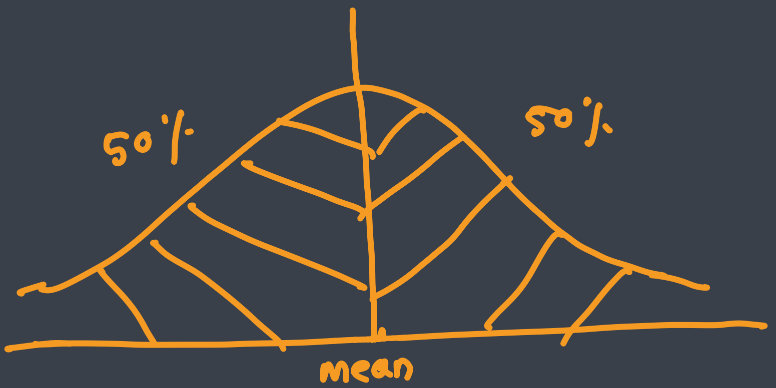
Skewness

Kurtosis

Measures of Asymmetry



Symmetrical

Need to Measures of Asymmetry

- Two distributions may have same mean and standard deviation but may differ widely in their overall appearance

D1

X	Y
0-5	10
5-10	30
10-15	60
15-20	60
20-25	30
25-30	10

mean = 15

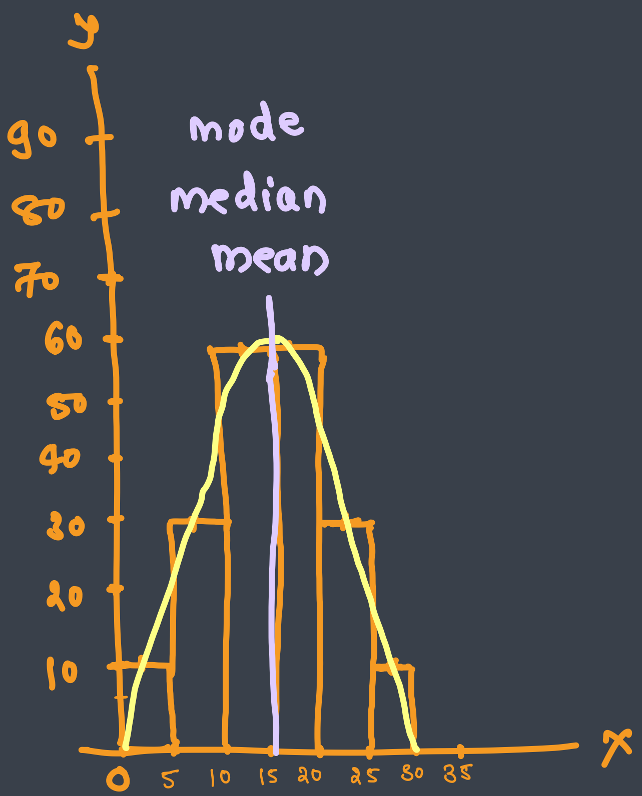
SD (σ) = 6.02

D2

X	Y
0-5	10
5-10	40
10-15	30
15-20	90
20-25	20
25-30	10

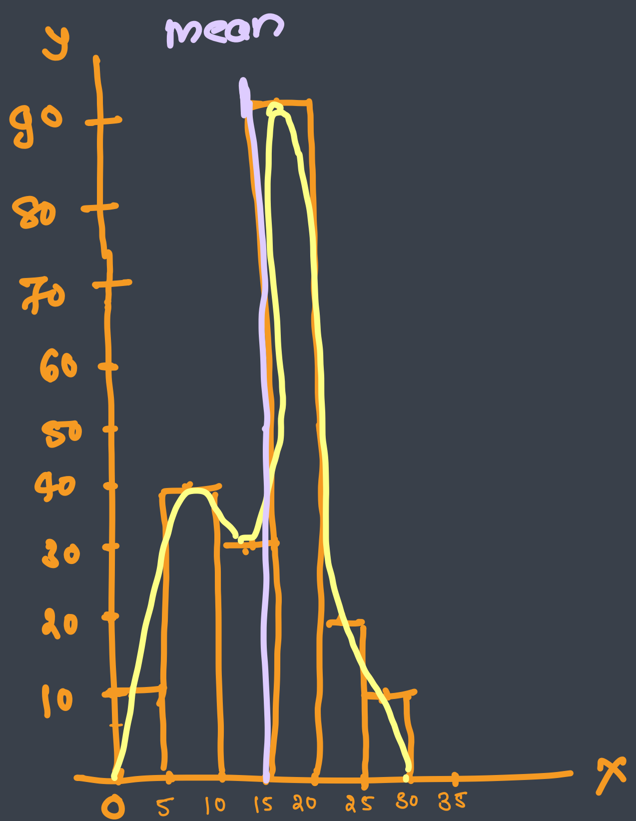
mean = 15

SD (σ) = 6.02



D1

Symmetrical

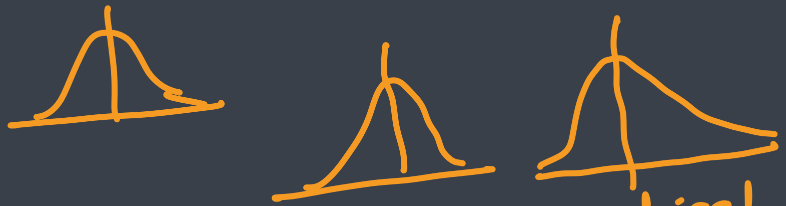


D2

assymmetrical / skewed

Symmetrical distribution

- When series is not symmetrical, it is said to be asymmetrical or skewed
- Skewness refers to the asymmetry or lack of symmetry in the shape of frequency distribution
- Measure of skewness tell us the direction and the extent of skewness
- In symmetrical distribution the mean, mode and median are identical where are in asymmetrical distribution they are not
- Distribution types
  - Symmetrical distribution
  - Asymmetrical distribution
    - Positively skewed distribution
    - Negatively skewed distribution



assymmetrical

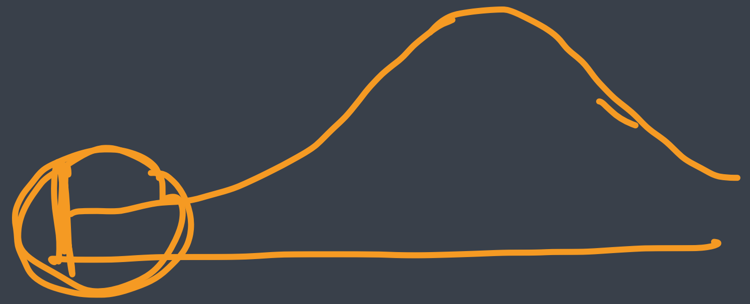
Skewness

- Skewness measures the deviation of a random variable's given distribution from the normal distribution, which is symmetrical on both sides.
- A given distribution can be either be skewed to the left or the right. Skewness risk occurs when a symmetric distribution is applied to the skewed data.
- Investors take note of skewness while assessing investments' return distribution since extreme data points are also considered.

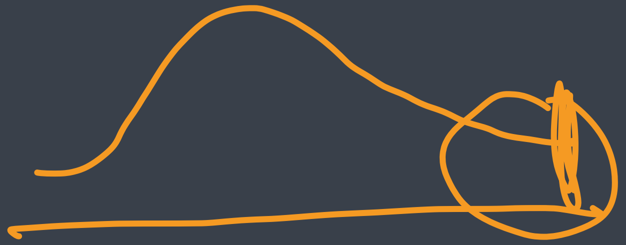
symmetrical distribution

Normal distribution

bell shaped curve



Skew



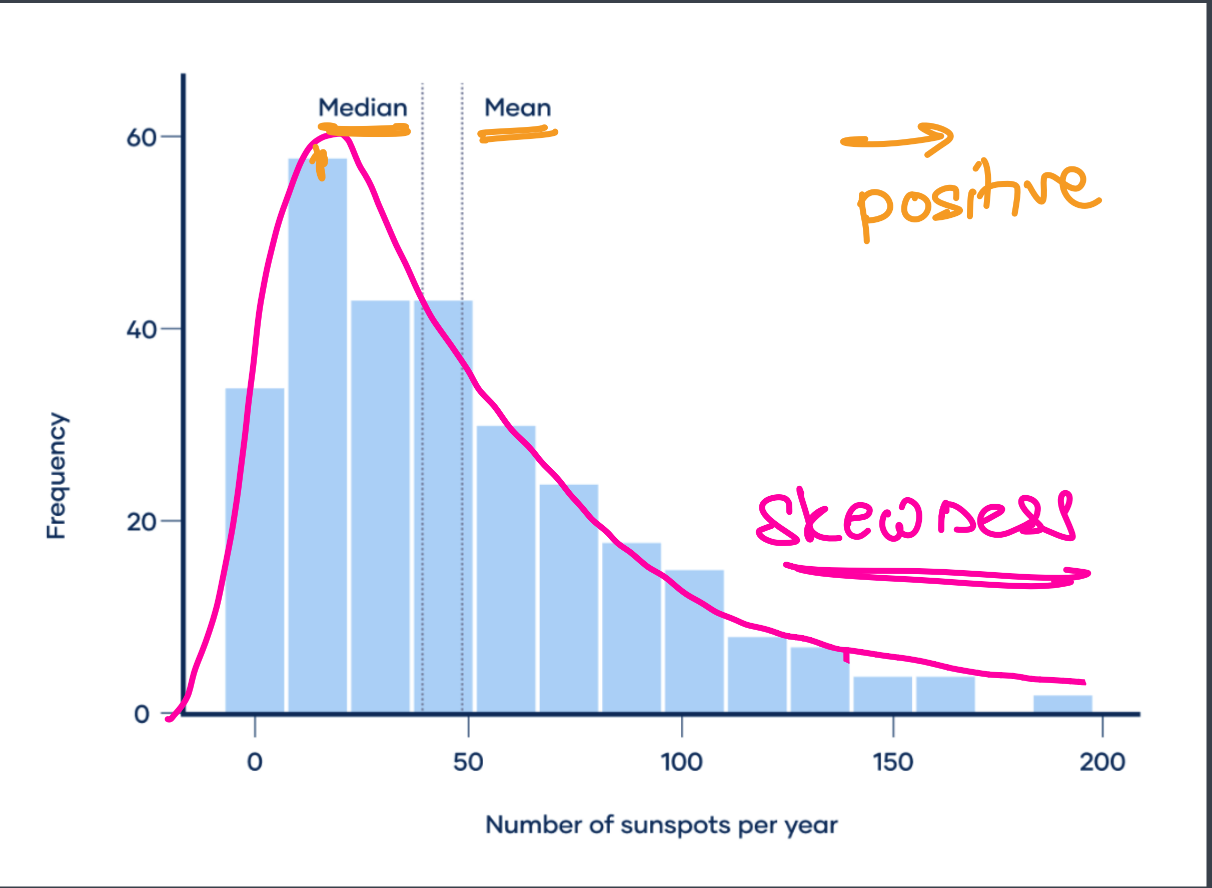
Skew

Positive Skewness → Right skewness

mean - median = +ve

mean > median

- If the given distribution is shifted to the left and with its tail on the right side, it is a positively skewed distribution
- It is also called the right-skewed distribution
- A tail is referred to as the tapering of the curve differently from the data points on the other side
- As the name suggests, a positively skewed distribution assumes a skewness value of more than zero
- Since the skewness of the given distribution is on the right, the mean value is greater than the median and moves towards the right, and the mode occurs at the highest frequency of the distribution

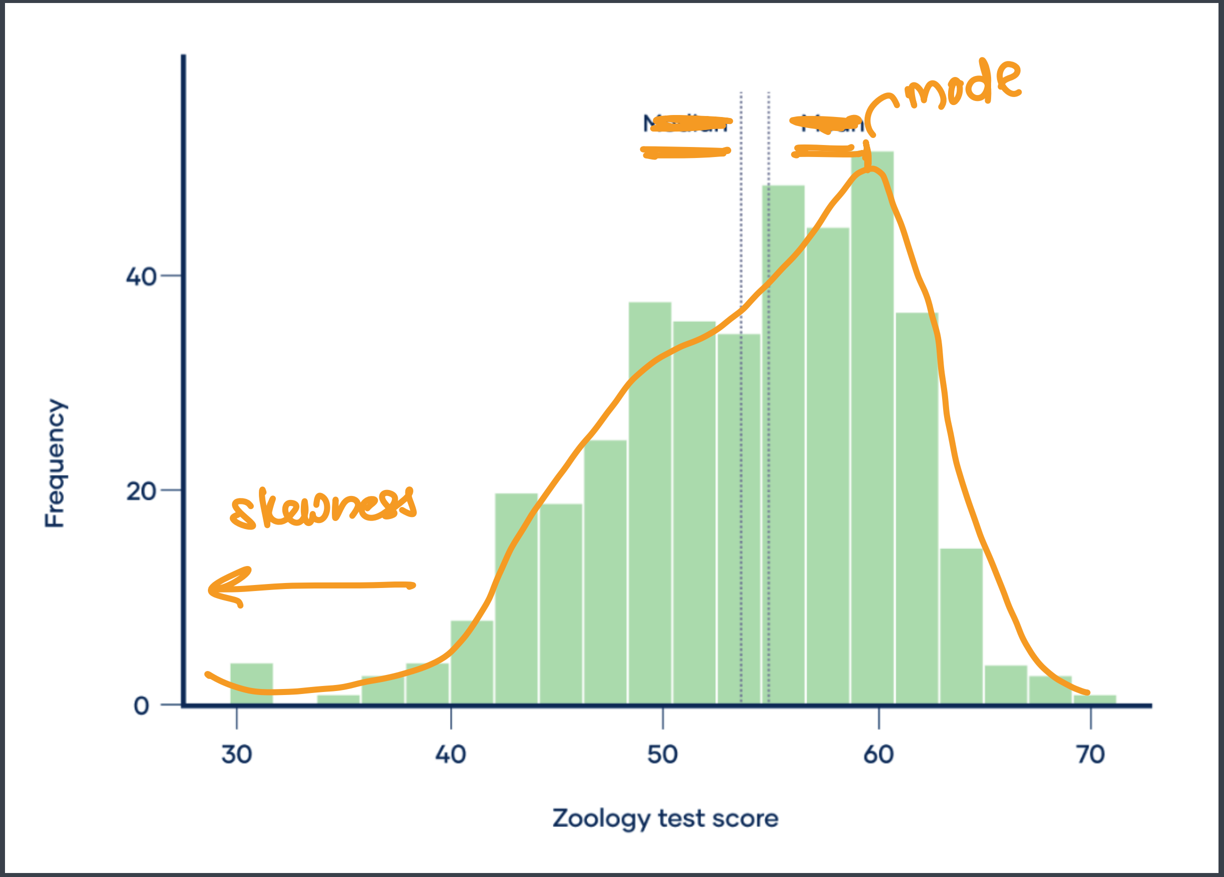


Negative Skewness → left skewed

mean − median = −ve value

mean < median

- If the given distribution is shifted to the right and with its tail on the left side, it is a negatively skewed distribution
- It is also called a left-skewed distribution
- The skewness value of any distribution showing a negative skew is always less than zero
- The skewness of the given distribution is on the left; hence, the mean value is less than the median and moves towards the left, and the mode occurs at the highest frequency of the distribution

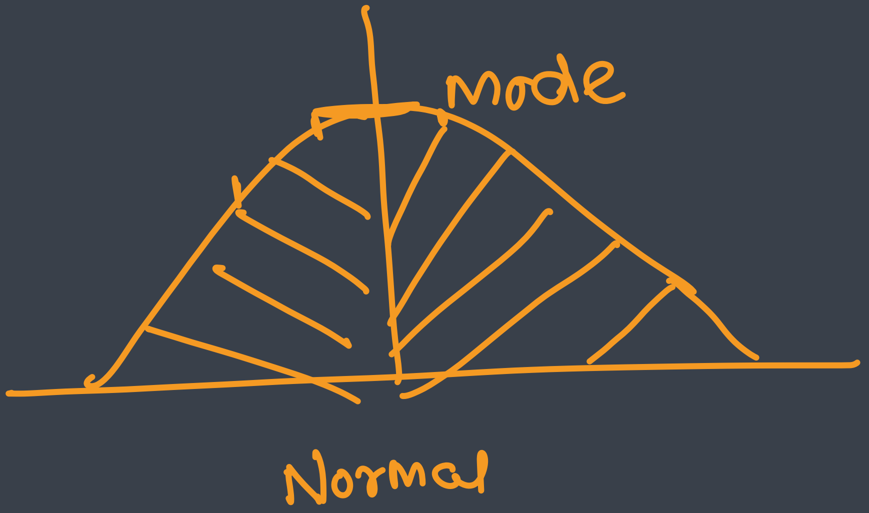


Dispersion vs Skewness

- Dispersion is concerned with the amount of variation rather than its direction
- Skewness tells us the the direction of variation or departure from symmetry → Normal distribution
- Measures of symmetry are dependent upon the amount of dispersion
- Although skewness is an important characteristic for defining precise pattern or shape of distribution, it is rarely calculated in business and economic series
- Variation is by far the most important characteristic of a distribution

Tests of skewness

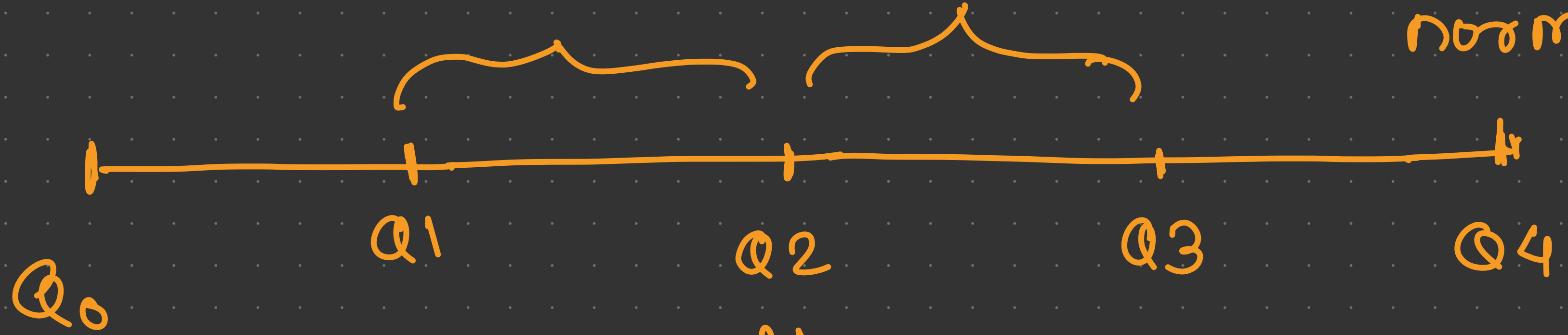
- In order to assert if a distribution is skewed or not, the following tests may be applied
  - The values of mean, mode and median do not coincide
  - When data are plotted on the graph they do not give the normal bell shaped curve i.e. when cut along the vertical line through the center, the two halves are not equal
  - The positive sum of deviations is not exactly equal to the sum of negative deviations from the deviation
  - Quartiles are not equidistant from the median
  - Frequencies are not equally distributed at points of equal deviation from the mode



$Q2 - Q1 == Q3 - Q2$

symmetrical

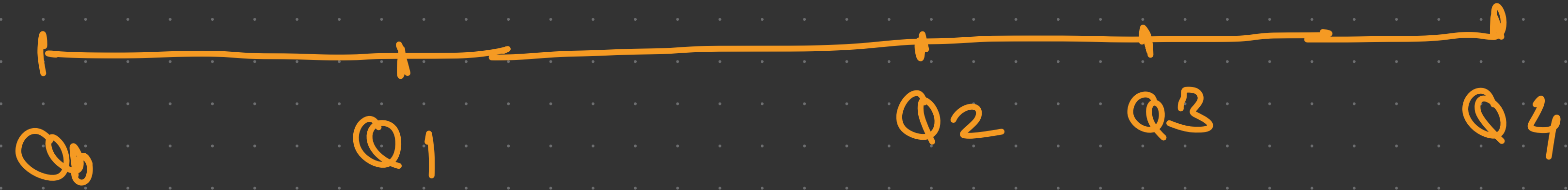
normal



median

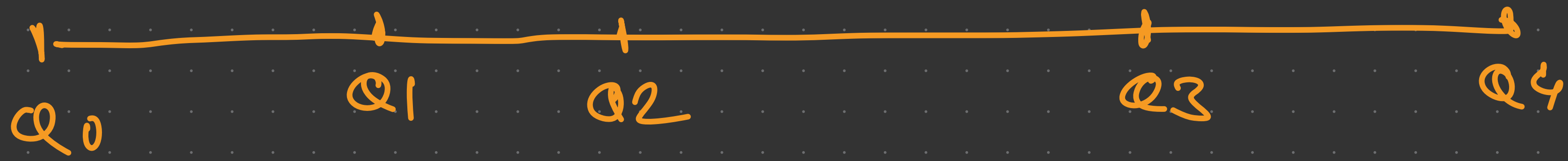
mean > median

+ve skewed



mean < median

-ve skewed



Properties of good measures of skewness

- It should be a pure number in the sense that its value should independent of the units of series and also the degree of variation in the series
- It should have a zero value when the distribution is symmetrical
- It should have some meaningful scale of measure so that we could easily interpret the measured value

Measures of Skewness

- Absolute measure
  - Skewness can be measured in absolute terms by taking the difference between mean and ~~mode~~ median
  - Symbolically mean - median
    - Absolute skewness = mean – mode
  - If the value of mean is greater than mode, the skewness will be positive (we shall get plus sign in the result)
  - If the value of mode is greater than mean, the skewness will be negative (we shall get the negative result)
- Relative measures
  - Karl Pearson's coefficient of skewness ***
  - Bowley's coefficient of skewness
  - Kelly's coefficient of skewness
  - Measure of skewness based on the moments

Karl Pearson's Coefficient of Skewness

- It is also known as Pearsonian Coefficient of skewness
- It is based upon the difference between mean and mode
- This difference is divided by standard deviation to give relative measure

$$Sk_p = \frac{mean - mode}{\sigma}$$

$$Sk_p = \frac{3(mean - median)}{\sigma}$$

- There is no limit to this measure in theory and this is a slight drawback
- But in practice, the value given by this formula is rarely high and usually lies between + or - 1
- When the distribution is symmetrical, mean mode and median will coincide and result will be zero
- The coefficient will have plus sign if distribution is positively skewed and negative if the distribution is negatively skewed

108 99 112 111 108

mean = 107.6

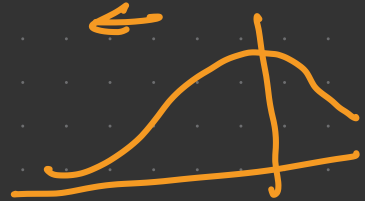
$$mean = \frac{\Sigma N}{N} = \frac{538}{5} = 107.6$$

mode = 108

median = 108

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}} = 4.58$$

-vely



$$Skp = \frac{mean - mode}{\sigma} = -0.08$$

Q1 = 103.5

Q2 = 108

Q3 = 111.5

$$Skp = \frac{3(mean - median)}{\sigma} = -0.26$$

Bowley's Coefficient of Skewness

- An alternative measure of skewness has been proposed by late professor Bowley
- It is based on quartiles
- In symmetrical distribution, first and third quartiles are equidistant from the median
- If distribution is positively skewed, the top 25% of values will tend to be farther from median than the bottom 25% values i.e. Q3 will be farther from median than Q1 is from median and reverse for negatively skewed distribution

$$Sk_B = \frac{Q3 + Q1 - 2 * median}{Q3 - Q1}$$

$$= \frac{111.5 + 103.5 - 2 * 108}{111.5 - 103.5}$$

$$= -0.125$$

- Drawback: It ignores two extreme quartiles of the distribution

Kelly's Coefficient of Skewness

- This coefficient is better than Bowley's as it can be used with any two deciles or percentiles equidistant from median
- That is it can be extended by taking any two deciles or percentiles equidistant from median

$$S_K = \frac{D1 + D9 - 2 \ * median}{D9 \ - D1}$$

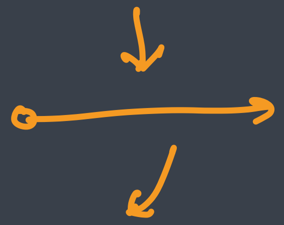
$$S_K = \frac{P10 + P90 - 2 \ * median}{P90 \ - P10}$$

- This method is not popular in practice and generally Karl Pearson's method is used

Moments

- In mechanical perspective, it is the measure of a force with respect to its tendency to provide rotation
- The strength of the tendency depends upon the amount of force and distance from the origin
- For example, if a number of forces F1, F2, F3 … Fn at distances X1, X2, X3 … Xn are applied, then
  - the moment of the first force about the origin is F1X1
  - the moment of the second force about the origin is F2X2
  - the moment of the nth force about the origin is FnXn
- If the total moment is divided by the sum of force, the equivalent is termed as 'a moment'

$$\text{Moment} = \frac{\sum FX}{\sum F}$$



Moments in statistics

- The moments in statistics are used to describe the various characteristics of frequency distribution like the central tendency, variation, skewness and kurtosis
- It can be seen that the formula for moment is identical to the arithmetic mean, which is why the arithmetic mean is considered as "first moment about the origin"
- The arithmetic mean of various powers of the deviations in any distribution is called as moment of that distribution
  - The mean of first power of deviation – first moment about the mean (μ_1)
  - The mean of second power of deviation – second moment about the mean (μ_2)
  - The mean of third power of deviation – third moment about the mean (μ_3)
  - The mean of forth power of deviation – forth moment about the mean (μ_4)

$$\mu_r = \frac{\Sigma(x-\bar{x})^r}{N}$$

$$\mu_1 = \left[\frac{\Sigma(x-\bar{x})^1}{N}\right]$$

$$\mu_2 = \frac{\Sigma(x-\bar{x})^2}{N}$$

$$\mu_3 = \frac{\Sigma(x-\bar{x})^3}{N}, \quad \mu_4 = \frac{\Sigma(x-\bar{x})^4}{N}$$

Characteristics of moments

- Two important characteristics of a distribution can be calculated from μ_2, μ_3 and μ_4

$$\beta_1 = \frac{{\mu_3}^2}{{\mu_2}^3}$$

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2}$$

$$\mu_3 = \frac{\Sigma (x-\bar{x})^3}{N}$$

$$\mu_4 = \frac{\Sigma (x-\bar{x})^4}{N}$$

$$\mu_2 = \frac{\Sigma (x-\bar{x})^2}{N}$$

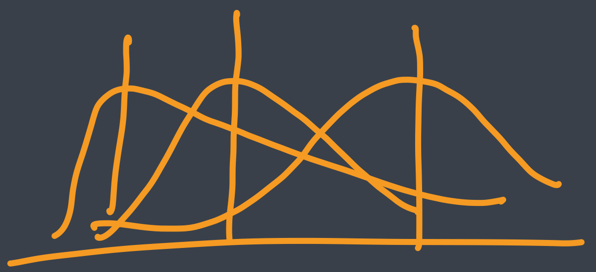
- β_1 measures skewness
- β_2 measures kurtosis

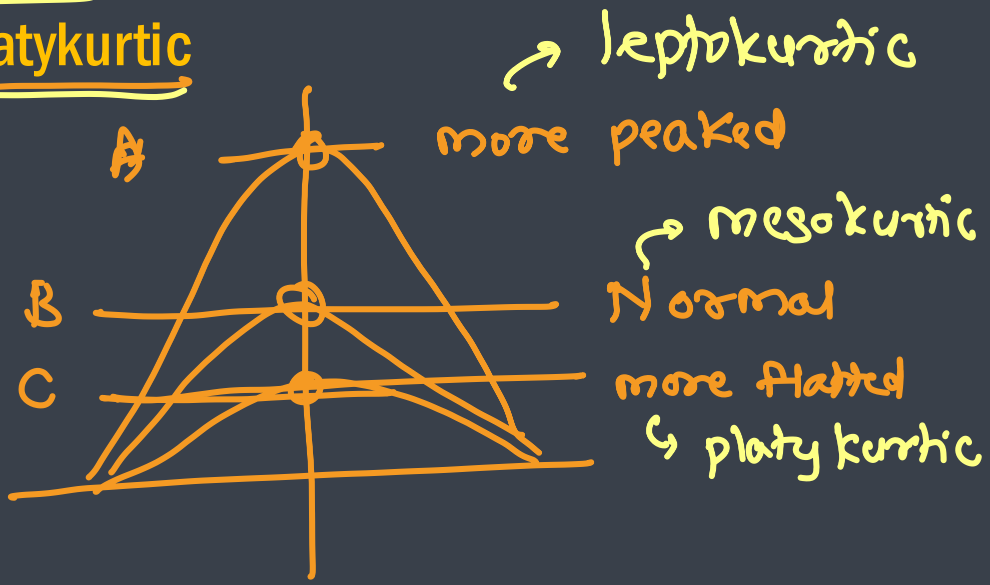
Purpose of Moments

- With the help of moments, one can measure the central tendency of a set of observations, their asymmetry and the height of peak of their curve that they make
- Because of the great convenience in obtaining measure of various characteristics of a frequency distribution, the calculation of the first four moments about the mean may well be made in the first step of statistical analysis
- Following are the summary of first four moments
  - First moment about the origin => Mean
  - Second moment about the origin => Variance
  - Third moment about the origin => Skewness
  - Forth moment about the origin => Kurtosis

Kurtosis

- Kurtosis in Greek means ‘bulginess’
- In statistics, kurtosis refers to the degree of flatness or peakedness in the region about the mode of frequency curve
- The degree of Kurtosis of a distribution is measured relative to the peakedness of normal curve
- It tells us the extent to which distribution is more peaked or flat-topped compare to normal
- If the curve is more peaked than normal, it is called as leptokurtic
- If the curve is more flat-topped than normal, it is called as platykurtic
- The normal curve itself is called as Mesokurtic





Examples

- Calculate mean, mode, median, mean deviation, coefficient of MD, quartile deviation, coefficient of QD, variance, standard deviation, coefficient of variance, skewness, kurtosis and forth moment of the following dataset

- since skewness is 0.052, the data set is +vely skewed
- since the kurtosis is 1.56, dataset is leptokurtic

$Sk_P = 0.61$, $Sk_B = 0.8$

individual

X
10
20
34
56
85
30
30
69
71

discrete

X	F
12.5	28
17.5	42
22.5	54
27.5	108
32.5	129
37.5	61
42.5	45
47.5	32

continuous

Profits	#companies
70-80	12
80-90	18
90-100	35
100-110	43
110-120	50
120-130	45
130-140	30
140-150	8

x	$\lvert x-\bar{x}\rvert$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$(x-\bar{x})^4$
10	-35			
20	-25			
34	-11			
56	11			
85	40			
30	-15			
30	-15			
69	24			
71	26			
405	202	5394	30150	5370534

$$\bar{x} = \frac{\Sigma x}{N} = \frac{405}{9} = 45$$

$$\text{mode} = 30$$

$$Q2 = \text{median} = 34$$

$$MD = \frac{\Sigma \lvert x-\bar{x}\rvert}{N} = 22.44$$

$$\text{coe. of } MD = \frac{MD}{\bar{x}} = 0.49$$

$$Q_1 = 30, \quad Q3 = 71$$

$$IQR = Q3 - Q1 = 41$$

$$QD = IQR/2 = 20.5$$

$$\text{coe of } QD = \frac{IQR}{Q3+Q1} = \frac{41}{101} = 0.40$$

$$\sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2}{N}} = \sqrt{\frac{5394}{9}} = 24.48$$

$$var = \sigma^2 = 599.33$$

$$CV = \frac{\sigma}{\bar{x}} \times 100 = \frac{24.48}{45} \times 100 = 54.4$$

$$\mu_2 = \frac{\Sigma (x-\bar{x})^2}{N} = var = 599.33$$

$$\mu_3 = \frac{\Sigma (x-\bar{x})^3}{N} = \frac{30150}{9} = 3350$$

$$\mu_4 = \frac{\Sigma (x-\bar{x})^4}{N} = 596726$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0.052$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = 1.66$$