Subject: **Introduction to Statistics**

BS Mathematics Morning/Evening program spring semester 2020

Chapter # 04 Measures of Dispersion, Moments, Skewness and Kurtosis

Measures of Dispersion: Sometimes when two or more different data sets are to be compared using measure of central tendency or averages, we get the same result.

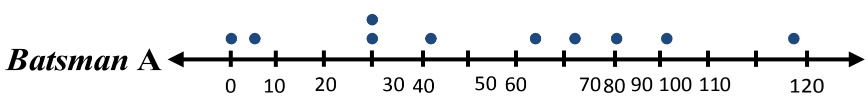
Consider the runs scored by two batsmen in their last ten matches as follows:

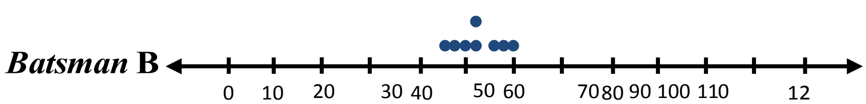
Batsman A: 30, 91, 0, 64, 42, 80, 30, 5, 117, 71
Batsman B: 53, 46, 48, 50, 53, 53, 58, 60, 57, 52

Clearly, mean of the runs scored by both the batsmen A and B is same i.e. 53

Can we say that the performance of two players is same? Clearly No, because the variability in the scores of batsman A is from 0 to 117, whereas, the variability of the runs scored by batsman B is from 46 to 60.

Let us now plot the above scores as dots on a number line. We find the following diagrams:





We can see that the dots corresponding to batsman B are close to each other and is clustering around the measure of central tendency (mean), while those corresponding to batsman A are scattered or more spread out. Thus, the measures of central tendency are not sufficient to give complete information about a given data. In such a situation the comparison becomes very difficult. We therefore, need some additional information for comparison, concerning with, how the data is dispersed about (more spread out) the average. This can be done by measuring the dispersion. Like „measures of central tendency" we want to have a single number to describe variability. This single number is called a **'measure of dispersion'**.

Dispersion: "The variability (spread) that exists between the value of a data is called dispersion"

OR

"The extent to which the observations are spread around an average is called dispersion or scatter ".

As we know that, there are quite a few ways of measuring the central tendency of a data set i.e. A.M, G.M, H.M, Mode and Median. Similarly, we have different ways of measuring and comparing the dispersion of the distribution(s). There are two important types of measures of dispersion.

❖ **Types of Measures of Dispersion:**

➢ **There are two types of measure of dispersion**

I) Absolute Measure of Dispersion

II) Relative Measure of Dispersion

➢ **Absolute Measure of Dispersion: "An absolute measure of dispersion measures the variability in terms of the same units of the data"** e.g. if the units of the data are Rs, meters, kg, etc. The units of the measures of dispersion will also be Rs, meters, kg, etc.

The common absolute measures of dispersion are:

- Range
- Quartile Deviation or Semi Inter-Quartile Range
- Average Deviation or Mean Deviation
- Standard Deviation

➢ **Relative Measure of Dispersion: "A relative measure of dispersion compares the variability of two or more data that are independent of the units of measurement"**

➢ In other word **"A relative measure of dispersion, expresses the absolute measure of dispersion relative to the relevant average and multiplied by 100 many times" i.e.**

$$Relative\ Dispersion = \frac{Absolute\ Dispersion}{Average}$$

$$Relative\ Dispersion = \frac{Absolute\ Dispersion}{Average} \times 100$$

This is a pure number and independent of the units in which the data has been expressed. It is used for the purpose to compare the dispersion of a data with the dispersion of another data.

The common relative measures of dispersion are:

- Coefficient of Dispersion or Coefficient of Range
- Coefficient of Quartile Deviation
- Coefficient of Mean Deviation
- Coefficient of Standard Deviation or Coefficient of Variation (C.V)

The major difference b/w Absolute and Relative Measures of Dispersion is that the Absolute measure of dispersion measures only the variability of the data, further it has the unit of measurement; on the other hand Relative measure of dispersion is used to compare the variation of two or more distributions, further it is unit less.

- **Range: "The difference between the largest and the smallest value in a set of data is called range"**

OR

"In continuous grouped data the difference between the upper class boundary of the highest class and lower class boundary of the lowest class is called range"

Formula: $R = X_m - X_0$ OR R=Largest value –smallest value

Where R is the range, X_m is the largest value and X_0 is the smallest value.

- **Coefficient of Range or Coefficient of Dispersion:** The coefficient of range or coefficient of dispersion is a relative measure of dispersion and is given by:

$$Coefficient\ of\ Range = \frac{X_m - X_0}{X_m + X_0}$$

- **Numerical example of Range and Coefficient of range**

Ex # The marks obtained by 9 students are given below:

x_i	45	32	37	46	39	36	41	48	36

$$R = X_m - X_0$$

$$X_m = 48, \quad X_0 = 32$$

$$R = 48 - 32 \quad \Rightarrow R = 16 \text{ marks}$$

- Coefficient of Range

$$C.o.f = \frac{X_m - X_0}{X_m + X_0}$$

$$C.o.f = \frac{48-32}{48+32} \quad \Rightarrow C.o.f = \frac{16}{80} \quad \Rightarrow C.o.f = 0.2$$

- **Quartile Deviation or Semi-inter-quartile Range:** "half of the difference between the upper quartile and lower quartile is called the semi-inter quartile range or quartile deviation"i.e.

$$Quartile\ deivation = \frac{Q_3 - Q_1}{2}$$

- **Coefficient of Quartile Deviation:** The coefficient of quartile deviation is a relative measure of dispersion and is given by:

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

- **Numerical example of quartile deviation and coefficient of quartile deviation**

Ex # calculate quartile deviation and coefficient of quartile deviation for ungrouped data.
The marks obtained by 9 students are given below

x_i	45	32	37	46	39	36	41	48	36

$$Quartile\ deviation = \frac{Q_3 - Q_1}{2}$$

"n=9" is odd then we use odd case formulae

$$Q_1 = Marks\ obtained\ by\ \left[\left(\frac{n}{4}\right)+1\right]^{th} student \qquad Q_1 = 36 \text{ marks}, \quad Q_3 = 45$$

$$Quartile\ deviation = \frac{45-36}{2} \quad = 4.5 \text{ marks} \qquad \text{(Answer)}.$$

- Coefficient of Quartile deviation:

$$Cofficient\ of\ Q.D = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$Cofficient\ of\ Q.D = \frac{45-32}{45+32} \quad = 0.11 \qquad \text{(Answer).}$$

Ex # calculate quartile deviation and coefficient of quartile deviation for continuous grouped data.

Class boundaries	Midpoints (x_i)	Frequency (f_i)	Cumulative frequency $(c.f)$
29.5---39.5	34.5	8	8
39.5---49.5	44.5	87	95
49.5---59.5	54.5	190	285
59.5---69.5	64.5	304	589
69.5---79.5	74.5	211	800
79.5---89.5	84.5	85	885
89.5---99.5	94.5	20	905
		$\sum_{i=1}^{n} f_i$ =905	

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - C\right) \quad \Rightarrow Q_1 = 56.40 \text{ marks}$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - C\right) \quad \Rightarrow Q_3 = 73.76 \text{ marks}$$

Calculate Quartile deviation and coefficient of quartile deviation from above results?

- **Mean Absolute Deviation or Mean Deviation (Average Deviation):** *"The arithmetic mean of the absolute deviation from an average (mean, median etc.) is called mean deviation or average deviation"*

	Ungrouped Data	Grouped Data
M.D from Mean	$M.D = \frac{\sum \lvert x_i - \bar{x} \rvert}{n}$	$M.D = \frac{\sum f \lvert x_i - \bar{x} \rvert}{n}$
M.D from Median	$M.D = \frac{\sum \lvert x_i - Med \rvert}{n}$	$M.D = \frac{\sum f \lvert x_i - Med \rvert}{n}$

Coefficient of Mean Deviation: The coefficient of mean deviation is a relative measure of dispersion and is given by:

$$Coefficient\ of\ M.D = \frac{M.D(from\ mean)}{Mean}$$

$$Coefficient\ of\ M.D = \frac{M.D(from\ median)}{Median}$$

- **Numerical example of Mean deviation and Coefficient of Mean deviation for both ungrouped and grouped data.**

- Calculate the mean deviation and coefficient of mean deviation from (i) the mean, (ii) the median , in the ungrouped data case, of the following set of examination marks:

x_i	45 32 37 46 39 36 41 48 36

n= 9

x_i	$x_i - \overline{x}$, $\overline{x} = 40$	$\lvert x_i - \overline{x}\rvert$	$\lvert x_i - median\rvert$, $median = 39$
32	-8	8	7
36	-4	4	3
36	-4	4	3
37	-3	3	2
39	-1	1	0
41	1	1	2
45	5	5	6
46	6	6	7
48	8	8	9
$\sum x_i$ =360		$\sum \lvert x_i - \overline{x}\rvert = 40$	$\sum \lvert x_i - median\rvert$ =39

$\overline{x} = \frac{\sum x_i}{n} = \frac{360}{9}$ =40 marks ,

$Median = marks\ obtained\ by \left(\frac{n+1}{2}\right)^{th} student$ $\quad Median = 39$ marks

- Mean deviation from mean is given by

$$M.D = \frac{\sum \lvert x_i - \overline{x}\rvert}{n} \Rightarrow M.D = \frac{40}{9} \Rightarrow M.D = 4.4 \text{ marks} \quad \text{(Answer).}$$

- Mean deviation from median is given by

$$M.D = \frac{\sum \lvert x_i - median\rvert}{n} \Rightarrow M.D = \frac{39}{9} \Rightarrow M.D = 4.3 \text{ marks} \quad \text{(Answer).}$$

➢ Calculate coefficient of mean deviation for both mean and median.

$$C.o.M.D = \frac{M.D}{mean\ (\overline{x})} \Rightarrow C.o.M.D = \frac{4.4}{40} \Rightarrow C.o.M.D = 0.11 \quad \text{(Answer)}.$$

$$C.o.M.D = \frac{M.D}{median} \Rightarrow C.o.M.D = \frac{4.3}{39} \Rightarrow C.o.M.D = 0.11 \quad \text{(Answer)}.$$

➢ Calculate mean deviation and coefficient of mean deviation from mean in continuous grouped case, showing the weights of 60 apples.

Weights (grams)	65--84	85--104	105--124	125--144	145--164	165--184	185--204
Frequency	09	10	17	10	05	04	05

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	$f_i x_i$	$\lvert x_i - \overline{x} \rvert$	$f_i \lvert x_i - \overline{x} \rvert$
65----84	74.5	09	670.5	-48.0	432.0
85----104	94.5	10	945.0	-28.0	280.0
105----124	114.5	17	1946.5	-8.0	136.0
125----144	134.5	10	1345.0		120.0
145----164	154.5	05	772.5	12.0	160.0
165----184	174.5	04	698.0		208.0
185----204	194.5	05	972.5	32.0	360.0
				52.0	
				72.0	
		$\sum_{i=1}^{n} f_i = 60$	$\sum_{i=1}^{n} f_i x_i =$ 7350.0		$\sum f_i \lvert x_i - \overline{x} \rvert$ = 1696.0

$$\overline{x} = \frac{\sum f_i x_i}{\sum f_i} \Rightarrow \overline{x} = \frac{7450.0}{60} \Rightarrow \overline{x} = 122.5 \text{ grams}$$

$$M.D = \frac{\sum f_i \lvert x_i - \overline{x} \rvert}{\sum f_i} \Rightarrow M.D = \frac{1696.0}{60} \Rightarrow M.D = 28.27 \text{ grams} \quad \text{(Answer)}.$$

Standard Deviation: *"*The positive square root of variance is called as standard deviation".

OR

"The positive square root of the arithmetic mean of the squared deviations from the mean is called the standard deviation"

	Ungrouped Data	Grouped Data
S.D for Population	$\sigma = \sqrt{\frac{\sum (xi-\mu)^2}{N}}$	$\sigma = \sqrt{\frac{\sum f(xi-\mu)^2}{N}}$
S.D for Sample	$S = \sqrt{\frac{\sum (xi-\bar{x})^2}{n}}$	$S = \sqrt{\frac{\sum f(xi-\bar{x})^2}{n}}$

❖ **Methods of Calculating Variance and Standard Deviation.**

Methods	Ungrouped Data	
	Variance	Standard Deviation
Direct Method	$S^2 = \frac{\sum xi^2}{n} - \left(\frac{\sum xi}{n}\right)^2$	$S = \sqrt{\frac{\sum xi^2}{n} - \left(\frac{\sum xi}{n}\right)^2}$
Short cut Method	$S^2 = \frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2$	$S = \sqrt{\frac{\sum D^2}{n} - \left(\frac{\sum D}{n}\right)^2}$
Step-deviation Method	$S^2 = h^2\left[\frac{\sum ui^2}{n} - \left(\frac{\sum ui}{n}\right)^2\right]$	$S = h\sqrt{\frac{\sum ui^2}{n} - \left(\frac{\sum ui}{n}\right)^2}$
Methods	**Grouped Data**	
	Variance	**Standard Deviation**
Direct Method	$S^2 = \frac{\sum fxi^2}{n} - \left(\frac{\sum fxi}{n}\right)^2$	$S = \sqrt{\frac{\sum fxi^2}{n} - \left(\frac{\sum fxi}{n}\right)^2}$
Short cut Method	$S^2 = \frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2$	$S = \sqrt{\frac{\sum fD^2}{n} - \left(\frac{\sum fD}{n}\right)^2}$
Step-deviation Method	$S^2 = h^2\left[\frac{\sum fui^2}{n} - \left(\frac{\sum fui}{n}\right)^2\right]$	$S = h\sqrt{\frac{\sum fui^2}{n} - \left(\frac{\sum fui}{n}\right)^2}$

Coefficient of Standard Deviation OR Coefficient of Variation: The coefficient of standard deviation is a relative measure of dispersion and is given by:

$$Coefficient\ of\ S.D = \frac{Standard\ Deviation}{Mean}$$

The coefficient of standard deviation is also called the coefficient of variation, denoted by C.V and is given by:

$$C.V = \frac{Standard\ Deviation}{Mean} \times 100$$

Coefficient of Variation was introduced by **Karl Pearson**. It is used to compare the variation or to compare the performance of two sets of data. A large value of C.V indicates that there is greater variability and vice versa. Similarly, the smaller the C.V the more consistent is the performance and vice versa.

- ❖ Numerical example of Standard deviation (S.D),Variance and Coefficient of Variation (C.V) in case of ungrouped data, using direct method:

Ex # Calculate the variance, S.D and C.V from the following marks obtained by 9 students.

x_i	45 32 37 46 39 36 41 48 36

x_i	$x_i - \overline{x}$	$\left(x_i - \overline{x}\right)^2$	x_i^2
45	5	25	2025
32	-8	64	1024
37	-3	09	1369
46	6	36	2116
39	-1	01	1521
36	-4	16	1296
41	1	01	1681
48	8	64	2304
36	-4	16	1296
$\sum x_i$ =360		$\sum\left(x_i - \overline{x}\right)^2$ =232	$\sum x_i^2$ =14632

$\overline{x} = \dfrac{\sum x_i}{n} \Rightarrow \overline{x} = \dfrac{360}{9} \Rightarrow \overline{x}$ =40 marks

Variance or $S^2 = \dfrac{\sum\left(x_i - \overline{x}\right)^2}{n} \Rightarrow S^2 = \dfrac{232}{9} \Rightarrow S^2$=25.78 $(marks)^2$

Standard deviation or S.D or $S = \sqrt{\dfrac{\sum\left(x_i - \overline{x}\right)^2}{n}} \Rightarrow S = \sqrt{25.78} \Rightarrow S$ =5.08 marks

(Answer).

Using the alternative method.

Calculate Variance:

$$S^2 = \frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2 \quad \Rightarrow S^2 = \frac{14632}{9} - \left(\frac{360}{9}\right)^2 \quad \Rightarrow S^2 \; = 1625.78 \; - \; 1600$$

$\Rightarrow S^2 = 25.78 \; (marks)^2$ (Answer).

Calculate Standard deviation:

$$S = \sqrt{\frac{\sum x_i^2}{n} - \left(\frac{\sum x_i}{n}\right)^2} \quad \Rightarrow S = \sqrt{\frac{14632}{9} - \left(\frac{360}{9}\right)^2} \quad \Rightarrow S = \sqrt{25.78} \quad \Rightarrow S = 5.08 \text{ marks}$$

(Answer).

- Calculate Coefficient of Variation (C.V):

$$C.V = \frac{S.D}{\bar{x}} \times 100 \quad \Rightarrow C.V = \frac{5.08}{40} \times 100 \quad \Rightarrow C.V = 12.70 \qquad \text{(Answer).}$$

- **Calculate the Variance, Standard deviation and Coefficient of Variation from the following weight of 60 apples in Continuous grouped data:**

Weight (grams)	Midpoints (x_i)	Frequency (f_i)	$f_i x_i$	$f_i x_i^2$
65----84	$(65+84)/2 = 74.5$	09	9×74.5=670.	49 952.25
85----104	94.5	10	5	89 302.50
105----124	114.5	17	945.0	222 874.25
125----144	134.5	10	1946.5	180 902.50
145----164	154.5	05	1345.0	119 351.25
165----184	174.5	04	772.5	121 801.00
185----204	194.5	05	698.0	189 151.25
			972.5	
		$\sum_{i=1}^{n} f_i = 60$	$\sum_{i=1}^{n} f_i x_i = 735$ 0.0	$\sum f_i x_i^2 = 973\ 335.00$

Calculate Variance:

$$S^2 = \frac{\sum f_i x_i^2}{\sum f_i} - \left(\frac{\sum f_i x_i}{\sum f_i}\right)^2 \qquad \Rightarrow S^2 = \frac{973335.00}{60} - \left(\frac{7350.0}{60}\right)^2$$

$\Rightarrow S^2 = 16222.25 - 15006.25$

$\Rightarrow S^2 = 1216 \; (grams)^2$ (Answer).

Calculate Standard deviation:

$$S=\sqrt{\frac{\sum f_i x_i^2}{\sum f_i}-\left(\frac{\sum f_i x_i}{\sum f_i}\right)^2} \quad \Rightarrow S=\sqrt{1216} \quad \Rightarrow S = 34.87 \text{ grams} \qquad \text{(Answer).}$$

Calculate Coefficient of Variation:

$$C.V=\frac{S.D}{\bar{x}}\times 100 \quad \Rightarrow C.V=\frac{34.87}{122.5}\times 100 \quad \Rightarrow C.V = 28.46 \qquad \text{(Answer).}$$

Question # Calculate Variance, Standard deviation and Coefficient of Variation using direct, shortcut and step deviation method for Continuous grouped data, the data are given below:

Income	35--39	40--44	45--49	50--54	55--59	60--64	65--69
Frequency	13	15	17	28	12	10	05

Properties of Variance

- Variance of a constant is zero i.e. Var(c) = 0, where “c” is any constant.

  Proof: By definition of variance:

$$Var(x) = \frac{\sum(x-\bar{x})^2}{n}$$

$$\Rightarrow Var(c) = \frac{\sum(c-c)^2}{n} = 0 \qquad (\text{if } x \to c \Rightarrow \bar{x} \to c)$$

- The variance is unaffected by the change of origin i.e. when a constant is added to or subtracted from each value of a variable, the variance remains unchanged i.e. $Var(x \pm c) = Var(x)$

  Proof: By definition of variance:

$$Var(x) = \frac{\sum(x-\bar{x})^2}{n}$$

$$\Rightarrow Var(x-c) = \frac{\sum[(x-c)-(\bar{x}-c)]^2}{n} \qquad (\text{if } x \to x-c \Rightarrow \bar{x} \to \bar{x}-c)$$

$$= \frac{\sum(x-c-\bar{x}+c)^2}{n}$$

$$= \frac{\sum(x-\bar{x})^2}{n}$$

$$= Var(x)$$

Similarly, $Var\ (x+c) = Var(x)$

- Variance is affected by the change of scale i.e. when each observation of a variable is multiplied or divided by a constant, then variance is multiplied or divided by square of that constant i.e.

$$Var(cx) = c^2 Var(x) \quad \text{OR} \quad Var\left(\frac{x}{c}\right) = \frac{1}{c^2} Var(x)$$

Proof: By definition of variance:

$$Var(x) = \frac{\sum (x-\bar{x})^2}{n}$$

$$\Rightarrow Var(cx) = \frac{\sum \left[(cx)-(c\bar{x})\right]^2}{n} \qquad (\text{if } x \to cx \Rightarrow \bar{x} \to c\bar{x})$$

$$= \frac{\sum (cx - c\bar{x})^2}{n}$$

$$= \frac{\sum c^2 (x-\bar{x})^2}{n}$$

$$= c^2 \frac{\sum (x-\bar{x})^2}{n}$$

$$= c^2 Var(x)$$

Similarly $$Var\left(\frac{x}{c}\right) = \frac{1}{c^2} Var(x)$$

- The variance of the sum or difference of two independent variables is equal to the sum of their respective variances i.e. $Var(x \pm y) = Var(x) + Var(y)$

Proof: By definition of variance:

$$Var(x) = \frac{\sum (x-\bar{x})^2}{n} \quad \text{And} \quad Var(y) = \frac{\sum (y-\bar{y})^2}{n}$$

$$\Rightarrow Var(x-y) = \frac{\sum \left[(x-y)-(\bar{x}-\bar{y})\right]^2}{n}$$

$$= \frac{\sum (x - y - \bar{x} + \bar{y})^2}{n} = \frac{\sum \left[(x-\bar{x}) - (y-\bar{y})\right]^2}{n}$$

$$= \frac{\sum \left[(x-\bar{x})^2 + (y-\bar{y})^2 - 2(x-\bar{x})(y-\bar{y})\right]}{n}$$

$$= \frac{\sum (x-\bar{x})^2}{n} + \frac{\sum (y-\bar{y})^2}{n} - \frac{2\sum (x-\bar{x})(y-\bar{y})}{n}$$

Where the term $\frac{\sum (x-\bar{x})(y-\bar{y})}{n}$ is called the covariance and we know that for independent variables covariance is zero.

$$\therefore \quad Var(x-y) = \frac{\sum (x-\bar{x})^2}{n} + \frac{\sum (y-\bar{y})^2}{n}$$

$$\Rightarrow \quad Var(x-y) = Var(x) + Var(y)$$

Similarly, $$Var(x-y) = Var(x) + Var(y)$$

❖ **Moments: "The arithmetic mean of the r^{th} power of deviations taken either from mean, zero or from any arbitrary origin (provisional ,mean) are called moments".**

- When the deviations are computed from the arithmetic mean, then such moments are called moments about mean (mean moments) or sometimes called **central moments**, denoted by m_r and given as follows:

	Ungrouped Data	**Grouped Data**
For Sample	$m_r = \frac{\sum (x_i - \bar{x})^r}{n}$	$m_r = \frac{\sum f(x_i - \bar{x})^r}{n}$
For Population	$\mu_r = \frac{\sum (x_i - \mu)^r}{N}$	$\mu_r = \frac{\sum f(x_i - \mu)^r}{N}$
Where r = 1,2,3,4…		

- When the deviations of the values are computed from origin or zero, then such moments ar called the **moments about origin**, denoted by m'_r and are given by:

	Ungrouped Data	**Grouped Data**
For Sample	$m'_r = \frac{\sum x_i^r}{n}$	$m'_r = \frac{\sum f x_i^r}{n}$
For Population	$\mu'_r = \frac{\sum x_i^r}{N}$	$\mu'_r = \frac{\sum f x_i^r}{N}$
Where r = 1,2,3,4…		

- When the deviations of the values are computed from any arbitrary value say A (provisional mean), then such moments are called moments about provisional mean, denoted by m'_r or μ'_r. There are two methods for finding moments about provisional mean: i.e. Short-cut Method and Step-deviation Method.

	Ungrouped Data		Grouped Data	
	Short cut Method	Step-deviation Method	Short cut Method	Step-deviation Method
For Sample	$m'_r = \frac{\sum D^r}{n}$	$m'_r = h^r \left(\frac{\sum ui^r}{n}\right)$	$m'_r = \frac{\sum fD^r}{n}$	$m'_r = h^r \left(\frac{\sum fui^r}{n}\right)$
For Population	$\mu'_r = \frac{\sum D^r}{N}$	$\mu'_r = h^r \left(\frac{\sum ui^r}{N}\right)$	$\mu'_r = \frac{\sum fD^r}{N}$	$\mu'_r = h^r \left(\frac{\sum fui^r}{N}\right)$
Where r = 1,2,3,4… D= x – A and where h is the common width of the class intervals, $ui = \frac{xi - A}{h}$				

All the raw moments can then be converted into central moments or mean moments or moments about mean, by using the following relations:

For Sample:

$$m_1 = 0$$
$$m_2 = m'_2 - (m'_1)^2$$
$$m_3 = m'_3 - 3m'_1 m'_2 + 2(m'_1)^3$$
$$m_4 = m'_4 - 4m'_1 m'_3 + 6(m'_1)^2 m'_2 - 3(m'_1)^4$$

For Population:

$$\mu_1 = 0$$
$$\mu_2 = \mu'_2 - (\mu'_1)^2$$
$$\mu_3 = \mu'_3 - 3\mu'_1 \mu'_2 + 2(\mu'_1)^3$$
$$\mu_4 = \mu'_4 - 4\mu'_1 \mu'_3 + 6(\mu'_1)^2 \mu'_2 - 3(\mu'_1)^4$$

- Numerical example of first Moments:
- Calculate first four moments about mean for ungrouped data for the following set of examination marks:

General formula for moment about mean are given below:

$m_r = \frac{\sum (x_i - \bar{x})^r}{n}$, where *r=1, 2, 3, 4.*

Put r=1 $m_1 = \frac{\sum (x_i - \bar{x})}{n}$, put r=2 $m_2 = \frac{\sum (x_i - \bar{x})^2}{n}$, put r=3

$m_3 = \frac{\sum (x_i - \bar{x})^3}{n}$ and put r=4

$m_4 = \frac{\sum (x_i - \bar{x})^4}{n}$

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^4$
32	-8	64	-512	4096
36	-4	16	-64	256
36	-4	16	-64	256
37	-3	09	-27	81
39	-1	01	-1	1
41	1	01	1	1
45	5	25	125	625
46	6	36	216	1296
48	8	64	512	4096
$\sum x_i = 360$	$\sum(x_i - \bar{x}) = 0$	$\sum(x_i - \bar{x})^2 = 232$	$\sum(x_i - \bar{x})^3 = 186$	$\sum(x_i - \bar{x})^4 = 10708$

$m_1 = 0$, $m_2 = 25.78$ (marks)^2, $m_3 = 20.67$ (marks)^3 , $m_4 = 1189.78$ (marks)^4 (Answer).

❖ Question # Calculate first four moments about mean for grouped data (using a continuous grouped case formula). The following distribution relates to the number of assistants in 50 retail establishments, the data are given below:

No. of assistants	0	1	2	3	4	5	6	7	8	9
f	3	4	6	7	10	6	5	5	3	1

Using these formulae

$$m_1 = \frac{\sum f_i (x_i - \bar{x})}{\sum f_i}, \quad m_2 = \frac{\sum f_i (x_i - \bar{x})^2}{\sum f_i}, \quad m_3 = \frac{\sum f_i (x_i - \bar{x})^3}{\sum f_i} \quad \text{and}$$

$$m_4 = \frac{\sum f_i (x_i - \bar{x})^4}{\sum f_i}$$

❖ Numerical example of Moment in continuous grouped data:

❖ Compute the first four moments and measure of Skewness and Kurtosis for the following distribution of wages using a short cut method:

Weekly earnings (Rupees)	5	6	7	8	9	10	11	12	13	1 4	15
No. of men	1	2	5	10	20	51	22	11	5	3	1

Earnings in Rs. (x_i)	Men f_i	$D_i = (x_i - A)$ $A = 10$	$f_i D_i$	$f_i D_i^2$	$f_i D_i^3$	$f_i D_i^4$
5	1	-5	-5	25	-125	625
6	2	-4	-8	32	-128	512
7	5	-3	-15	45	-135	405
8	10	-2	-20	40	-80	160
9	20	-1	-20	20	-20	20
10	51	0	0	0	0	0
11	22	1	22	22	22	22
12	11	2	22	44	88	176
13	5	3	15	45	135	405
14	3	4	12	48	192	768
15	1	5	5	25	125	625
	$\sum f_i = 131$		$\sum f_i D_i = 8$	$\sum f_i D_i^2 = 346$	$\sum f_i D_i^3 = 74$	$\sum f_i D_i^4 = 3718$

$m_1' = \dfrac{\sum f_i D_i}{\sum f_i} \Rightarrow m_1' = \dfrac{8}{131} \Rightarrow m_1' = 0.06,$ $\quad m_2' = \dfrac{\sum f_i D_i^2}{\sum f_i} \Rightarrow m_2' = \dfrac{346}{131} \Rightarrow m_2' =$ 2.64

$m_3' = \dfrac{\sum f_i D_i^3}{\sum f_i} \Rightarrow m_3' = \dfrac{74}{131} \Rightarrow m_3' = 0.56,$ $\quad m_4' = \dfrac{\sum f_i D_i^4}{\sum f_i} \Rightarrow m_4' = \dfrac{3718}{131}$ $\Rightarrow m_4' = 28.38$

$m_1 = m_1' - m_1' = 0$ (always zero), $\quad m_2 = m_2' - (m_1')^2 \Rightarrow m_2 = 2.64 - (0.06)^2$ $\Rightarrow m_2 = 2.64;$

$m_3 = m_3' - 3m_2'm_1' + 2(m_1')^3 \quad \Rightarrow m_3 = 0.56 - 3(2.64)(0.06) + 2(0.06)^3 \quad \Rightarrow m_3 = 0.08;$

$m_4 = m_4' - 4m_3'm_1' + 6m_2'(m_1')^2 - 3(m')^4$

$\Rightarrow m_4 = 28.38 - 4(0.56)(0.06) + 6(2.64)(0.06)^2 - 3(0.06)^4$

$\Rightarrow m_4 = 28.30$ (Answer).

- Calculate measure of Skewness:

$$b_1 = \frac{m_3}{m_2^3} \Rightarrow b_1 = \frac{0.08}{(2.64)^2} \Rightarrow b_1 = 0.0114 \quad \text{(Answer).}$$

- Calculate measure of Kurtosis:

$$b_2 = \frac{m_4}{m_2^2} \Rightarrow b_2 = \frac{28.30}{(2.64)^2} \Rightarrow b_2 = 4.0604 \quad \text{(Answer).}$$

Question # Calculate first four moment about the mean and Measure of Skewness and Kurtosis for the following data are given below,

Using a step deviation method for grouped data?

Age nearest birth day	22	27	32	37	42	47	52
No. of men	1	2	26	22	20	15	14

Symmetrical Distribution

- A distribution in which the values of mean, median and mode are equal is called symmetrical distribution i.e.

$$Mean = Median = Mode$$

- A distribution is which the two quartiles are equidistant from the median is called a symmetrical distribution i.e.

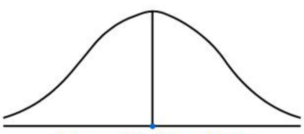
$$Q_3 + Q_1 - 2\ Median = 0$$

- A distribution is said to be symmetrical if:

$$b_1 = 0$$

- A distribution in which the two tails are equal in length from the central value then it is called symmetrical distribution. The symmetrical distribution is always in the form of a bell.

Moment-Ratios	1st Moments Ratio	2nd Moments Ratio
Sample	$b_1 = \frac{m_3^2}{m_2^3}$	$b_2 = \frac{m_4}{m_2^2}$
Population	$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$	$\beta_2 = \frac{\mu_4}{\mu_2^2}$
Moment-Ratios are independent of the origin and units of measurements i.e. they are dimensionless quantities.		



Skewness: We know that for symmetrical distribution the values of mean, median and mode are equal and that the two tails of the distribution are equal in length from the central value etc.

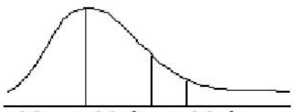
"Skewness is the degree of asymmetry" OR

"Skewness is the lack (absence) of symmetry around central value (average)"

The presence skewness tells us that a particular distribution is not symmetrical or in other words it is skewed. In skewed distribution the curve is turned more to one side than the other.

Positive Skewness

- Skewness is said to be positive, if mean is greater than the median and median is greater than mode i.e. *Mean > Median > Mode*
- Skewness is said to be positive, if: $Q_3 + Q_1 - 2\ Median > 0$
- In terms of moments, skewness is said to be positive if: $\alpha_3 > 0$
- Skewness is said to be positive, if the right tail of a distribution is longer than its left tail.

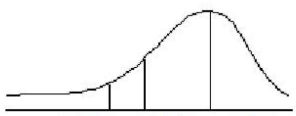


Mean > Median > Mode

$$\alpha_3 = \sqrt{b_1} = \frac{m_3}{\sqrt{m_2^3}}$$

Negative Skewness

- Skewness is said to be negative, if mean is smaller than the median and median is smaller than mode i.e. *Mean > Median > Mode*
- Skewness is said to be negative, if: $Q_3 + Q_1 - 2\ Median > 0$
- In terms of moments, skewness is said to be negative if: $\alpha_3 < 0$
- Skewness is said to be negative, if the left tail of a distribution is longer than its right tail.



Mean < Median < Mode

Karl Pearson's measures of Skewness: It is defined as:

$$S_k = \frac{Mean - Mode}{Standard\ Deviation}$$

It is to be noted that, this measure is suggested by Karl Pearson (1857-1936) and is known as Pearsonian coefficient.

Since in many cases mode is ill-defined, therefore we replace (Mean – Mode) by its equivalent from the empirical relation i.e. 3 (Mean – Median) and hence:

$$S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$$

This coefficient usually varies between –3 and +3.

Bowley's measures of Skewness: It is defined as:

$$S_k = \frac{Q_3 + Q_1 - 2Median}{Q_3 - Q_1}$$

It is also to be noted that, this measure is suggested by Bowley (1869-1957) and is known as Bowley's coefficient.

This coefficient usually varies between –1 and +1.

Coefficient of Skewness based on Moments: It is defined by:

$$\alpha_3 = \sqrt{b_1} = \frac{m_3}{\sqrt{m_2^3}} \qquad \text{(For sample)}$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}} \qquad \text{(For population)}$$

Normal Distribution: A distribution is said to be normal if its b_1=0 and b_2=3 respectively. The curve of the normal distribution is Bell-shaped and symmetric.

For a Bell-shaped symmetric distribution:

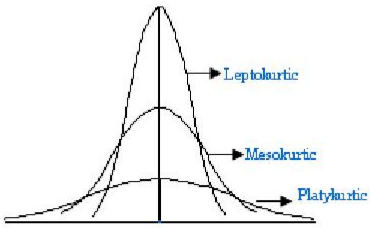
- ✓ 68.27% area of the normal curve lies under the range $\mu \pm \delta$
- ✓ 95.45% area of the normal curve lies under the range $\mu \pm 2\delta$
- ✓ 99.73% area of the normal curve lies under the range $\mu \pm 3\delta$
- ✓ Mean Deviation = 4/5 Standard Deviation
- ✓ Quartile Deviation = 2/3 Standard Deviation
- ✓ Quartile Deviation = 5/6 Mean Deviation

Kurtosis: *"The degree of peakedness or flatness of a frequency distribution relative to normal distribution is called Kurtosis". OR*

"The characteristic by which we compare the "hump" of a distribution with normal distribution is called kurtosis".

Kurtosis indicates whether a particular distribution is flatter or more peaked than the normal curve. Kurtosis is measured by the b2

- If $b_2 > 3$, then the distribution is known as **leptokurtic**
- If $b_2 = 3$, then the distribution is known as **mesokurtic**
- If $b_2 < 3$, then the distribution is known as **platykurtic**



To measure the skewness we will use: $b_2 = \frac{m_4}{m_2^2}$

Importance of Moments: It is a known fact that Statistics is divided into two types i.e. descriptive statistics and inferential statistics. Moments are the only concept in whole of the Statistics, which belong to both descriptive as well as inferential statistics. Now the point arises in what way moments are descriptive and in what way moments are inferential. To understand in what way moments are descriptive, let us consider the following table:

Moments as Descriptive Measures	
1st moment about origin	m_1' or μ_1' = *Arithmetic Mean*
2nd moment about mean	m_2 or μ_2 = *Variance*
Coefficient of skewness	b_1 or β_1 = *Measure of Skewness*
Coefficient of kurtosis	b_2 or β_2 = *Measure of Kurtosis*

So far we have discussed four characteristics of a frequency distribution given as follows:

- Measure of Central Tendency
- Measure of Dispersion
- Measure of Skewness
- Measure of Kurtosis

All of these Descriptive characteristics can be obtained by the moments.

On the other hand, Moments are included in inferential statistics in the sense that there is a very old method of estimation, commonly known as estimation by method of moments. In estimation, by the method of moments we find as many moments as the number of parameters desired to be estimated both from the theoretical distribution and from the sample data. By equating corresponding moments and solving the equations obtained in this we, we get the estimators for the parameters under consideration.

Q: Explain the term skewness and kurtosis.

Skewness: *"Skewness is the lack of symmetry around the average".*

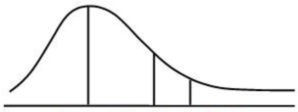
The presence of skewness tells us that a particular distribution is not symmetrical. In skewed distribution the curve is turned more to one side than the other.

There are two types of skewness:

- Positive skewness
- Negative skewness

Positive Skewness

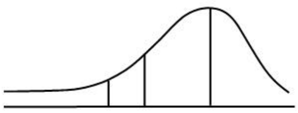
- Skewness is positive, if Mean > Median > Mode
- In terms of moments, skewness is positive if: $\sqrt{b_1}$ or $\sqrt{\beta_1} > 0$
- Skewness is positive, if the right tail of a distribution is longer than its left tail.



Mean > Median > Mode

Negative Skewness

- Skewness is negative, if Mean < Median < Mode
- In terms of moments, skewness is negative if: $\sqrt{b_1}$ or $\sqrt{\beta_1} < 0$
- Skewness is negative, if the left tail of a distribution is longer than its right tail.



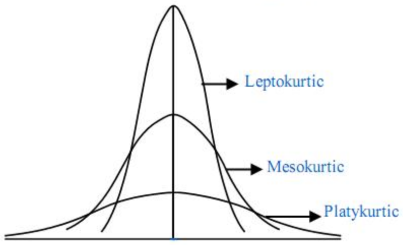
Mean < Median < Mode

Kurtosis: *"The characteristic by which we compare the "hump" of a distribution with normal distribution is called kurtosis".*

Kurtosis indicates whether a particular distribution is flatter or more peaked than the normal curve.

Kurtosis is measured by the moment ratio b_2 or β_2

- If $\beta_2 > 3$, then the distribution is known as leptokurtic
- If $\beta_2 = 3$, then the distribution is known as mesokurtic
- If $\beta_2 < 3$, then the distribution is known as platykurtic.



Q: Explain the difference between absolute and relative measure of dispersion.

Absolute Measure of Dispersion: "An absolute measure of dispersion is one that measures the dispersion in terms of the same units or in the square of units as the units of the data" e.g. if the units of the data are Rs, meters, kg, etc. The units of the measures of dispersion will also be Rs, meters, kg, etc.

The common absolute measures of dispersion are:

- Range
- Quartile Deviation
- Mean Deviation
- Standard Deviation

Relative Measure of Dispersion: "A relative measure of dispersion is one; that is expressed in the form of a ratio or percentage and independent of the units of measurements"
This is a pure number and independent of the units in which the data has been expressed. It is used for the purpose to compare the dispersion of one data set with the dispersion of another.

The common relative measures of dispersion are:

- Coefficient of Dispersion or Coefficient of Range
- Coefficient of Quartile Deviation
- Coefficient of Mean Deviation
- Coefficient of Standard Deviation or Coefficient of Variation (C.V)

Q: State the measures commonly employed to define skewness and kurtosis. What aspects of the frequency curve are measured by them?

There are various measures to define skewness and kurtosis but the following are commonly employed measures to define skewness and kurtosis:

To define skewness we use $\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}}$

The following aspects of the frequency curve are measured by it:

- If $\gamma_1 = 0$ then the frequency curve is symmetric.
- If $\gamma_1 < 0$ then the frequency curve is negatively skewed i.e. the left tail of the frequency curve is longer.
- If $\gamma_1 > 0$ then the frequency curve is positively skewed i.e. the right tail of the frequency curve is longer.

It is to be noted that Karl Pearson and Bowley's coefficients can also be used to define skewness.

Similarly to define kurtosis we use $\beta_2 = \frac{\mu_4}{\mu_2^2}$

The following aspects of the frequency curve are measured by it:

- If $\beta_2 = 3$ then the frequency curve is mesokurtic i.e. it is similar to the Normal curve
- If $\beta_2 > 3$ then the frequency curve is leptokurtic i.e. it is peak as compared to the Normal curve
- If $\beta_2 < 3$ then the frequency curve is platykurtic i.e. it is flatter as compared to the Normal curve

Q: Define moment's ratios b_1 and b_2 and state the purpose for which they are used.

Moments Ratios: There are certain ratios in which both the numerator and denominators are the moments about mean. These ratios are called moment's ratios and are denoted by β_1 and β_2 or b_1 and b_2, where:

	1stMoments Ratio	2nd Moments Ratio
For Sample	$b_1 = \frac{m_3^2}{m_2^3}$	$b_2 = \frac{m_4}{m_2^2}$
For Population	$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$	$\beta_2 = \frac{\mu_4}{\mu_2^2}$

The moment ratios are independent of origin and units of measurements.

Purpose: It is to be noted that, $\sqrt{\beta_1}$ or $\sqrt{b_1}$ is the measure of skewness while β_2 or b_2 is the measure of kurtosis. Hence moment ratios are use to determined the shape of the distribution.