Unit - I

Measures of Central Tendency (Averages)

Averages are " statistical constants which enable us to comprehend in a single effort the significance of the whole".

— Bowley

They give us an idea about the concentration of the values in the central part of the distribution. In other words, an average of a statistical series is the value of the variable which is representative of the entire distribution. The following are the five commonly used measures of central tendency, ~~that are~~

(i) Arithmetic mean (or) Simple mean
(ii) median
(iii) Mode
(iv) Geometric Mean and
(v) Harmonic Mean.

1. Arithmetic Mean:

Arithmetic mean of a set of obsⁿˢ is their sum divided by the number of obsⁿˢ. E.g: The A.M $\bar{x}$ of n obsⁿˢ $x_1, x_2 \dots x_n$ is given by

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For raw data

Incase of frequency distribution:

$$\bar{x} = \frac{\Sigma f x}{N} \; ; \; N = \text{total frequency}$$

f = frequency

Short-cut method: (i) Raw data:

$$AM, \bar{x} = A + \frac{\Sigma d}{n} \; ; \; d = x - A, \; A = \text{Assumed mean}$$

(ii) Discrete frequency distribution:

$$AM, \bar{x} = A + \frac{\Sigma fd}{N} \; ; \; d = x - A, \; A = \text{Assumed mean}$$

N = total frequency

(iii) Continuous frequency distribution:

$$AM, \bar{x} = A + \frac{\Sigma fd}{N} \; ; \; d = x - A, \; A = \text{Assumed mean}$$

$$\bar{x} = A + \frac{\Sigma fd}{N} \times c \quad \text{Step deviation method}$$

$$d = \frac{x - A}{c} \; ; \; A = \text{Assumed mean}$$

c = Class Interval

Ex:

1. The monthly income (in rupees) of 10 employees working in a firm is as follows
4487, 4493, 4502, 4446, 4475, 4492, 4572, 4516, 4468, 4489. (Ans, $\bar{x} = 4494$)

2. Find the arithmetic mean of the following frequency distribution:

x:	1	2	3	4	5	6	7
f:	5	9	12	17	14	10	6

(Ans, $\bar{x} = 4.09$)

3. Calculate the A.M of the marks from the following table:

Marks :	0-10	10-20	20-30	30-40	40-50	50-60
No. of students:	12	18	27	20	17	6

(Ans: $\bar{x} = 28$)

4. calculate the mean for the following frequency distribution:

C.I :	0-8	8-16	16-24	24-32	32-40	40-48
f :	8	7	16	24	15	7

(Ans, $\bar{x} = 25.404$)

Variables:

(i) Individual observations → 4, 8, 10 ...

(ii) Discrete variable →

x :	1	2	3	4	5
f :	2	5	10	8	5

(iii) Continuous variable →

x :	0-10	10-20	20-30
f :	1	5	4

Frequency distribution:

"Frequency distribution is a statistical table which shows the set of all distinct values of the variable arranged in order of magnitude, either individually or in groups, with their corresponding frequencies side by side".

Discrete frequency distribution:

Example: No of rooms in the houses kept by 25 families.

1 2 4 3 4 2 5 3 2 2 4 1 2 3 5

1 3 5 1 3 3 1 3 1 1

No of rooms	Tallies	frequency
1	~~||||~~ ||	7
2	~~||||~~	5
3	~~||||~~ ||	7
4	|||	3
5	|||	3
		25

Scanned by CamScanner

Thus from the above table, it is clear that out of 25 families 7 were living in 1 room each, 5 in 2 rooms each, 7 in 3 rooms each, 3 in 4 rooms and 3 in 5 rooms each.

— x —

Continuous frequency distribution:

* Class limits — Ex: 20 - 40
* class Interval — Ex: 20-40 ; C.I = 20
* class frequency — Ex: 20-40 40 - f
* class mid point — $M.P = \frac{UL + LL}{2}$
* Exclusive method — 0-10, 10-20, 20-30 ...
* Inclusive method — 0-9 10-19 20-29 ...
* Number of classes:

$$K = 1 + 3.322 \log N$$

* Magnitude of class interval

$$i = \frac{Range}{1 + 3.322 \log N}$$

(ie) Range = L - S.

Ex: The profits (in lakshs of rupees) of 30 Companies for the year 1999-2000 are given below

20, 22, 35, 42, 37, 42, 48, 53, 49, 65, 39

48, 67, 18, 16, 23, 37, 35, 49, 63, 65, 55

45, 58, 57, 69, 25, 29, 58, 65

Classify the above data taking a suitable class Interval.

Sol: W.K.T

$K = 1 + 3.322 \log N = 1 + 3.322 \log 30 = 5.91$

$i = \frac{\text{Range}}{K} = \frac{69-16}{5.91} = 8.97 \text{ or } 9$

Profits (Rs. laksh)	Tally mark	frequency
15-25	IIII	5
25-35	II	2
35-45	IIII II	7
45-55	IIII I	6
55-65	IIII	5
65-75	IIII	5

Median :

Median of a distribution is the value of the variable which divides it into two equal parts. It is the value which exceeds and is exceeded by the same number of observations. That is, it is the value such that the number of obsns above it is equal to the number of observations below it.

Formula:

I. Raw data :

Median = Size of $\left(\frac{n+1}{2}\right)^{th}$ observation.

[First arrange the data in ascending or descending order of Magnitude]

where, n = no. of obsn.

II Discrete data :

Median = Size of $\left(\frac{N+1}{2}\right)^{th}$ obsn.

[After calculating the cumulative frequency]

where N = total frequency.

III Continuous frequency distribution:

$$\text{Median} = l + \frac{N/2 - m}{f} \times c$$

where, l – is the lower limint of the median class

f – is the frequency of the "

m – Cumulative frequency above the median class

c – class width of the M.C.

N – total frequency.

Ex:

1. From the following data of wages of 7 workers, compute the median wage:

4600, 4650, 4580, 4690, 4660, 4606, 4640.

$\boxed{\text{Ans: } 4640}$

2. Obtain the median for the following frequency distribution

x:	1	2	3	4	5	6	7	8	9
f:	8	10	11	16	20	25	15	9	6

$\boxed{\text{Ans: median} = 5}$

3. Calculate the median from the following data:

Profits (Rs. Crore) :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Companies	4	12	24	36	20	16	8	5

$\boxed{\text{Ans : median} = 36.25}$

Mode: (Raw data)

Mode is the value which occurs most frequently in a set of observations. It is the value of the variable which predominant in the series.

Discrete frequency distribution:

Mode is the value of x corresponding to maximum frequency.

But in any one of the following cases:

(i) If the maximum frequency is repeated

(ii) If the maximum frequency occurs in the very beginning or at the end of the distribution, and

(iii) if there are irregularities in the distn, Ex: Find the mode of the frequency distribution

x:	1	2	3	4	5	6	7	8
f:	4	9	16	25	22	15	7	3

Ans: mode = 4

Ex: Find the mode of the following frequency dist.

x:	1	2	3	4	5	6	7	8	9	10	11	12
f:	3	8	15	23	35	40	32	28	20	45	14	6

Ans:

(i) original frequency (ii) Combine 2×2 (iii) leave 1st and combine 2×2 (iv) Combine 3×3 (v) leave 1st and then combine 3×3 (vi) leave 1st two and combine 3×3.

Then prepare the follow table

column Number (1)	max. frequency (2)	combination of values of x (3)

∴ mode = 6

Ex: Find the mode for the following distribution

C.I :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
f :	5	8	7	12	28	20	10	10

Ans: mode = 46.67

Ex:
In a moderately symmetrical distribution, the mode and mean are 32.1 and 35.4 respectively. calculate median.

Ans: mode = 3 median − 2mean , median = 34.3

Ex:
If in a moderately asymmetrical frequency distn. the values of median and A.M are 72 and 78 respectively, estimate the value of the mode.

Ans: mode = 60.

PROBLEMS:

1. Calculate mean from the following data:

(a) 40, 50, 55, 78, 58, 60, 73, 35, 43, 48, Ans: 54

(b)

x :	1	2	3	4	5	6	7	8	9	10
f :	21	30	28	40	26	34	40	9	15	57

Ans: 5.72

(c)

C.I :	100-200	200-300	300-400	400-500	500-600	600-700	700-800
f :	10	18	20	26	30	28	18

Ans: 486

(d)

Ex: The average marks secured by 36 students was 52. But it was discovered that an item 64 was misread as 46. Find the correct mean of marks.

Sn Given, $n = 36$, $\bar{x} = 52$

W.K.T, $\bar{x} = \frac{\Sigma x}{n} \Rightarrow \bar{x} = \frac{\Sigma x}{36} \Rightarrow \bar{x} \times 36 \Rightarrow \Sigma x$

$\Rightarrow 52 \times 36 = 1872$

$\therefore \Sigma x = 1872$ (wrong)

Correct:

Σx = incorrect Σx − wrong item + Correct item

$= 1872 - 46 + 64 = 1890$

$\therefore \bar{x} = \frac{\Sigma x}{n} = \frac{1890}{36} = 52.5$.

Ex: An average rainfall on a city from monday to Saturday is 0.3 inch. Due to heavy rainfall on Sunday, the average rainfall for the week increased to 0.5 inch. what was the rainfall on Sunday?

Sn $\bar{x} = \frac{\Sigma x}{n}$ (or) $\Sigma x = n\bar{x}$

For 6 days (ie) Monday to Saturday

$\Sigma x = 0.3 \times 6 = 1.8$ inch

For 7 days (ie) Monday to Sunday

$\Sigma x = 0.5 \times 7 = 3.5$ inch

$\therefore$ Rainfall on Sunday

$= 3.5$ inch $- 1.8$ inch

$= 1.7$ inch.

Geometric Mean:

Geometric mean of a set of n observations is the n^{th} root of their product. Let's $x_1, x_2 \dots x_n$ are the given n observation, then their G.M is given by

$$G.M = (x_1 \cdot x_2 \dots x_n)^{1/n}$$

taking log on both sides

$$\log G.M = \frac{1}{n} \log [x_1 \cdot x_2 \dots x_n]$$

$$= \frac{1}{n} [\log x_1 + \log x_2 + \dots + \log x_n]$$

$$= \frac{1}{n} \sum_{i=1}^{n} \log x_i$$

$$\therefore G.M = \text{Antilog} \left[\frac{1}{n} \sum_{i=1}^{n} \log x_i \right]$$

For raw data

For frequency distribution (Discrete or Continuous)

$$G.M = \text{Antilog} \left[\frac{\sum_{i=1}^{n} f_i \log x_i}{N} \right]$$

—x—

Ex: Calculate G.M from the following data
25, 32, 50, 42, 45, 30, 70, 65, 48, 51

Ans: G.M = 43.6707.

Ex: Calculate G.M of the following
50, 72, 54, 82, 93

Ans: G.M = 68.26

Ex: The following table gives the weight of 31 persons in sample survey. Calculate G.M.

weight (lbs)	130	135	140	145	146	148	149	150	157
No of persons	3	4	6	6	3	5	2	1	1

Ans: G.M = 142.5 lbs

Ex: Find out the G.M

x:	7.5–10.5	10.5–13.5	13.5–16.5	16.5–19.5	19.5–22.5	22.5–25.5	25.5–28.5
f:	5	9	19	23	7	4	1

Ans: G.M = 16.02

Harmonic Mean:

Harmonic mean is the reciprocal of the arithmetic mean of the reciprocals of the values of the items in a series.

I Raw data:

$$H.M = \frac{n}{\Sigma\left(\frac{1}{x}\right)}$$

II For frequency distribution:

$$H.M = \frac{N}{\Sigma(f/x)}$$

Ex:

calculate H.M form the following data

10 20 25 40 50

$\boxed{\text{Ans : } 21.24.}$

Ex: calculate harmonic mean from the following frequency dist

X :	0-10	10-20	20-30	30-40	40-50
f :	8	15	20	4	3

Ans : HM = 13.96.

Measures of Dispersion:

Dispersion:

"The degree to which numerical data tend to spread about an average value is called variation or dispersion of the data"

The various measures of dispersion are

(i) Range

(ii) Quartile deviation or Semi Inter Quartile range

(iii) Mean deviation or Average deviation

(iv) Standard deviation.

Range :

The Range is the simplest method of studying variation. It is defined as the difference between the value of the smallest observation and the value of the largest observation included in the distribution

(ie) Range $= L - S$

$L =$ Largest Value and

$S =$ Smallest value

and Co-efficient of range is

$$C \cdot R = \frac{L-S}{L+S} \Big\} \text{ Relative measure}$$

Ex: Find the range from the following data

200, 210, 208, 160, 220, 250

Ans: Range = 90

C.R = 0.219.

Ex: Calculate the Co-efficient of range from the following frequency data

x :	10-20	20-30	30-40	40-50	50-60
f :	8	10	12	8	4

Ans: Range = 50

C.R = 0.714.

Scanned by CamScanner

Quartile deviation:

Quartile deviation or Semi-inter Quartile range is defined as

$$Q.D = \frac{Q_3 - Q_1}{2}$$

where Q_1 and Q_3 are the first and Third quartiles of the distribution respetively.

Also

$$\text{Co-efficient of } Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

note:

Quartile is the measure which divides the whole distribution into four equal Parts.

Formula:

I Raw data:

Q_i = size of $i\left(\frac{n+1}{4}\right)^{th}$ observation.

put $i = 1, 2, 3$ and set

Q_1 = size of $\left(\frac{n+1}{4}\right)^{th}$ obsn

Q_2 = size of $\left(\frac{n+1}{2}\right)^{th}$ obsn [equal to median]

Q_3 = size of $3\left(\frac{n+1}{4}\right)^{th}$ obsn / After arranging the data

II Discrete frequency distn:

$$Q_i = \text{size of } i\left(\frac{N+1}{4}\right)^{th} \text{ obsn}$$

put $i = 1, 2, \text{and } 3$, we get

$$Q_1 = \text{size of } \left(\frac{N+1}{4}\right)^{th} \text{ obsn}$$

$$Q_2 = \text{size of } \left(\frac{N+1}{2}\right)^{th} \text{ obsn}$$

$$Q_3 = \text{size of } 3\left(\frac{N+1}{4}\right)^{th} \text{ obsn}$$

III Continuous frequency distn

$$Q_i = l_i + \frac{i\,N/4 - m_i}{f_i} \times c_i$$

Put $i = 1, 2, 3$, we get

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

$$Q_2 = l_2 + \frac{N/2 - m_2}{f_2} \times c_2$$

$$Q_3 = l_3 + \frac{3 \cdot N/4 - m_3}{f_3} \times c_3$$

Ex: Calculate quartile deviation and its Co-efficient of A's monthly earnings for a year.

239, 250, 251, 251, 257, 258, 260, 261, 262, 262, 273, 275

Ans: $Q_1 = 251$, $Q_3 = 262$, $Q.D = 5.5$, Co-eff $= 0.0214$

Ex: Calculate the Semi Interquartile range and Quartile Co-efficient from the following:

Age in yrs:	20	30	40	50	60	70	80
no. of members:	3	61	132	153	140	51	3

Ans: $Q_1 = 40$, $Q_3 = 60$, $Q.D = 10$, Co-eff $Q.D = 0.2$

Ex: Calculate the range and Q.D of wages:

wages (Rs.):	30-32	32-34	34-36	36-38	38-40	40-42	42-44
Labourers:	12	18	16	14	12	8	6

Ans: Range $= 14$, $Q_1 = 33.06$, $Q_3 = 38.25$, $Q.D = 2.85$
Co-effi. $Q.D = 0.08$

Mean deviation: (Average deviation)

Mean deviation is the A.M of the deviations of a series computed from any measure of central tendency (ie) mean, median or mode; all the deviations are taken as positive (ie) + and - signs are ignored.

Formula:

I. Raw data

$$\text{Mean deviation about mean} = \frac{\Sigma |x_i - \bar{x}|}{n}$$

$$\text{mean deviation about median} = \frac{\Sigma |x_i - md|}{n}$$

$$\text{mean deviation about mode} = \frac{\Sigma |x_i - mode|}{n}$$

II For frequency distt (Discrete and Continuous)

$$\text{Mean deviation about mean} = \frac{\Sigma f_i |x_i - \bar{x}|}{N}$$

$$\text{Mean deviation about median} = \frac{\Sigma f_i |x_i - md|}{N}$$

$$\text{Mean deviation about Mode} = \frac{\Sigma f_i |x_i - mode|}{N}$$

and

$$\text{Co-efficient of M.D} = \frac{M.D}{\text{mean (or) median (or) Mode}}$$

Scanned by CamScanner

(or) Average from which it is calculated.

Ex: Calculate mean deviation from mean and median for the following data:

100, 150, 200, 250, 360, 490, 500, 600, 671

Also calculate Co-efficient of Mean deviation.

Ans:

i) MD about mean = 174.44, Co-efft M.D = 0.47

$\because \bar{x} = 369$

ii) MD about median = 173.44 ; co-efft M.D = 0.48.

$\because$ median = 360

Ex: Calculate MD from the following data:

x:	2	4	6	8	10
f:	1	4	6	4	1

Ans: M.D about mean = 1.5, $\bar{x} = 6$.

Ex: Find out the co-efficient of M.D from the following series:

Age in years	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Persons	20	25	32	40	42	35	10	8

Ans:
M.D about mean = 15.1, $\bar{x} = 36.5$

co-efficient of M.D = 0.41,,

Standard deviation:

It is defined as positive square root of the arithmetic mean of the squares of the deviation of the given observation from their arithmetic mean. The S.D is denoted by the Greek letter (Sigma) σ.

Formula:

I Raw data

Direct method

$$\text{Standard deviation} = \sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

Sm

$$S.D(\sigma) = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} \quad ; d = X - A$$

A = Assumed mean

II Discrete frequency distn

Dm

$$S.D(\sigma) = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}}$$

Sm

$$S.D(\sigma) = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2}$$

when

$d = X - A$, N = total frequency

Scanned by CamScanner

III continuous frequency distd

DM

$$S.D(\sigma) = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{N}}$$

SM

$$S.D(\sigma) = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fy}{N}\right)^2}$$

Step deviation method

$$S.D(\sigma) = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \times C$$

Co-efficient of S.D $= \sigma / \bar{x}$

Example: The table below gives the marks obtained by 10 B.com students in statistics examination. Calculate standard deviation.

Nos:	1	2	3	4	5	6	7	8	9	10
Marks:	43	48	65	57	31	60	37	48	78	59

Ans: S.D = 13.26

Ex: calculate standard deviation from the following:

Marks:	10	20	30	40	50	60
No of students:	8	12	20	10	7	3

Ans: SD = 13.45

$\bar{x} = 30.8$

Ex: Compute the Standard deviation and mean deviation from the following.

C.I :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
f :	8	12	17	14	9	7	4

Ans: $\bar{X} = 30.775$, M.D = 13.906, S.D = 16.67

Co-efficient of Variation:

100 times the co-efficient of dispersion based upon standard deviation is called co-efficient of variation (C.V)

(ie) $C.V = 100 \times \frac{\sigma}{\bar{X}}$

Ex:

The prices of 2 Commodities over 10 weeks are given below. Find out which price shows less variation.

A :	54	55	53	56	52	52	58	49	50	51
B :	108	107	105	106	105	103	102	104	104	101

Sln

For A : $\bar{x} = 53$, $\sigma_x = 2.65$, C.V = 5

For B : $\bar{y} = 104.5$, $\sigma_y = 2.06$, CV = 1.97

Since C.v for B is less than that of A, price of B shows less variation.

Note:

For comparing the variability of two series, we calculate the co-efficient of variation for each series. The series having greater C.V is more variable than the other and the series having lesser C.V is said to be more consistent or more homogeneous than the other.

Ex: The number of employes, daily wages per employee and the variance of the wages per employee for two factories are given below:

	Factory A	Factory B
Number of employees	50	100
Average daily wages per employee (Rs)	120	85
Variance of the daily wages per employee (Rs)	9	16

Q) In which factory is there greater variation in the distribution of daily wages per employee?

Ans:

Factory B.

Skewness:

The term 'Skewness' refers to lack of Symmetry or departure from Symmetry. we study skewness to have an idea about the shape of the curve which we can draw with the help of the given data. A distn is said to be skewed if,

(i) mean, median and mode fall at different points,
(ie) mean = median = mode;

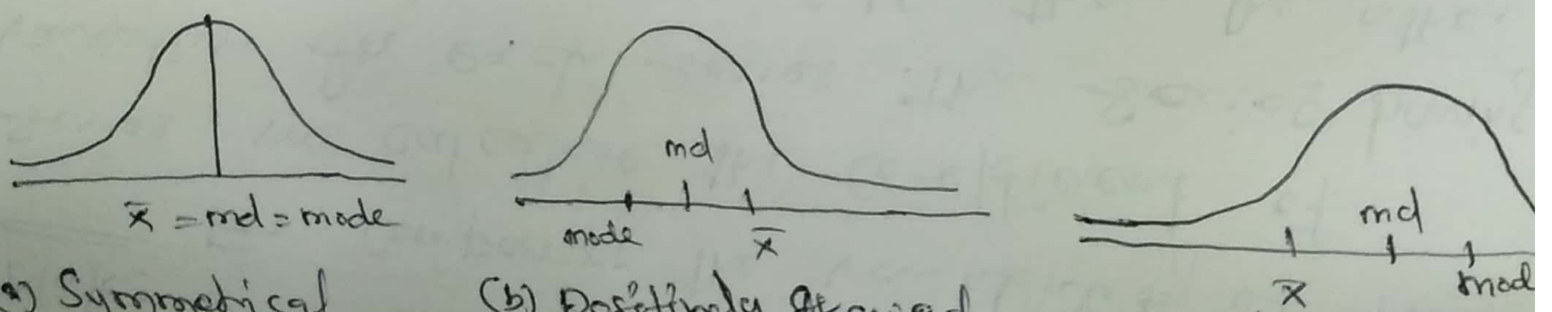
(ii) Quartiles are not equidistant from median and

(iii) The curve drawn with the help of the given data is not symmetrical but stretched more to one side than to the other.

"When a series in not symmetrical it is said to be asymmetrical or skewed".

— Croxton & Cowden.

The following diagrams would clarify the meaning of skewness



(a) Symmetrical distribution
(b) positively skewed distribution
(c) negatively skewed distribution

Difference between measures of dispersion and Skewness

(i) Dispersion deals with the spread of values ~~around~~ around central value. Skewness on the other hand deals with symmetry of distribution of a central value.

(ii) Dispersion deals with the amount of variation and skewness deals with the direction of variation.

Measures of Skewness:

Measures of skewness can be both absolute as well as relative.

(i) S_K = Mean − Median, (ii) S_K = Mean − Mode

(ii) $S_K = (Q_3 - \text{median}) - (\text{median} - Q_1)$.

These are the absolute measures of skewness.

As in dispersion, for comparing two series we do not calculate these absolute measures but we calculate the relative measures called the co-efficient of skewness which are pure numbers independent of units of measurement.

The following are the important ~~measures~~ Co-efficient of Skewness.

1. Karl pearson's Co-efficient of Skewness:

$$S_{kp} = \frac{\text{mean} - \text{mode}}{\sigma}, \quad \sigma = \text{Standard deviation.}$$

2. Bowley's Co-efficient of Skewness:

$$Sk_B = \frac{Q_3 + Q_1 - 2\,\text{median}}{Q_3 - Q_1}$$

3. Based upon moments, co-efficient of Skewness is

$$S_K = \frac{\sqrt{\beta_1} \cdot (\beta_2 + 3)}{2(5\beta_2 - 6\beta_1 - 9)}$$

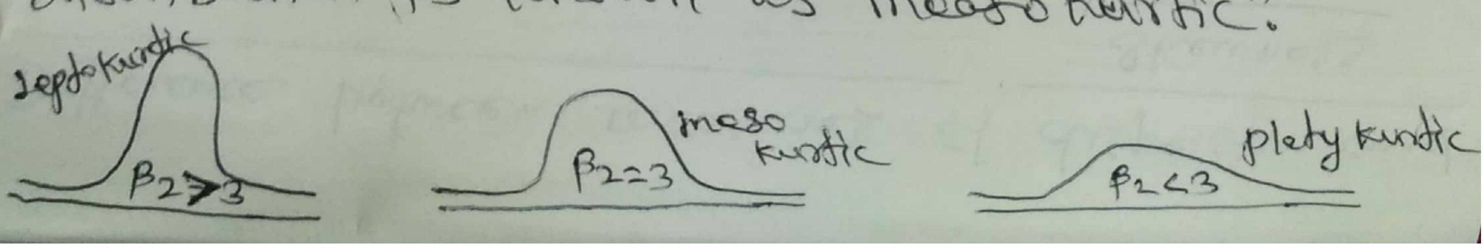
where symbols have their usual meaning.

Kurtosis:

"A measure of Kurtosis Indicates the degree to which a curve of a frequency distribution is peaked or flat topped"
— Croxton and Cowden

If a distribution is more peaked than the normal distribution it is called "Leptokurtic".

If the distn is more flat than the normal distribution is "platykurtic". The normal distribution is known as "mesokurtic".



Kurtosis is measured by the co-efficient β_2.

$$\text{(ie) } \beta_2 = \frac{\mu_4}{\mu_2^2}$$

Moments :

" Moment is a familiar mechanical term for the measure of a force with reference to its tendency to produce rotation. The strength of this tendency depends, obviously, on the amount of the force and the distance from the origin of the point at which the force is exerted".

— Frederic Mills.

Moment is a term generally used in physics, Mechanics and refers to the turning effect or rotating effect of a force. When it is applied in Statistics, it describes the various characteristics of the frequency distribution, Viz.. Central tendency, dispersion, skewness and kurtosis. Moments can be defined as the arithmetic mean of various powers of deviations taken from the mean of a distribution.

The deviations taken from any average, the moments are called raw moments or moments about origin and is denoted by μ_r'

(i.e) $\mu_r' = \frac{1}{n}\sum_{i=1}^{n}(x_i - A)^r$; For raw data

and

$\mu_r' = \frac{1}{N}\sum_{i=1}^{n} f_i (x_i - A)^r$; For frequency distn

If the deviation taken from the A.M, the moments are called central moments, it is denoted by μ_r.

(i.e) $\mu_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$; For raw data

and

$\mu_r = \frac{1}{n}\sum_{i=1}^{n} f_i (x_i - \bar{x})^r$; For frequency distn

<u>Conversion formula</u>:

(i) $\mu_1 = 0$ (always)

(ii) $\mu_2 = \mu_2' - (\mu_1')^2$

(iii) $\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2(\mu_1')^3$

(iv) $\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$

and

$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$; $\beta_2 = \frac{\mu_4}{\mu_2^2}$

Ex: Calculate first four moments from the following data and find out β_1 and β_2.

X:	0	1	2	3	4	5	6	7	8
f:	5	10	15	20	25	20	15	10	5

Sn

$\bar{x} = 4$; $\Sigma d = 0$; $d = x-4$; $\Sigma fd = 0$, $\Sigma fd^2 = 500$

$\Sigma fd^3 = 0$; $\Sigma fd^4 = 4700$

$\mu_1 = 0$, $\mu_2 = 4$, $\mu_3 = 0$, $\mu_4 = 37.6$ $\beta_1 = 0$, $\beta_2 = 2.3$

The value of β_2 is less than 3, hence the curve is platykurtic.

Ex: Calculate the first four moments about the mean from the following data. Also calculate the value of β_1 and β_2.

Marks:	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No. of Students:	5	12	18	40	15	7	3

Sn

$\mu_1' = -1.9$, $d = \frac{x-A}{c} = \frac{x-35}{10}$; $\mu_2' = 181$, $\mu_3' = -970$

$\mu_4' = 98500$,

$\mu_2 = 177.39$, $\mu_3 = 47.982$, $\mu_4 = 95009.364$

$\beta_1 = 0.0004$, $\beta_2 = 3.02$.

PROBLEMS:

1. The frequency distribution of weight in grams of mangoes of a given variety is given below. Calculate the A.M, median and mode.

weight (in grams)	:	410-419	420-429	430-439	440-449	450-459	460-469	470-479
No. of Mangoes	:	14	20	42	54	45	18	7

Ans: mean = 443.4, median = 443.94, mode = 445.21

2. Lives of two models of refrigerators turned in for new models in a recent survey are:

Life (no. of years)	:	0-2	2-4	4-6	6-8	8-10	10-12
no. of refrigerators model A	:	5	16	13	7	5	4
Model B	:	2	7	12	19	9	1

What is the average life of each model of these refrigerators? Which model has more uniformity?

[Ans: $\bar{x}_A = 5.12$, $\bar{x}_B = 6.6$, $CV_A = 54.92\%$. $CV_B = 36.2\%$.
B model has greater Uniformity].

PROBLEMS: 1. The frequency distribution of weight in grams of mangoes of a given variety is given below. Calculate the A.M, median and mode.

weight (in gms)	410-419	420-429	430-439	440-449	450-459	460-469	470-479
No. of Mangoes	14	20	42	54	45	18	7

Ans: mean = 443.4, median = 443.94, mode = 445.21

2. Lives of two models of refrigerators turned in for new models in a recent survey are:

Life (no. of years)		0-2	2-4	4-6	6-8	8-10	10-12
no. of refrigerators	model A	5	16	13	7	5	4
	Model B	2	7	12	19	9	1

What is the average life of each model of these refrigerators? Which model has more uniformity?

Ans: $\bar{x}_A = 5.12$, $\bar{x}_B = 6.6$, $CV_A = 54.92\%$. $CV_B = 36.2\%$.
B model has greater Uniformity].

Combined Arithmetic Mean:

If we know the means and the number of items in two or more related groups, the combined or composite mean can be computed with the help of the following formula

$$\bar{X}_{12} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2}{n_1+n_2}$$

$$\bar{X}_{123} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1+n_2+n_3}$$

$\bar{x}_{12}, \bar{x}_{123}$ = The combined means

$\bar{x}_1\ \bar{x}_2\ \bar{x}_3$ = Arithmetic mean of first group, second group and third group.

$n_1\ n_2\ n_3$ = Number of items in first, second and third groups.

PROBLEM:

There are two branches of a company, employing 100 and 80 persons respectively. If the arithmetic mean of the monthly salaries paid by the two companies are Rs.275 and Rs.225 respectively, find the A.M of the salaries of the employees of the companies as a whole.

Sol: w.k.T $\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2} = \frac{(100\times 225)+(80\times 225)}{100+80}$

$= 252.78$