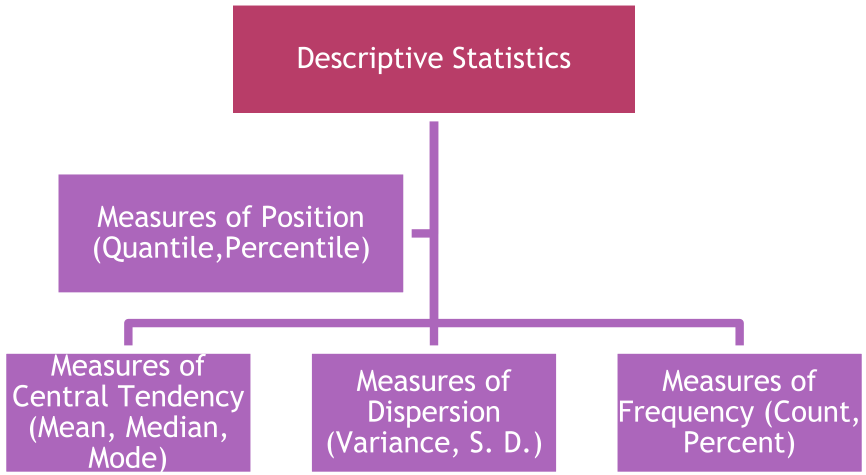
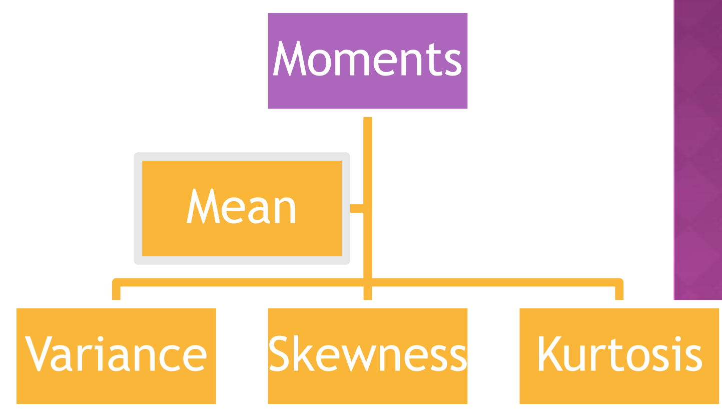
SKEWNESS, KURTOSIS AND QUANTILE

SKEWNESS



Descriptive Statistics
Measures of Position (Quantile,Percentile)
Measures of Central Tendency (Mean, Median, Mode)
Measures of Dispersion (Variance, S. D.)
Measures of Frequency (Count, Percent)



Moments
Mean
Variance
Skewness
Kurtosis

SKEWNESS

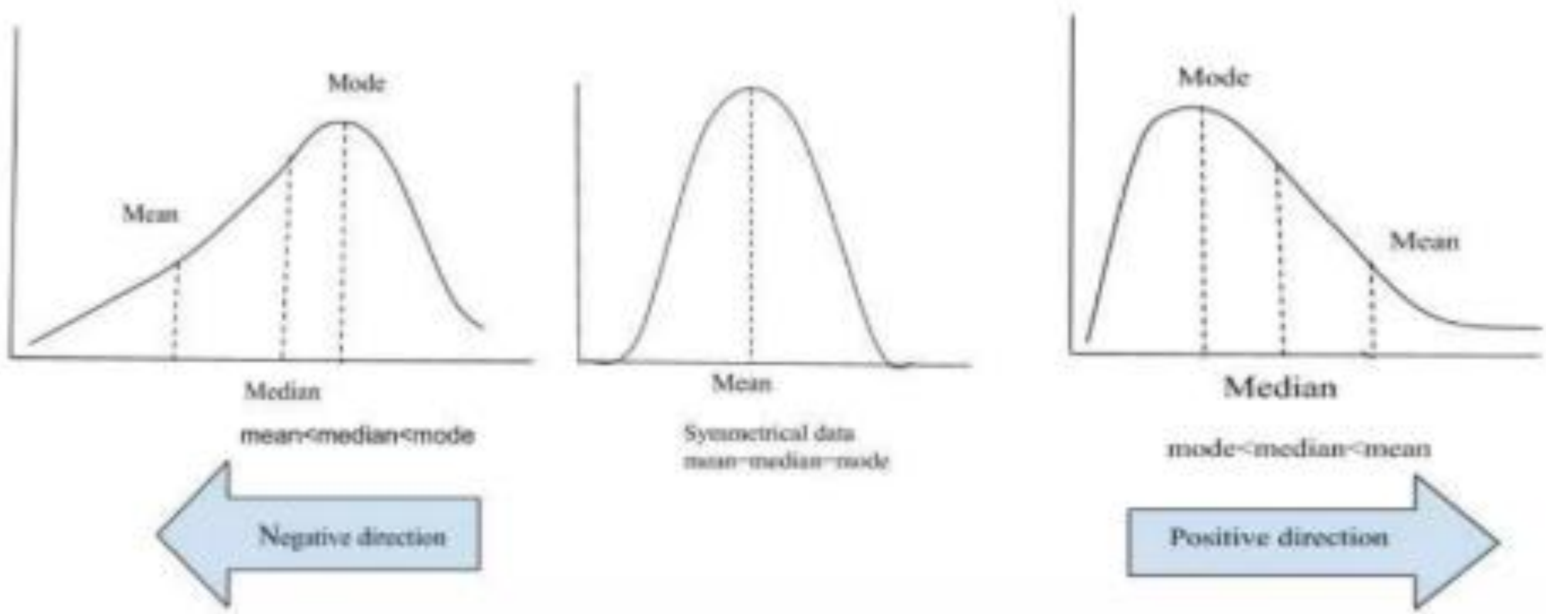
- In Probability Theory and Statistics, Skewness is a measure of Asymmetry of the probability distribution of a real valued random variable about its mean (Mean and Median fall at different points in the distribution)
- The Skewness values can be zero, negative or unidentified.
- Skewness discovered by Karl Pearson (1894)

- If one tail is longer than another, the distribution is skewed.
- These distribution are sometimes called asymmetrical distribution as they don’t show any kind of symmetry.
- Within each graph, the values on the right side of the distribution taper differently from the values on the left side.
- These tapering sides are called tails, and they provide a visual means to determine which of the two kinds of skewness a distribution has :

1) Positive Skewness , 2) Negative Skewness

Negative Skewness	Positive Skewness
The left tail is longer, the mass of the distribution is concentrated on the right of the figure.	The right tail is longer, the mass of the distribution is concentrated on the left of the figure.
Long tail in the negative direction on the number line.	Long tail in the positive direction on the number line.
The mean is also to the left of the peak.	The mean is also to the right of the peak.
The distribution is left-skewed, negatively-skewed distributions.	The distribution is right-skewed, positively-skewed distributions.

PICTORIAL REPRESENTATION



RULES FOR SKEWNESS

- If the Skewness is between -0.5 to 0.5 : Approximately Symmetrical
- If the Skewness is between -1 to 1 : Moderately Skewed
- If the Skewness is less than -1 or greater than $+1$: Highly Skewed

SKEWED DISTRIBUTION

There are Three types:

- Symmetrical Distribution

  A. M. = Median = Mode

- Positively Skewed Distribution

  A. M. > Median > Mode

- Negatively Skewed Distribution

  A. M. < Median < Mode

MEASURES OF SKEWNESS

- Karl Pearson's Coefficient of Skewness
- Bowley's Coefficient of Skewness
- Kelly's Coefficient of Skewness

KURTOSIS

- Kurtosis refers to Peakedness or flatness or curvedness of the distribution
- The larger the Kurtosis, the more peaked will be distribution.
- Kurtosis is always positive number.

- In some distributions, the values of Mean, Median and Mode are the same.
- But if a curve is drawn from the distribution then the height of the curve is either more or less than the normal probability curve, since such type of deviation is related with the crest of the curve, it is called Kurtosis.

TYPES OF KURTOSIS

Three types of kurtosis that can be exhibited by a distribution

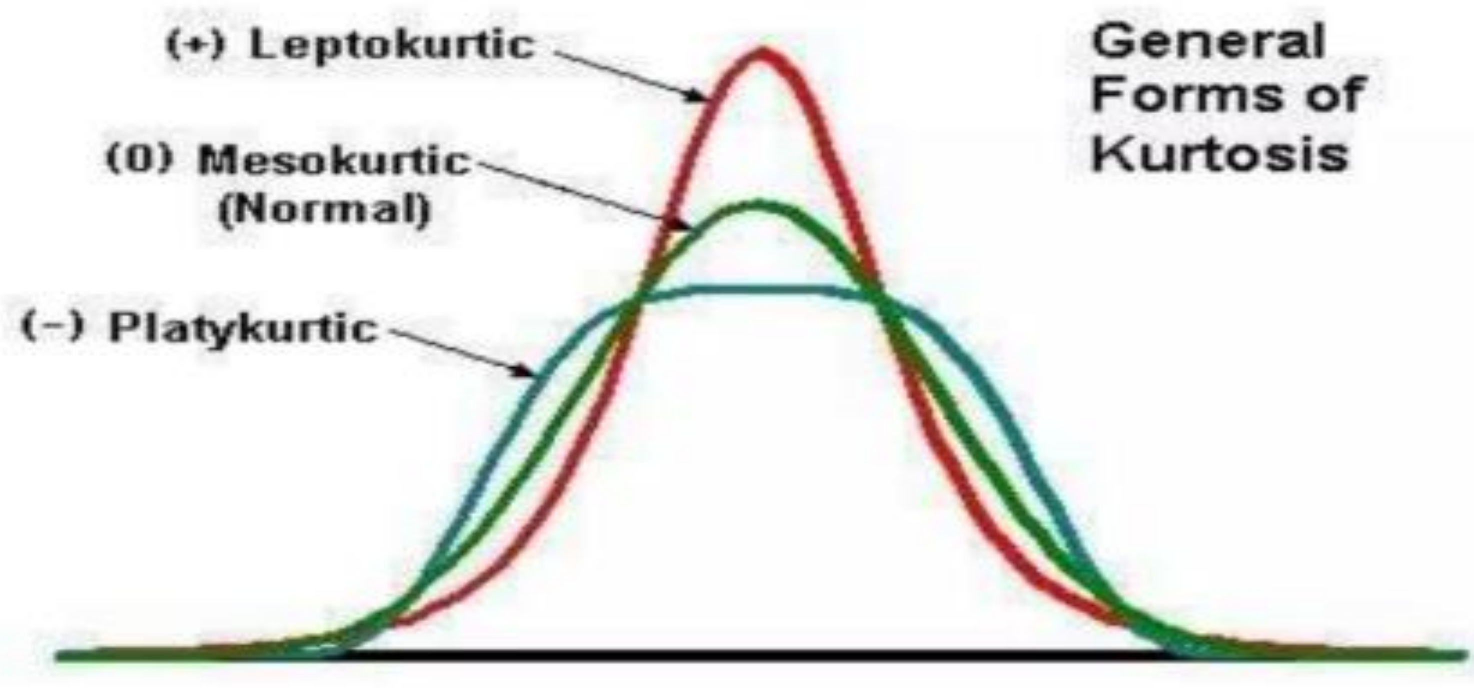
- Mesokurtic – same as the normal distribution with zero
- Platykurtic – less than normal distribution, short tailed distribution, thin tail, negative kurtosis
- Leptokurtic - more than normal distribution, heavy tailed distribution, fatter tail, positive kurtosis

CALCULATION OF KURTOSIS

Kurtosis is measured by β_2

- If the value of $\beta_2 > 3$, the curve is more peaked than the normal i.e. Leptokurtic
- If the value of $\beta_2 < 3$, the curve is less peaked than the normal i.e. platykurtic
- If the value of $\beta_2 = 3$, then the curve is having normal peak i.e. Mesokurtic

GENERAL FORMS OF KURTOSIS



QUANTILES

- Percentile
- Quartile
- Decile

QUANTILES

Unlike Mean, Median and Mode which generally describe the centre of distribution, Percentile, Quartile and Decile characterize a specific location of the distribution.

- For example, we want to know the score of the rank 30 students in a class of 100. we can not use the median formula since it will give us the score of the middle rank student which is most probably the 50^{th} and 51^{st} student in rank. We need to use another measure of location which can be either percentile or decile.

QUARTILES

- Quartiles are three values that split a data in four equal parts.
- The three quartiles are named as Q_1, Q_2 and Q_3, in which $Q_1$is the 25th of the data, Q_2 is the 50th of the data or the median of the data, Q_3 is the 75th of the data.

QUARTILES

$$Q_1 = [\frac{N+1}{4}]^{th}\ item$$

$$Q_2 = [\frac{N+1}{2}]^{th}\ item$$

$$Q_3 = [\frac{3(N+1)}{4}]^{th}\ item$$

where, n is the total number of observations, Q_1 is First Quartile, Q_2 is Second Quartile, and Q_3 is Third Quartile.

EXAMPLE - 1

Calculate the lower and upper quartiles of the following weights in the family: 25, 17, 32, 11, 40, 35, 13, 5, and 46.

Solution:

First of all, organize the numbers in ascending order.

5, 11, 13, 17, 25, 32, 35, 40, 46

Lower quartile, $Q_1 = \left[\frac{N+1}{4}\right]^{th}$ term

$Q_1 = \left[\frac{9+1}{4}\right]^{th}$ term

$Q_1 = 2.5^{th}$ term

As per the quartile formula;

$$Q_1 = 2nd\ term + 0.5(3rd\ term\ - 2nd\ term)$$

$$Q_1 = 11 + 0.5(13 - 11) = 12$$

Upper quartile, $Q_3 = \left[\frac{3(N+1)}{4}\right]^{th}$ term

$Q_3 = \left[\frac{30}{4}\right]^{th}$ term

Q_3 = 7.5th term

As per the quartile formula;

$$Q_3 = 7th\ term + 0.5(8th\ term\ - 7th\ term)$$

$$Q_3 = 35 + 0.5(40 - 35) = 37.5$$

EXAMPLE-2

- Calculate Q1 and Q3 for the data related to the age in years of 99 members in a housing society.

Age (in years)	Number of Members
10	20
18	5
25	10
35	30
40	20
45	14

EXAMPLE-2

Age	Number of Members	Cumulative frequency (cf)
10	20	20
18	5	25
25	10	35
35	30	65
40	20	85
45	14	99

$Q_1 = \left[\frac{N+1}{4}\right]^{th}$ term

$Q_1 = \left[\frac{99+1}{4}\right]^{th}$ term

Q_1 = 25^{th} term

Now, the 25^{th} item falls under the cumulative frequency of 25 and the age against this cf value is 18.

$Q_1 = 18$ years

$Q_3 = \left[\frac{3(N+1)}{4}\right]^{th}$ term

$Q_3 = \left[\frac{300}{4}\right]^{th}$ term

Q_3 = 75^{th} term

Now, the 75^{th} item falls under the cumulative frequency of 85 and the age against this cf value is 40.

Q_3 = 40 years

EXAMPLE - 3

Determine the quartiles Q_1 and Q_3 for the company's salary listed below.

Salaries (per day in ₹)	Number of Employees
500-600	10
600-700	12
700-800	16
800-900	14
900-1000	8

EXAMPLE - 3

Salaries (per day in ₹)	Number of Employees	Cumulative Frequency (c.f.)
500-600	10	10 (m_1)
600-700	**12 (f_1)**	22
700-800	16	38 (m_3)
800-900	**14 (f_3)**	52
900-1000	8	60

EXAMPLE - 3

$Q1\ Class = \frac{N}{4}$

Q1 Class $= \frac{60}{4}$ **= 15th item**

Now, the 15th item falls under the cumulative frequency 22 and the salary against this cf value lies in the group 600-700.

$$\boldsymbol{Q_1 = l_1 + \frac{\frac{N}{4} - m_1}{f_1} \times c_1}$$

$$\boldsymbol{Q_1 = 600 + \frac{\frac{60}{4} - 10}{12} \times 100}$$

Q_1 = ₹641.67

EXAMPLE - 3

$Q3\ Class = \frac{3N}{4}$

Q3 Class $= \frac{180}{4}$ **= 45th item**

Now, the 45th item falls under the cumulative frequency 52 and the salary against this cf value lies in the group 800-900.

$$\boldsymbol{Q_3 = l_1 + \frac{\frac{3N}{4} - m_3}{f_3} \times c_3}$$

$$\boldsymbol{Q_3 = 800 + \frac{\frac{180}{4} - 38}{14} \times 100}$$

Q_3 = ₹ 850

DECILES

- Deciles are 9 values that split data in 10 equal parts.
- Each decile represents a multiple of 10 of the data of the total data.
- D_1, D_2, D_3 and so on until D_9 represents the 10^{th} until the 90^{th} of the data.

DECILES

$$D_1 = [\frac{N+1}{10}]^{th}\ item$$

$$D_2 = [\frac{2(N+1)}{10}]^{th}\ item$$

$$............D_9 = [\frac{9(N+1)}{10}]^{th}\ item$$

Where, n is the total number of observations, D_1 is First Decile, D_2 is Second Decile,..........D_9 is Ninth Decile.

EXAMPLE-1

Calculate D_1 and D_5 from the following weights in the family: 25, 17, 32, 11, 40, 35, 13, 5, and 46.

Solution:

First of all, organize the numbers in ascending order.

5, 11, 13, 17, 25, 32, 35, 40, 46

$$D_1 = \left[\frac{N+1}{10}\right]^{th} \text{term}$$

$$D_1 = \left[\frac{9+1}{10}\right]^{th} \text{term}$$

D_1 = 1st term $= 5$

$$D_5 = \left[\frac{5(N+1)}{10}\right]^{th} \text{term}$$

$$D_5 = \left[\frac{5(9+1)}{10}\right]^{th} \text{term}$$

D5 = 5th term $= 25$

EXAMPLE-2

- Calculate D2 and D6 for the data related to the age in years of 99 members in a housing society.

Age (in years)	Number of Members
10	20
18	5
25	10
35	30
40	20
45	14

EXAMPLE-2

Age	Number of Members	Cumulative frequency (cf)
10	20	20
18	5	25
25	10	35
35	30	65
40	20	85
45	14	99

$D_2 = \left[\frac{2(N+1)}{10}\right]^{th}$ term

$D_2 = \left[\frac{2(99+1)}{10}\right]^{th}$ term = 20^{th} term

Now, the 20^{th} item falls under the cumulative frequency of 20 and the age against this cf value is 10.

D_2 = 10 years

$D_6 = \left[\frac{6(N+1)}{10}\right]^{th}$ term

$D_6 = \left[\frac{6(99+1)}{10}\right]^{th}$ term = 60^{th} term

Now, the 60^{th} item falls under the cumulative frequency of 65 and the age against this cf value is 35.

D_6 = 35 years

EXAMPLE - 3

Determine D_4 for the company's salary listed below.

Salaries (per day in ₹)	Number of Employees
500-600	10
600-700	12
700-800	16
800-900	14
900-1000	8

Salaries (per day in ₹)	Number of Employees	Cumulative Frequency (c.f.)
500-600	10	10
600-700	12	22 (m)
700-800	16 (f)	38
800-900	14	52
900-1000	8	60

In case N is an even number, the following formula is used:

$$D_4 = \left(\frac{4N}{10}\right) th \text{ item}$$

$$D_4 = \left(\frac{4 \times 60}{10}\right) th \text{ item} = 24^{th} \text{ Item}$$

Now, the 24^{th} item falls under the cumulative frequency 38 and the salary against this cf value lies in the group 700-800.

$$\boldsymbol{D_4 = l + \frac{\frac{4N}{10} - m}{f} \times c}$$

$$\boldsymbol{D_4 = 700 + \frac{\frac{4 \times 60}{10} - 22}{16} \times 100}$$

D_4 = ₹712.5

PERCENTILE

- Percentiles are 99 values that split data in 100 equal parts.
- Each Percentile represents a multiple of 1 of the total value.
- P_1, P_2, P_3 and so on until P_{99} represents the 1^{st} until the 99^{th} of the data.

PERCENTILE

$$P_1 = [\frac{N+1}{100}]^{th}\ item$$

$$P_2 = [\frac{2(N+1)}{100}]^{th}\ item$$

$$P_3 = [\frac{3(N+1)}{100}]^{th}\ item$$

$$..................P_{99} = [\frac{99(N+1)}{100}]^{th}\ item$$

Where, n is the total number of observations, P_1 is First Percentile, P_2 is Second Percentile, P_3 is Third Percentile, ...P_{99} is Ninety Ninth Percentile.

EXAMPLE-1

Calculate P_{20} and P_{90} from the following weights in the family: 25, 17, 32, 11, 40, 35, 13, 5, and 46.

Solution:

First of all, organize the numbers in ascending order.

5, 11, 13, 17, 25, 32, 35, 40, 46

$P_{20} = \left[\frac{20(N+1)}{100}\right]^{th}$ term

$P_{20} = \left[\frac{20(9+1)}{100}\right]^{th}$ term

P_{20} = 2nd term $=11$

$P_{90} = \left[\frac{90(N+1)}{100}\right]^{th}$ term

$P_{90} = \left[\frac{90(9+1)}{100}\right]^{th}$ term

P_{90} = 9th term $= 46$

EXAMPLE-2

- Calculate P10 and P75 for the data related to the age in years of 99 members in a housing society.

Age (in years)	Number of Members
10	20
18	5
25	10
35	30
40	20
45	14

EXAMPLE-2

Age	Number of Members	Cumulative frequency (cf)
10	20	20
18	5	25
25	10	35
35	30	65
40	20	85
45	14	99

$$P_{10} = \left[\frac{10(N+1)}{100}\right]^{th} \text{term}$$

$$P_{10} = \left[\frac{10(99+1)}{100}\right]^{th} \text{term} = 10^{\text{th}} \text{ term}$$

Now, the 10^{th} item falls under the cumulative frequency of 20 and the age against this cf value is 10.

P_{10} = 10 years

$$P_{75} = \left[\frac{75(N+1)}{100}\right]^{th} \text{term}$$

$$P_{75} = \left[\frac{75(99+1)}{100}\right]^{th} \text{term} = 75^{\text{th}} \text{ term}$$

Now, the 75^{th} item falls under the cumulative frequency of 85 and the age against this cf value is 40.

$$\boldsymbol{P_{75} = 40\ years}$$

EXAMPLE - 3

Determine P_{50} for the company's salary listed below.

Salaries (per day in ₹)	Number of Employees
500-600	10
600-700	12
700-800	16
800-900	14
900-1000	8

- Solution

Salaries (per day in ₹)	Number of Employees	Cumulative Frequency (c.f.)
500-600	10	10
600-700	12	22 (m)
700-800	16 (f)	38
800-900	14	52
900-1000	8	60

In case N is an even number, the following formula is used:

$$P_{50} = \left(\frac{50N}{100}\right) th \text{ item}$$

$$P_{50} = \left(\frac{50\times 60}{100}\right) th \text{ item} = 30^{th} \text{ Item}$$

Now, the 30^{th} item falls under the cumulative frequency 38 and the salary against this cf value lies in the group 700-800.

$$\boldsymbol{P_{50} = l + \frac{\frac{50N}{100} - m}{f} \times c}$$

$$\boldsymbol{P_{50} = 700 + \frac{\frac{50\times 60}{100} - 22}{16} \times 100}$$

P_{50} = ₹750