UNIT - 2

Business Statistics

Ms. Jyoti Ma'am

B.COM 1st

List of Contents

Created by Jyoti Yadav

Measures of Central Tendency

The representative value of a data set, generally the central value or the most occurring value that gives a general idea of the whole data set is called Measure of Central Tendency.

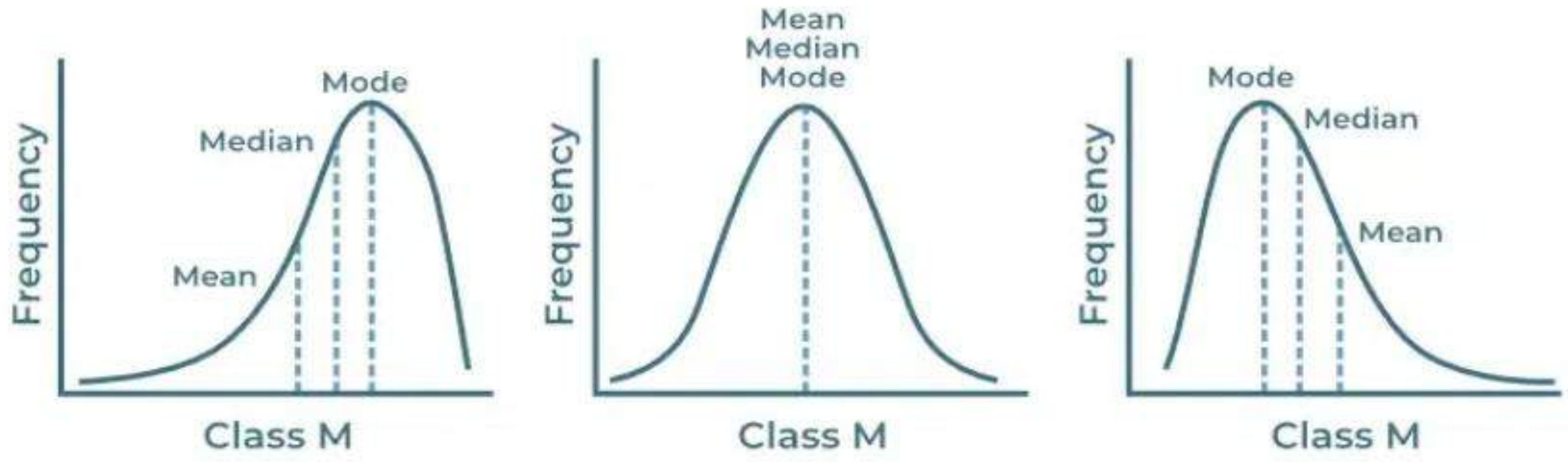
Central tendency is a fundamental concept in statistics that refers to the measure that identifies the center or average value of a dataset. It provides a single value that attempts to describe an entire set of data by indicating where most of the values lie. The mean, median and mode are the three commonly used measures of central tendency.

Created by Jyoti Yadav

Following are some of the important measures of central tendency which are commonly used in business and industry.

- Arithmetic Mean
- Weighted Arithmetic Mean
- Geometric Mean
- Harmonic Mean
- Median
- Mode
- Quantiles(quartiles, deciles and percentiles)

The three most commonly used measures of central tendency are mean, median, and mode



Created by Jyoti Yadav

Definition of Mean in Statistics

This is the arithmetic average of all values in a dataset. It is calculated by adding all the values and then dividing the sum by the total number of values. Mean is nothing but the average of the given set of values. it can be affected by extremely high or low values (outliers).

but other than the arithmetic mean there are geometric mean and harmonic mean as well that are calculated using different formulas

Mean = (Sum of all the observations/Total number of observations)

$$2 + 3 + 5 + 6 = \frac{16}{4} = 4$$

The Mean Number is 4

Examples

What is the mean of 2, 4, 6, 8 and 10?

Solution:

First, add all the numbers.

2 + 4 + 6 + 8 + 10 = 30

Now divide by 5 (total number of observations).

Mean = 30/5 = 6

Mean Symbol (X Bar)

The symbol of mean is usually given by the symbol '$\bar{x}$'. The bar above the letter x, represents the mean of x number of values.

$\bar{X}$ = (Sum of values ÷ Number of values)

$\bar{X} = (x_1 + x_2 + x_3 + ... + x_n)/n$

Created by Jyoti Yadav

ARITHMETIC MEAN

The arithmetic mean (or mean or average) is the most commonly used and readily understood measure of central tendency. In statistics, the term average refers to any of the measures of central tendency.

1. **Ungrouped data/Raw data:**

The arithmetic mean is defined as being equal to the sum of the numerical values of each and every observation divided by the total number of observations. Symbolically, it can be represented as:

$$\bar{X} = \Sigma x/n$$

Where,

, ∑X indicates the sum of the values of all the observations, and N is the total number of observations.

For example,

Q.1) the monthly salary (Rs.) of 10 employees of a firm ABC

2500, 2700, 2400, 2300, 2550, 2650, 2750, 2450, 2600, 2400

Answer: If we compute the arithmetic mean, then

2500+2700+2400+2300+2550+2650+2750+2450+2600+2400 = 25300

Mean=25300 /10= Rs. 2530.

Therefore, the average monthly salary is Rs. 2530.

2. Discrete data

When the observations are classified into a frequency distribution, Therefore, for discrete data; the arithmetic mean is defined as

$$\bar{X} = \Sigma f x / f$$

Where,

f is the frequency for corresponding variable x and N is the total frequency, i.e. N = Σf

Formula of Mean Using Assumed Mean

$$\bar{x} = A + \frac{\sum f d}{\sum f}$$

Where:

- $\bar{x}$ =Mean
- A= Assumed mean (a value from the data, often a midpoint of a class)
- f= Frequency of each class
- $d = x - A$= Deviation of each class midpoint $x$from the assumed mean A
- $\sum f\ d$ =Sum of the products of frequency and deviation
- $\sum f$ =Total frequency

Step-by-step:

1. Choose an assumed mean A(usually the midpoint of a class near the center).
2. Calculate deviations $d = x - A$for each class midpoint x.
3. Multiply each deviation by the corresponding frequency $f$to get fd.
4. Sum all fdvalues.
5. Divide $\sum f\ d$by total frequency $\sum f$.
6. Add this result to the assumed mean $A$to get the mean $\bar{x}$.

What is Combined Mean?

When two or more series having different arithmetic means and number of items are combined together, the combined mean of all the series can be calculated using,

$$\text{Combined Mean i.e. } \bar{X}_{12} = \frac{\bar{X}_1.N_1+\bar{X}_2.N_2}{N_1+N_2}$$

$$\text{Combined Mean i.e. } \bar{X}_{123} = \frac{\bar{X}_1.N_1+\bar{X}_2.N_2+\bar{X}_3.N_3}{N_1+N_2+N_3}$$

Where,

$\bar{X}_{12}$ = *Combined Mean*

$\bar{X}_1$ = *Mean of the first series*

N_1 = *Number of items in the first series*

$\bar{X}_2$ = *Mean of the second series*

N_2 = *Number of items in the second series*

$\bar{X}_3$ = *Mean of the third series*

N_3 = *Number of items on the third series*

Harmonic Mean Definition

The Harmonic Mean (HM) is defined as the reciprocal of the average of the reciprocals of the data values.. It is based on all the observations, and it is rigidly defined. Harmonic mean gives less weightage to the large values and large weightage to the small values to balance the values correctly. In general, the harmonic mean is used when there is a necessity to give greater weight to the smaller items. It is applied in the case of times and average rates.

Harmonic Mean Formula

Since the harmonic mean is the reciprocal of the average of reciprocals, the formula to define the harmonic mean "HM" is given as follows:

If x_1, x_2, x_3,..., x_n are the individual items up to n terms, then,

Harmonic Mean, HM = n / [(1/x_1)+(1/x_2)+(1/x_3)+...+(1/x_n)]

For frequency data: $HM = \frac{\sum f}{\sum \frac{f}{x}}$

How to Find a Harmonic Mean?

If a, b, c, d, ... are the given data values, then the steps to find the harmonic mean are as follows:

Step 1: Calculate the reciprocal of each value (1/a, 1/b, 1/c, 1/d, ...)

Step 2: Find the average of reciprocals obtained from step 1.

Step 3: Finally, take the reciprocal of the average obtained in step 2.

Example 1:
Find the harmonic mean for data 2, 5, 7, and 9.

Example 2:

Calculate the harmonic mean for the following data:

x	1	3	5	7	9	11
f	2	4	6	8	10	12

Solution — harmonic mean (step-by-step)

For data with frequencies the harmonic mean is

$$H = \frac{N}{\sum_i \frac{f_i}{x_i}}$$

where $N = \sum_i f_i$

1. Compute total frequency:

$$N = 2 + 4 + 6 + 8 + 10 + 12 = 42.$$

2. Compute each $\frac{f_i}{x_i}$:

$$\frac{2}{1} = 2.000000, \frac{4}{3} = 1.333333, \frac{6}{5} = 1.200000, \frac{8}{7} \approx 1.142857, \frac{10}{9} \approx 1.111111, \frac{12}{11} \approx 1.090909.$$

3. Sum them:

$$\sum \frac{f_i}{x_i} = 2 + 1.333333 + 1.200000 + 1.142857 + 1.111111 + 1.090909$$

$$\approx 7.87821067821.$$

4. Compute harmonic mean:

$$H = \frac{42}{7.87821067821} \approx 5.3311597919.$$

Geometric Mean Definition

In Mathematics, the **Geometric Mean (GM)** is the average value or mean which signifies the central tendency of the set of numbers by finding the product of their values. Basically, we multiply the numbers altogether and take the nth root of the multiplied numbers, where n is the total number of data values. For example: for a given set of two numbers such as 3 and 1, the geometric mean is equal to $\sqrt{(3\times1)} = \sqrt{3} = 1.732$.

Formula:

$$G.M = \sqrt[n]{x_1 \times x_2 \times \ldots x_n}$$

or

$$G.M = (x_1 \times x_2 \times \ldots x_n)^{\frac{1}{n}}$$

This can also be written as;

$$Log\ GM = \frac{1}{n}\log(x_1 \times x_2 \times \ldots . x_n)$$
$$= \frac{1}{n}(\log x_1 + \log x_2 + \ldots . + \log x_n)$$
$$= \frac{\sum \log x_i}{n}$$

Therefore, Geometric Mean,

$$GM = Antilog \frac{\sum \log x_i}{n}$$

Where n = $f_1 + f_2 + \ldots + f_n$

It is also represented as:

$$G.M. = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

For any Grouped Data, G.M can be written as;

$$GM = Antilog \frac{\sum f \log x_i}{n}$$

Merits

1.It is rigidly defined.

2.It is based on all the observations of the series.

3.It is suitable for measuring the relative changes.

4.It gives more weights to the small values and less weights to the large values.

5.It is used in averaging the ratios, percentages and in determining the rate gradual increase and decrease.

Demerits

1.It is not easy to understand by a man of ordinary prudence as it involves logarithmic operations. As such it is not popular like that of arithmetic average.

2.It is difficult to calculate as it involves finding out of the root of the products of certain values either directly, or through logarithmic operations.

3.It cannot be calculated, if the number of negative values is odd.

4.It cannot be calculated, if any value of a series is zero.

5.At times it gives a value which may not be found in the series, and may even be assured or impracticable.

Difference Between Arithmetic Mean and Geometric Mean

Arithmetic Mean	Geometric Mean
The arithmetic mean or mean can be found by adding all the numbers for the given data set divided by the number of data points in a set.	It can be found by multiplying all the numbers in the given data set and take the nth root for the obtained result.
For example, the given data sets are: 5, 10, 15 and 20 Here, the number of data points = 4 Arithmetic mean or mean = (5+10+15+20)/4 Mean = 50/4 =12.5	For example, consider the given data set, 4, 10, 16, 24 Here n= 4 Therefore, the G.M = 4th root of (4 ×10 ×16 × 24) = 4th root of 15360 G.M = 11.13

Relation Between AM, GM and HM

To learn the relation between the AM, GM and HM, first we need to know the formulas of all these three types of the mean.

Assume that "x" and "y" are the two number and the number of values = 2, then

$AM = (a+b)/2$

$\Rightarrow 1/AM = 2/(a+b)$ .... (1)

$GM = \sqrt{(ab)}$

$\Rightarrow GM^2 = ab$ .... (2)

$HM = 2/[(1/a) + (1/b)]$

$\Rightarrow HM = 2/[(a+b)/ab$

$\Rightarrow HM = 2ab/(a+b)$ ..... (3)

Now, substitute (1) and (2) in (3), we get

$HM = GM^2 / AM$

$\Rightarrow GM^2 = AM \times HM$

Or else,

$GM = \sqrt{[AM \times HM]}$

Hence, the relation between AM, GM and HM is $GM^2 = AM \times HM$

Median

A second measure of central tendency is the median. Median is that value which divides the distribution into two equal parts. Fifty per cent of the observations in the distribution are above the value of median and other fifty per cent of the observations are below this value of median. The median is the value of the middle observation when the series is arranged in order of size or magnitude (Ascending order).

Ungrouped Data

If the number of observations is odd, then the median is equal to one of the original observations (Middle).

$$\text{Median} = \left(\frac{N+1}{2}\right) \text{th value}$$

UNGROUPED DATA

If the number of observations is **odd**, then the median is equal to one of the original observations (Middle).

$$\text{Median} = \left(\frac{N+1}{2}\right) \text{ th value}$$

For example, if the income of seven persons in rupees is 1100, 1200, 1350, 1500, 1550, 1600, 1800, then

Median $=((7+1)/2) = 4^{th}$ value

Median =1500

If the number of observations is even, then the median is the arithmetic mean of the two middle observations.

$$\text{Median} = \frac{\left\{\left(\frac{N}{2}\right) + \left(\frac{N}{2}+1\right)\right\}}{2} \text{ th value}$$

For example, if the income of eight persons in rupees is 1100, 1200, 1350, 1500, 1550, 1600, 1800,1850, then the median income of eight persons would be 1500+1550/2= 1525

Find out the median from the following:
57 58 61 42 38 65 72 66 80

Step 1: Arrange the Series

Step 2: Use Formula:

Step 3: Find Median

$$\text{Median} = \text{Size of } \frac{N+1}{2} \text{th Item}$$

$$= \text{Size of } \frac{9+1}{2} \text{ the item}$$

$= 10 / 2 = 5^{th}$ item

Median = 62

DISCRETE SERIES

First we find cumulative frequency.then locate (N+1/2) the value in cumulative frequency.corresponding that value of x is median.

X	f	cf
10	12	12
20	23	35
30	35	70
40	47	117
50	38	155
60	29	184
70	16	200
Sum	200	

N=200

N+1/2=100.5

Median= 40

Discrete Series

Compute the median for the following distribution of weeks of wagers of 65 employees of the xyz company

Weekly wages in Rs	55	65	785	85	95	105	115
Number of employees	8	10	16	14	10	5	2

Solution

Weekly wages in Rs	No of Employees	Cumulative frequency (cf)
55	8	8
65	10	18
75	16	34
85	14	48
95	10	58
105	5	63
115	2	65

$$\text{Median} = \text{Size of } \frac{N+1}{2} \text{th Item}$$

$$= \text{size of } \frac{65+1}{2} \text{th Item}$$

= 33' which is nearer to 34

Cf of 34 = 75

Median weekly wages = 75

CONTINUOUS DATA

For continuous data, First we find cumulative frequency. then locate (N+1/2) the value in cumulative frequency. corresponding class interval is median class.the following formula may be used to locate the value of median.

$$\text{Median} = l_1 + \{(N/2 - cf)*h\} / f$$

where l_1 is the lower limit of the median class, cf is the preceding cumulative frequency to the median class, f is the frequency of the median class and h is the width of the median class.

Consider the following data which relate to the age distribution of 1000 workers in an industrial establishment.
The location of median value is facilitated by the use of a cumulative frequency distribution as shown below in the table.

Age (Years)	No. of workers f	Cumulative frequency c.f
Below 25	120	120
25-30	125	245
30-35	180	**425(cf)**
35-40	**160(f)**	585
40-45	150	735
45-50	140	875
50-55	100	975
55 and Above	25	1000

N=1000
Median Class=(1000+1)/2=500.5th =(35-40)
l_1 =35,cf =425,f=160,N/2=500,h=5

Median =35+{(500-425)*5}/160 =35+375/160=35+2.34=**37.34**

MERITS OF MEDIAN

[1] It is easy to understand and calculate.

[2] It is rigidly defined.

[3] It is not affected by extreme values.

[4] It is calculated in case of open-end interval.

[5] It is located by graphically also.

DEMERITS OF MEDIAN

[1] It is not based on all observations.

[2] It is affected by sampling fluctuation.

[3] It is not capable of further algebraic treatment.

MODE :

The mode is the typical or commonly observed value in a set of data. It is defined as the value which ***occurs most*** often or with the greatest frequency.

For example, in the series of numbers 3, 4, 5, 5, 6, 7, 8, 8, 8, 9,

The mode is 8 because it occurs the maximum number of times. That means in ungrouped data mode can find by inspection only.

The mode is the most frequently occurring measurement in a data set.

There may be **one mode; multiple modes**

To determine the mode:

1. **Put the data in order from smallest to largest, as you did to find your median.**
2. **Look for any value that occurs more than once.**
3. **Determine which of the values from Step 2 occurs most frequently.**

Example:

Consider the data set: 17, 10, 9, 14, 13, 17, 12, 20, 14

Step

Step 1: Put the data in order from smallest to largest.

Step 2: Look for any number that occurs more than once. Step

Step 3: Determine which of those occur most frequently.

The modes of this data set are 14 and 17.

DISCRETE DATA

Mode is the value of X which has highest frequency.
For example,

X	f
10	12
20	23
30	35
40	47
50	38
60	29
70	16
Sum	200

Mode=40

CONTINUOUS DATA

First, we find Modal class = corresponding to highest frequency.

$$\text{MODE} = l_1 + \left\{\frac{f1-f0}{2f1-f0-f2}\right\} * h$$

where l_1 is lower limit of the modal class, f_1 is the frequency of the modal class, f_0 the frequency of the preceding class, f_2 is the frequency of the succeeding class, h is the size of the modal class.

To illustrate the computation of mode, let us consider the following data.

Age (Years)	No. of workers f
Below 25	120
25-30	125**(f_0)**
30-35	180**(f_1)**
35-40	**160(f_2)**
40-45	150
45-50	140
50-55	100
55 and Above	25

Modal class=(30-35)

l_1 =30, f_1 =180, f_0 =125, f_2 =160,h=5

Mode =30+{(180-125)/(2*180-125-160)}*5 =30+(55/75)*5 =30+3.667=**33.67**

MERITS OF MODE

[1] It is easy to understand and calculate.

[2] It is calculated in case of open-end interval.

[3] It is located by graphically also.

DEMERITS OF MODE

[1] It is not based on all observations.

[2] It is not capable of further algebraic treatment

[3] It is not rigidly defined.

Thank You

Created by Jyoti Yadav

BCom 1st

Geometric Mean

Jyoti

1) Individual Series:

x	Log of x
–	–
–	–
–	–
	$\Sigma \log x$ (total)

we calculate log x with the help of log table

$$G.M = \text{Antilog}\left(\frac{\Sigma \log x}{N \text{ (no. of } x)}\right)$$

↳ with the help of A.L. Table

2) Discreate Series:

x	Log x (Table)		f	Log $x f$
–	–	x	–	–
–	–	x	–	–
–	–	x	–	–
–	–	x	–	–
		(gun)	Σf (n)	$\Sigma \log x f$

$$Gm = \text{Antilog}\left(\frac{\Sigma \log x f}{N \rightarrow (\Sigma f)}\right)$$

(3) Continuous Series:

Class	f	x ($\frac{L+U}{2}$)	Log x (Table)	Log $x f$
0-10	–	5	–	–
10-20	–	15	–	–
20-30	–	25	–	–
30-40	–	35	–	–
	Σf (n)			$\Sigma \log x f$

$$Gm = \text{Antilog}\left(\frac{\Sigma \log x f}{N}\right)$$

Log Table

Diff. column → mean diff

main	0	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
10																			
11																			
12																			
⋮																			

Log : 1456

- count [1456] : 4 − 1 = 3
- 3.1632

[14] → [5] → [6]

main   Diff.   Mean Diff.

Table   1614 + 18

- ⇒ 1632

इस प्रकार   1456 = x

3.1632 = $\log x$

Ex x	$\log x$
1.456	0.1632
14.56	1.1632
145.6	2.1632 (Point का कोई मतलब x)
1456.0	3.1632

- Point के पहले कैसे लिखना है, आया ? characteristic → mantissa

x	$\log x$
1234	3.0913
123.4	2.0913
12.34	1.0913
1.234	0.0913

[12]34 = 4 − 1

= 3.0913

0899 + 14

0913

Q: Find out GM (indi. series)

x	$\log x$
Q: 173.00	2.2380
182.00	2.2601
75.00	1.8751
+ 5.00	0.6990
0.8	$\bar{1}$.9031
0.08	$\bar{2}$.9031
0.8974	$\bar{1}$.9530
	$\Sigma \log x = 5.8314$

Jyoti

Count 1st ⇒ after dec

.2380 + (.2601 + .8751 + ... = 4.8314

+1

5.8314

+5
−4
1

1) when single digit before decimal, we can count one more digit after decimal bcz of mantissa

2) when 0 is before decimal we count the numbers (only digit not 0) before decimal and sign (−)

3) we can't add directly 2.2380 (2 → Ch, .2380 → mantissa)

$$GM = \text{Antilog}\left(\frac{\Sigma \log x}{N}\right)$$

$$= \frac{5.8314}{7}$$

(1) zero Antilog = (0.8331)

ignore 0 → 0.8331

83 3 1

6800 + 2 = 6810

Antilog = 6.81 → GM

Median –

(1) Individual Series:–

(A) Find out the median: x: 10, 15, 9, 25, 18

Ans: 15

(B) x: 12, 18, 9, 25

Ans: 15

[2] Discrete Series: (ungrouped data)

(A)

x	1	2	3	4	5	6
f	8	12	16	19	20	25

Ans: 4

[3] Continuous Series (grouped data)

$$\text{median } (M) = L_1 + \frac{m-c}{f} \times i \quad \text{(or)}$$

$$M = L_1 + \frac{m-c}{f} \times (L_2 - L_1)$$

where, L_1 = Lower limit, L_2 = upper limit, f = freq.

$i = L_2 - L_1$ (class size), m = median no = $\frac{N}{2}$

c = Cumulative freq. (Less then type) of class preceding of the median class

3) **Harmonic Mean:** The harmonic mean is the reciprocals of averaged numbers. It is defined as the reciprocal of the arithmetic mean of the reciprocal of the individual observations. Under certain conditions, harmonic mean is a better measure of central tendency e.g., computation of average speed, average price, etc.

Mathematically,

$$H.M. = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + \frac{1}{X_3} + ... + \frac{1}{X_N}}$$

Where, N = Number of observations;
X_N = Value of observation

2.1.4.2. Positional Average

Positional average is the average which depends on the position of the items, rather than the values of the items.

1) **Median:** When N observations are grouped and arranged in the sorting order (ascending or descending order) according to their values, then the central value of the observation is known as median. It is denoted by M or Me.

2) **Mode:** Mode is the value of the variable for which the frequency is maximum and it is denoted by Z or Mo.

3) **Quartiles:** The measure of central tendency which divides a group of data into four subgroups or parts, then it is called quartile. The three quartiles are denoted as Q_1, Q_2, and Q_3.

4) **Deciles:** If the values of the observation in a data set are arranged in ascending or descending order and divided it into ten equal parts by using nine points on the scale of observations, then it is called **decile**.

5) **Percentiles:** If the values of the observation in a data set are arranged in ascending or descending order and divided it into hundred equal parts by using ninety nine points on the scale of observations, then it is called **percentile**.

2.2. MEAN/ ARITHMETIC MEAN

2.2.1. Introduction

Mean is also known as arithmetic mean (A.M.). To calculate the mean, summate all the observations and divide it by the total number of observations. Mean is denoted by $\overline{X}$.

So, $$\overline{X} = \frac{\text{Sum of all observations}}{\text{Number of observations}} = \frac{\sum X}{N}$$

According to W. I. King, "The arithmetic average may be defined as the sum or aggregate of a series of items divided by their number".

For example, the arithmetic mean of the four numbers 32, 40, 44 and 36 is given by,

$$\overline{X} = \frac{32+40+44+36}{4} = \frac{152}{4} = 38$$

2.2.2. Advantages of Mean

1) Simple to calculate and understand.
2) Some value is always determined, i.e., it is never indefinite.
3) Can be used in other algebraic calculations.
4) No need of sorting or arrangement (ascending or descending order).
5) It is stable and not affected by the variation of sampling.

2.2.3. Disadvantages of Mean

1) Mean is greatly affected by the extreme values. **For example,** mean of 3, 7 and 200 is 70. Here, no value is present near this 70; hence this average is of no use.
2) Sometimes mean may provide confusing impressions. **For example,** in a hospital, per day average number of patients admitted is 5.7. Here given information is useful but doesn't provide the actual item because some values are of no use when expressed in fraction or decimal.
3) The 'mean' cannot be predicted by just inspecting the sample item.
4) If a single value is missing, mean cannot be calculated.
5) In case of open-end classes, mean cannot be calculated.
6) Graphical representation of mean is not possible.

2.2.4. Implications of Arithmetic Mean

1) The sum of the deviations of the individual items from the arithmetic mean is always zero. This means $\sum(X-\overline{X})=0$, where X is the value of an item and $\overline{X}$ is the arithmetic mean. 'Since the sum of the derivations in the positive direction is equal to the sum of the deviations in the negative direction, the arithmetic mean is regarded as a measure of location.'

2) The sum of the squared deviations of the individual items from the arithmetic mean is always minimum. In other words, the sum of the squared deviations taken from any value other than the arithmetic mean will be higher.

3) Since the arithmetic mean is based on all the observations in a series, a change in the value of any observation will lead to a change in the value of the arithmetic mean.

4) In the case of a highly skewed distribution, the arithmetic mean may get inaccurate on account of a few items with extreme values. In such a case, it may not represent the characteristic of the distribution.

2.2.5. Application of Arithmetic mean

1) Arithmetic mean is used to measure the standard deviation.
2) Arithmetic mean is used in the construction of index numbers.
3) It is also used in the hypotheses testing.

2.2.6. Types of Arithmetic Mean

There are two types of arithmetic mean:

Types of Arithmetic Mean
- Simple (or unweighted) Arithmetic Mean
- Weighted Arithmetic Mean

2.2.7. Method of Calculation of Arithmetic Mean

The methods for calculating arithmetic mean vary according to various types of series such as:

1) Calculation of Arithmetic Mean - Individual Series
2) Calculation of Arithmetic Mean - Discrete Series (Ungrouped Data)
3) Calculation of Arithmetic Mean - Continuous Series (Grouped Data)

2.2.7.1. Calculation of Arithmetic Mean - Individual Series

The methods used are as follows:

1) **Direct Method:** The direct method of calculating arithmetic mean in the case of individual series involves the following steps:
   i) Add the various given values of the variable X and find the total value which is denoted by ΣX.
   ii) Divide this total value (ΣX) by the number of observations (N) i.e.,

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{N} = \frac{\Sigma X}{N}$$

Example 1: The weekly wage of 5 workers is as given below:

1350, 1400, 1450, 1370 and 1480

Find arithmetic mean.

Solution: **Computation of Arithmetic Mean**

Serial Number	Weekly Wages (in Rupees)
1	1350
2	1400
3	1450
4	1370
5	1480
N = 5	**ΣX = 7050**

$$\overline{X} = \frac{\sum X}{N} = \frac{7050}{5} = ₹1410$$

2) **Short-Cut Method:** With the use of short cut method also, arithmetic mean can be calculated. The calculation of arithmetic mean by short cut method includes the following steps:
   i) Any value may be taken as an assumed mean of the data. This assumed value is also known as working mean or arbitrary average (A = Assumed mean).
   ii) Subtract assumed mean from each value of the observation (d = X – A).
   iii) All the deviations are added and it is denoted by (Σd).
   iv) Apply the formula:

$$\overline{X} = A + \frac{\sum d}{N}$$

Where,

$\overline{X}$ = Arithmetic mean ;

A = Assumed mean

$\sum d$ = Sum of the deviation;

N = Number of items

Example 2: Calculate mean from the following data:

Years	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Price of Rice (in ₹)	40	50	55	80	58	60	75	35	45	52

Solution: Let the assumed mean be A = 50

Years	Price of Rice (X)	d = (X – 50)
2001	40	–10
2002	50	0
2003	55	5
2004	80	30
2005	58	8
2006	60	10
2007	75	25
2008	35	–15
2009	45	–5
2010	52	2
N = 10		**Σd = 50**

$$\overline{X} = A + \frac{\sum d}{N} = 50 + \frac{50}{10} = 50 + 5 = ₹55.$$

2.2.7.2. Calculation of Arithmetic Mean - Discrete Series (Ungrouped Data)

The methods used are as follows:

1) **Direct Method:** In the discrete series, the sum of items is determined by multiplying each value with the respective frequency. The values got after multiplication are totalled up. To find the arithmetic mean, total value is divided by the total number of items.

   Following are the steps which are involved in the calculation of mean:
   i) Multiply each item by its frequency, denoted by (fX).
   ii) Add all the fX, denoted by (ΣfX).

(iii) ΣfX is divided by the total number of items.

The formula is $\overline{X} = \frac{\sum fX}{\sum f}$

Where, $\overline{X}$ = Arithmetic mean;
ΣfX = The sum of products;
Σf = Total number of items.

Example 3: Find average wages of 10 workers:

Daily Wage (in ₹)	4	8	10	11	14	Total
No. of Workers	2	1	4	2	1	10

Solution:

Daily Wage (X)	No. of Workers (f)	fX
4	2	8
8	1	6
10	4	40
11	2	22
14	1	14
Total	$\Sigma f = 10$	$\Sigma fX = 90$

∴ Arithmetic Mean (Average Wage) $= \frac{\sum fX}{\sum f}$

$= \frac{90}{10} = ₹9.00$

2) **Short-Cut Method:** For calculating the arithmetic mean by short-cut method in a discrete series the following steps are taken:

i) Any value may be taken as an assumed mean of the data. This assumed value is also known as working mean or arbitrary average (A = Assumed mean). For making the calculation easy, the assumed mean should be close to the middle of the frequency distribution.

ii) Subtract assumed mean from each value of the observation ($d_x = X - A$).

iii) Each value of deviation is multiplied by its respective frequency, denoted by (fd_x).

iv) Following is the formula which is used for the calculation of arithmetic mean:

$$\overline{X} = A + \frac{\sum fd_x}{\sum f} \quad \text{or } A + \frac{\sum fd_x}{\sum N}$$

Where, $\overline{X}$ = Arithmetic Mean
A = Assumed Mean
Σfd_x = Total of deviations multiplied with the respective frequencies
Σf = Total of frequencies (N)

Example 4: From the following frequency distributions find out the mean weight of the 100 persons:

Weight (in kg.)	64	65	66	67	68	69	70	71	72	73
No. of Persons	15	13	18	5	20	11	7	6	3	2

Solution: Let assumed mean (A) = 68;

Short-cut Method

Weight in Kg. (X)	No. of Persons (f)	Deviation (d_x) = (X − A)	fd_x
64	15	−4	−60
65	13	−3	−39
66	18	−2	−36
67	5	−1	−5
68 (A)	20	0	0
69	11	1	11
70	7	2	14
71	6	3	18
72	3	4	12
73	2	5	10
	$\Sigma f = 100$		$\Sigma fd_x = -75$

$$\overline{X} = A + \frac{\Sigma fd_x}{\Sigma f} = 68 + \frac{(-75)}{100} = 68 - 0.75 = 67.25 \text{ Kg.}$$

2.2.7.3. Calculation of Arithmetic Mean - Continuous Series (Grouped Data)

The methods used are as follows:

1) **Direct Method:** For calculating the arithmetic mean in a continuous series the following steps are taken:

i) First mid value of each class-interval is determined by adding the lower and upper limit of each class and dividing the total by two. **For example,** in a class interval say 0 – 10, the mid value is 5.

$\left(\frac{0+10}{2} = 10/2 = 5\right)$ (This is denoted by X).

ii) Multiply these mid values of each class with the respective frequency of each class. In other words X will be multiplied by f.

iii) Add up all the products and obtain ΣfX.

iv) ΣfX is divided by the sum of the frequencies i.e., Σf.

v) Apply the formula: $\overline{X} = \frac{\sum fX}{\sum f}$

Example 5: From the following find out the mean profits:

Profits per Share (₹)	100-200	200-300	300-400	400-500	500-600	600-700	700-800
Number of Shares	12	20	18	30	32	26	22

Solution: **Calculation of Mean**

Profits	Mid-Point (X)	No. of Shares (f)	fX
100-200	150	12	1800
200-300	250	20	5000
300-400	350	18	6300
400-500	450	30	13500
500-600	550	32	17600
600-700	650	26	16900
700-800	750	22	16500
		$\Sigma f = 160$	$\Sigma fX = 77600$

$$\overline{X} = \frac{\sum fX}{\sum f} = \frac{77600}{160} = 485$$

The average profit is ₹485.

2) **Step-Deviation Method:** For calculating the arithmetic mean by step-deviation method in case of grouped series the following steps are taken:
   i) Find the mid-point of each group or class, denoted by X.
   ii) Any value may be taken as an assumed mean of the data (A = Assumed mean).
   iii) Subtract assumed mean from the mid-point of each class, denoted by $d_x = (X - A)$.
   iv) Deviations (d_x) are divided by their common factor (i) and step deviation is determined for each class-interval.

$$d' = \frac{d_x}{i}.$$

Thus, $$d' = \frac{X - A}{i}$$

   v) The step deviation of each class is multiplied by respective frequency of that class to find the total value $\Sigma fd'$.
   vi) Find the total frequency, $N = \Sigma f$.

Use the formula: $$\overline{X} = A + \frac{\sum fd'}{\sum f} \times i$$

Where,
$\overline{X}$ = Arithmetic mean,
A = Assumed mean;
$\sum fd'$ = Total of product of step-deviations and frequencies;
$\sum f$ = Total number of frequencies;
i = Common factor in x or in d_x

Example 6: Calculate the mean from the following table:

Weights of Children (in kg.)	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90
Number of Children	4	12	8	21	32	28	10	3	2

Solution: Let assumed mean (A) = 55, let I = 10 (Common Factor)

C.I (Weight in kg.)	**Mid Value (X)**	**Frequency (f)**	$d_x = X - A$	$d' = d_x/10$	**fd'**
0 – 10	5	4	– 50	– 5	– 20
10 – 20	15	12	– 40	– 4	– 48
20 – 30	25	8	– 30	– 3	– 24
30 – 40	35	21	– 20	– 2	– 42
40 – 50	45	32	– 10	– 1	– 32
50 – 60	55	28	0	0	0
60 – 70	65	10	10	1	10
70 – 80	75	3	20	2	6
80 – 90	85	2	30	3	6
Total		$\Sigma f = 120$			$\sum fd' = -144$

$$A.M. = A + \frac{\sum fd'}{\sum f} \times i = 55 + \frac{(-144)}{120} \times 10.$$

$$= 55 - 12 \text{ or } = 43 \text{ Kg.}$$

2.2.8. Weighted Arithmetic Mean

In the economic studies, weighted arithmetic mean plays a very important role. In the calculation of weighted arithmetic mean, items are assigned weights according to their relative importance in the group. The relative importance of various items in a distribution may be different; some items may be relatively more important than others. **For example**, to calculate mean-expenditure of a family, it would be wrong to give equal importance to different items of family expenditure. The family may give different importance to the various items depending upon how much monthly expenditure is incurred on those items. They may give more importance to food and less importance to entertainment or clothes. Hence, the arithmetic mean should be calculated according to their relative importance.

Uses of Weighted Arithmetic Mean

Arithmetic mean cannot present the real picture of any situation because it gives equal importance to all the items.

Whereas weighted mean is an important measure because all items of a group do not have equal importance in the calculation.

Following are some specific cases where calculation of weighted mean will be useful:
1) When index numbers are constructed.
2) When the results of two or more universities are compared and the number of students differs.
3) When standardised death and birth rates are calculated.

Methods of Calculation of Weighted Arithmetic Mean

1) **Direct Method:** The following steps are taken in the calculation of weighted arithmetic mean by direct method:

   i) Multiply size of the variable (X) and their respective weights and add up the products.

   ii) Divide the total by sum of the weights.

   iii) Following formula is used for the calculation of weighted arithmetic mean:

$$\bar{X}_W = \frac{\sum WX}{\sum W}$$

Where,

$\bar{X}_w$ = Weighted arithmetic mean

W = Weights,

X = The variable

2) **Short-Cut Method:** The following steps are taken in the calculation of weighted arithmetic mean by short-cut method:

   i) Any value may be taken as an assumed weighted mean of the data (A_w = Assumed mean).

   ii) Subtract assumed mean from each value of the observation ($d_x = X - A_w$)

   iii) Each value of deviation is multiplied with its respective weight, denoted by Wd_x and then added up, $\sum(Wd_x)$.

   iv) Following is the formula which is used in the calculation of weighted arithmetic mean:

$$\bar{X}_w = A_w + \frac{\sum Wd_x}{\sum W}$$

Where,

$\bar{X}_w$ = Weighted arithmetic mean;

A_w = Assumed weighted mean

ΣWd_x = Total of deviations multiplied with the respective weights;

ΣW = Total of weights

Example 7: An examination was held to decide the award of a scholarship. The weights given to the various subjects were different. Only three applicants for the scholarship obtained over 50% marks in each subject. The marks were as follows:

Subject	Weight	Marks of A	Marks of B	Marks of C
Statistics	4	65	58	66
Accountancy	3	64	65	72
Economics	2	58	57	58
Mercantile Law	1	72	79	54

Of the candidates the one getting the highest marks is to be awarded the scholarship, who should get it?

Solution:

Subject	Weight W	Marks of A X_1	Marks of A WX_1	Marks of B X_2	Marks of B WX_2	Marks of C X_3	Marks of C WX_3
Statistics	4	65	260	58	232	66	264
Accountancy	3	64	192	65	195	72	216
Economics	2	58	116	57	114	58	116
Mercantile Law	1	72	72	79	79	54	54
Total	ΣW = 10		ΣWX_1 = 640		ΣWX_2 = 620		ΣWX_3 = 650

Weighted Mean of A, $\bar{X}_{w_1} = \frac{\sum WX_1}{\sum W} = \frac{640}{10} = 64$ Marks

Weighted Mean of B, $\bar{X}_{w_2} = \frac{\sum WX_2}{\sum W} = \frac{620}{10} = 62$ Marks

Weighted Mean of C, $\bar{X}_{w_3} = \frac{\sum WX_3}{\sum W} = \frac{650}{10} = 65$ Marks

The Weighted Mean of C is the highest; hence he is entitled for scholarship.

Example 8: An investor is fond of investing in equity shares. During a period of falling prices in the stock exchange, a stock is sold at ₹120 per share on one day, ₹105 on the next and ₹90 on the third day. The investor has purchased 50 shares on the first day, 80 shares on the second day and 100 shares on the third day. What average price per share did the investor pay?

Solution:

Calculation of Weighted Average Price

Day	Price per Share (₹)x	No. of Shares Purchased w	Amount Paid ₹ wx
1	120	50	6000
2	105	80	8400
3	90	100	9000
Total	–	**230**	**23,400**

$$\text{Weighted Average} = \frac{w_1x_1 + w_2x_2 + w_3x_3}{w_1 + w_2 + w_3}$$

$$= \frac{\sum wx}{\sum w} = \frac{6000+8400+9000}{50+80+100} = ₹101.7$$

Thus, the investor paid an average price of ₹101.7 per share.

It will be seen if merely prices of the shares for the three days (regardless of the number of shares purchased) are taken into consideration, then the average price would be:

$$\frac{120+105+90}{3} = ₹105$$

This is an unweighted or simple average and as it ignores the quantum of shares purchased, it fails to give a correct picture. A simple average, it may be noted, is also a weighted average where weight in each case is the same, that is, only 1. When we use the term 'average' alone, we always mean that it is an unweighted or simple average.

2.2.9. Combined Mean/Grouped Mean

If the means of two or more than two sub-groups or different samples from the same universe along with their numbers are given, then a combined mean can be calculated with the help of following formula:

$$\text{Combined Mean}(\overline{X}) = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3 ......+ N_n\overline{X}_n}{N_1 + N_2 + N_3 ......+ N_n}$$

N_1, N_2, N_3........ are the frequencies of different groups and $\overline{X}_1, \overline{X}_2, \overline{X}_3$.......are the arithmetic means of these groups.

Example 9: An analysis of the average speed of the buses in two transport companies A and B, gives the following results:

	Transport Company A	Transport Company B
Number of Buses	595	645
Average Speed (Km/h)	52.5	47.5

Find the combined average speed.

Solution: Combined Average Speed in the two transport companies would be,

$$\overline{X}_{12} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

Where, N_1 and N_2 = the number of items in the two series

$\overline{X}_1$ and $\overline{X}_2$ = the mean of the two series respectively

$\overline{X}_{12}$ = the combined mean of the two series.

Given, $N_1 = 595$ and $N_2 = 645$; $\overline{X}_1 = 52.5$ and $\overline{X}_2 = 47.5$

Substituting the values,

$$\text{Combined Average Speed} = \frac{(595 \times 52.5) + (645 \times 47.5)}{595 + 645}$$

$$= \frac{61875}{1240} = ₹49.9$$

2.2.10. Various Cases in the Calculation of Arithmetic Mean

The following are some typical examples on the calculation of arithmetic mean:

1) **Calculation of Arithmetic Mean of Inclusive Series:** A series in which the upper limit of class interval does not repeat itself as a lower limit of the next class interval then it is called an inclusive series.

For example, 0-9, 10-19, 20-19, etc., shows an inclusive series. One can calculate arithmetic mean in the case of inclusive series without converting it into exclusive series.

Example 10: Find average profit per shop from the following data:

Profit per shop (In thousand)	10-12	13-15	16-18	19-21	22-24	25-27	28-30
No. of shops	35	51	85	65	27	19	8

Solution: (Profit per Shop)

C.I.	Mid Value (X)	f	fX
10-12	11	35	385
13-15	14	51	714
16-18	17	85	1445
19-21	20	65	1300
22-24	23	27	621
25-27	26	19	494
28-30	29	8	232
		Σf = 290	**ΣfX = 5191**

Here, C.I. = class interval; f = frequency;

$\sum f = 290$; $\sum fX = 5191$

$$\text{Mean }(\overline{X}) = \frac{\sum fX}{\sum f} = \frac{5191}{290} = 17.9$$

2) **Calculation of Arithmetic Mean in a Cumulative Frequency Distribution:** For the calculation of arithmetic mean in the case of cumulative frequency distribution, given frequencies are converted into simple frequency distribution. Cumulative frequency distribution may be classified into the following two types:

i) **Less than or Below Cumulative Frequency Distribution:** In the case of less than or below cumulative frequency distribution, we first convert it into simple frequency distribution. Now, one can calculate arithmetic mean by using any method.

Example 11: Find the mean from the following frequency distribution.

Marks Less than	10	20	30	40	50	60
No. of Students	10	15	35	70	90	100

Solution: Assume A = 25.

Calculation of Mean by Step-Deviation Method

Marks	No. of Students (f)	Mid-Values (X)	Deviations $d_x = (X - A)$ A = 25	Step-Deviations $d' = \frac{dx}{i}$ when C = 10w	fd'
0-10	10	5	−20	−2	−20
10-20	5 (15 − 10)	15	−10	−1	−5
20-30	20 (35 − 15)	25	0	0	0
30-40	35 (70 − 35)	35	10	1	35
40-50	20 (90 − 70)	45	20	2	40
50-60	10 (100 − 90)	55	30	3	30
	Σf = 100				**Σfd' = 80**

$$\bar{X} = A + \frac{\sum fd'}{\sum f} \times i = 25 + \frac{80}{100} \times 10 = 33 \text{ Marks}$$

ii) **More than or Above Cumulative Frequency Distribution:** In this case, we obtain frequencies of each class by subtracting the second class frequency from first class frequency and third from second and so on as shown in the example given below:

Example 12: Find the arithmetic mean from the following data:

Age (In Years)	No. of Persons	Age	No. of Persons
Above 0	134	Above 60	36
Above 10	128	Above 70	22
Above 20	123	Above 80	17
Above 30	90	Above 90	13
Above 40	87		
Above 50	66		

Solution:

Age	No. of Persons (f)	Mid-Values (X)	(fX)
0-10	6(134 – 128)	5	30
10-20	5(128 – 123)	15	75
20-30	33(123 – 90)	25	825
30-40	3(90 – 87)	35	105
40-50	21(87 – 66)	45	945
50-60	30(66 – 36)	55	1650
60-70	14(36 – 22)	65	910
70-80	5(22 – 17)	75	375
80-90	4(17 – 13)	85	340
90-100	13	95	1235
	Σf = 134		**ΣfX = 6490**

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{6840}{144} = 48.3 \text{ Years}$$

3) **Calculation of Arithmetic Mean in Unequal Classes:** In the case of continuous series which have unequal class interval, there is no need of any adjustment in the classes and arithmetic mean can be calculated in its original form.

Example 13: Calculate the mean of the following series:

Weekly Income (₹)	No. of Families
0-20	18
20-50	25
50-90	40
90-140	16
140-200	32

Solution: Let the assumed mean (A) be 70.

Weekly Income (₹)	No. of Families (f)	Mid-Values (X)	$d_x = (X - A)$	$d' = \frac{d_x}{5}$	fd'
0-20	18	10	–60	–12	–216
20-50	25	35	–35	–7	–175
50-90	40	70	0	0	0
90-140	16	115	+45	+9	144
140-200	32	170	+100	+20	640
	Σf = 131				**Σfd' = + 393**

$$\bar{X} = A + \left(\frac{\Sigma fd'}{N} \times i\right) = 70 + \left(\frac{393}{131} \times 5\right) = 70 + 15 = ₹85$$

4) **Calculation of Arithmetic Mean in Open-Ended Classes:** A series in which the lower class limit of the first class interval or the upper limit of the last class interval is missing and 'less than' or 'below' and 'more than' or 'above' are specified in lower class limit of first class or upper limit of last class, following two situations are considered in this case:

i) **When Class Intervals are Equal:** In such situation, let us take the following example:

Classes	Below 20	20-40	40-60	60-80	80 and above
Frequencies	5	8	9	11	6

Here the class sizes are same and it is 20. For calculating the lower limit of first class, the class size, i.e., (20) is subtracted from the upper limit of first class, i.e., (20 – 20 = 0). For determining the upper limit of first class, the class size, i.e., (20) is added to the lower limit of the first class, i.e., (0 + 20 = 20). For determining the upper limit of last class, the class size, i.e., (20) is added to the upper limit of the previous class, i.e., (80 + 20 = 100). The first class will be (0 – 20) and the last (80 – 100). Then one can calculate mean by using any method.

ii) **When Class-Intervals are Unequal:**

Weights	Below 20	20-25	25-35	35-50	50-70	70 and above
No. of Persons	7	11	14	19	6	2

Here, the class magnitude is unequal. The size of second, third, fourth, fifth class is 5, 10, 15, 20 respectively, i.e., it is increasing by 5. For calculating the lower limit of first class, the value 5 is subtracted from the upper limit of second class, i.e., (20 – 5 = 15). For determining the upper limit of last class, the value 30 is added to the upper limit of the previous class, i.e., (70 + 30 = 100). The first class will be (15 – 20) and the last class will be (70 – 100). Then one can calculate mean by using any method.

5) **Calculation of Arithmetic Mean in Mid-Value Series:** A series in which only mid-values of the class intervals & the corresponding frequencies are given then it is called mid-value series. In the calculation of arithmetic mean for the continuous series one requires mid-values of the classes. But in this case mid-values are already given so there is no need to adjust the frequencies & construct classes.

Example 14: Calculate the mean of the following series:

Mid-Values (X)	No. of Families (f)
10	17
20	21
30	13
40	11
50	18

Solution: Let the assumed mean (A) is 30.

Mid-Values (X)	No. of Families (f)	$d_x = (X - A)$	$d' = \frac{d_x}{10}$	fd'
10	17	–20	–2	–34
20	21	–10	–1	–21
30	13	0	0	0
40	11	+10	+1	+11
50	18	+20	+2	+36
	80			$\Sigma fd' = -8$

$$\overline{X} = A + \left(\frac{\Sigma fd'}{N} \times i\right) = 30 + \left(\frac{-8}{80} \times 10\right) = 30 - 1 = 29$$

6) **Determination of Missing Value:** Let us assume that the missing values of variables or frequencies are α, β, γ, etc. Now, calculate the value of mean by using general formula. With the available information required equation can be made. Missing values can be determined by solving these linear equations.

Example 15: Find out the missing size from the following:

Marks	10	15	20	α	30	40
No. of Students	3	6	8	10	7	6

Given: $\overline{X} = 24.5$

Solution:

Marks (X)	No. of Student (f)	Total Marks (fX)
10	3	30
15	6	90
20	8	160
α	10	10 α
30	7	210
40	6	240
	$\Sigma f = 40$	$\Sigma fX = 730 + 10\alpha$

$$\overline{X} = \frac{\Sigma fX}{\Sigma f}$$

$$\therefore \quad 24.5 = \frac{730 + 10\alpha}{40}$$

$\Rightarrow 24.5 \times 40 = 730 + 10\alpha$; $\quad \Rightarrow 980 = 730 + 10\alpha$

$\Rightarrow 980 - 730 = 10\alpha$,

$\Rightarrow \alpha = \frac{250}{10} = 25$ marks

2.3. MEDIAN

2.3.1. Introduction

When N observations are grouped and arranged in the sorting order (ascending or descending order) according to their values, then the central value of the observation is known as median. It is denoted by M or Me.

So,

$$M = \left(\frac{N+1}{2}\right)^{th} \text{observation.}$$

According to Prof. L. R. Conner, "The median is that value of the variable which divides the group into two equal parts, one part comprising of all values greater and the other all values less than the median".

According to Croxton and Cowden, "The median is that value which divides a series so that one half or more of the items are equal to or less than it and one half or more of the items are equal to or greater than it."

2.3.2. Advantages of Median

1) It is unaffected by the extreme values.
2) In case of individual observation and discrete series, it is easy to understand and calculate.
3) It can be used in other algebraic calculations and also useful during calculation of mean deviation.
4) After sorting the values of variable, it can be located by inspection.
5) It can also be measured by using 'graphical representation'.
6) Median of open-end classes can be measured.
7) Median is clearly definite in nature i.e., it is always clearly defined.

2.3.3. Disadvantages of Median

1) Sorting (ascending and descending order) is necessary for the calculation.
2) It cannot be used to measure the combined mean of two or more groups.
3) When sampling is fluctuated then median also gets fluctuated (affected more than mean).
4) It is not based on all the data samples. It is a positional average.

2.3.4. Implications of Median

1) The median can be calculated from open-ended distributions while arithmetic mean cannot be calculated.
2) Median can be calculated graphically while arithmetic mean cannot be calculated.
3) Median is preferred in case when the distribution has extreme values, because median is not affected by extreme values.
4) It is the most appropriate measure of location when it is not possible to count the items, yet they are scored or ranked.

2.3.5. Method of Calculation of Median

1) Calculation of Median – Individual Series
2) Calculation of Median – Discrete Series (Ungrouped Data)
3) Calculation of Median – Continuous Series (Grouped Data)

2.3.5.1. Calculation of Median – Individual Series

In case of individual series, the following steps are taken for calculating the median:

1) If the given observations are not arranged in ascending or descending order of magnitude then it has to be arranged for calculation

2) Locate the middle value:

i) There will be a single value in the middle which is selected as median when the number of observation N is odd. Mathematically

$$\text{Median} = \text{Value of } \left(\frac{N+1}{2}\right)^{th} \text{items}$$

Example 16: Find out the median of the following items:

X : 10, 15, 9, 25, 18

Solution:

Computation of Median

S. No.	Size of Item Ascending Order (X)		Size of Item Descending Order (X)
1)	9		25
2)	10		18
3)	15	← [Median] →	15
4)	18		10
5)	25		9

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

$$= \text{Size of } \left(\frac{5+1}{2}\right)^{th} \text{item}$$

$$= 3^{rd} \text{ item} = 15$$

In the above example the number of items was odd so there was no difficulty in locating the middle item and its value.

ii) If N is even, there will be two middle values, $\left(\frac{N}{2}\right)^{th}$ and $\left(\frac{N}{2}+1\right)^{th}$. The arithmetic mean of these two middle values is taken as median. That is

$$\text{Median} = \frac{\left(\frac{N}{2}\right)^{th} \text{value} + \left(\frac{N}{2}+1\right)^{th} \text{value}}{2}$$

Example 17: Find out the median of the following items:

X : 12, 18, 9, 25.

Solution: **Computation of Median**

S. No.	Size of Item Ascending Order (X)
1)	9
2)	12
3)	18
4)	25

Here, N = 4 which is even, so

$$\text{Median} = \frac{\left(\frac{N}{2}\right)^{th} \text{value} + \left(\frac{N}{2}+1\right)^{th} \text{value}}{2}$$

$$= \frac{\left(\frac{4}{2}\right)^{th} \text{value} + \left(\frac{4}{2}+1\right)^{th} \text{value}}{2}$$

$$= \frac{2^{nd} \text{ value} + 3^{rd} \text{ value}}{2} = \frac{12+18}{2} = \frac{30}{2} = 15$$

2.3.5.2. Calculation of Median – Discrete Series (Ungrouped Data)

In case of discrete series, the following steps are taken for calculating the median:

1) If the given observations are not arranged in ascending or descending order of magnitude then it has to be arranged for calculation.
2) Calculate the total frequency, i.e., $\sum f = N$.
3) Calculate the cumulative frequency (in general, less than type, data arranged in ascending order).
4) Locate the median by using the following formula:

$$\text{Median}(M) = \left(\frac{N+1}{2}\right)^{th} \text{value,}$$

$$(\text{Where, } N = \sum f)$$

The value, for which the cumulative frequency includes $\left(\frac{N+1}{2}\right)^{th}$ value, is selected as median.

Example 18: To find the median of the following distribution:

X	1	2	3	4	5	6
f	8	12	16	19	20	25

Solution:

X	f	Cumulative frequency
1	8	8
2	12	20
3	16	36
4 ← Median	19	55 ← Cumulative frequency of median
5	20	75
6	25	100
	N = 100	

$$\text{Now median} = \text{Value of } \left(\frac{N+1}{2}\right)^{th} \text{item}$$

$$= \text{Value of } \left(\frac{100+1}{2}\right)^{th} \text{item or,}$$

Value of 50.5^{th} item. This value lies in cumulative frequency (55) for the value 4.

So,

Median = 4.

2.3.5.3. Calculation of Median – Continuous Series (Grouped Data)

In case of continuous series, the following steps are taken for calculating the median:

1) If the given series is in inclusive form then it has to be converted in exclusive form.

2) Calculate the total frequency,

$N = \sum f$ and median number $\left(\frac{N}{2}\right)^{th}$ value.

3) If the data is arranged in ascending order, then calculate "less than" type cumulative frequencies.

4) Calculate median class with the use of $m = \frac{N}{2}$. The class in which median falls is called the median class.

5) Median can be calculated with the use of following formula:

$$\text{Median}(M) = L_1 + \frac{m-c}{f} \times i \quad \text{or} \quad M = L_1 + \frac{m-c}{f} \times (L_2 - L_1)$$

Where, M = median,

m = median number $= \frac{N}{2}$,

L_1 = lower limit of the median class,

L_2 = upper limit of the median class,

f = frequency of the median class,

$i = L_2 - L_1$ = width of the median class (Class size).

c = cumulative frequency (less then type) of the class preceding to the median class.

Example 19: Find the median and median class of the following data:

Class-Interval	15-25	25-35	35-45	45-55	55-65	65-75
Frequency	5	10	18	15	0	3

Solution:

Class interval	Frequency	Cumulative Frequency
15 – 25	5	5
25 – 35	10	15 ← 'c' cumulative frequency of the class preceding the median class
35 – 45 ← M. Class	18 ← f	34 ← Cumulative frequency of median class
45 – 55	15	49
55 – 65	0	49
65 – 75	3	52
	N = 51	

Median number, $m = \frac{N}{2} = \frac{51}{2} = 26^{th}$ element, and this value lie in cumulative frequency (34) and class interval is (35 – 45). So the median class is (35– 45).

Here, m = 26, c = 15, f = 18 and $L_2 - L_1 = 10$.

Hence, $M = L_1 + \frac{m-c}{f} \times (L_2 - L_1)$

$$= 35 + \frac{26-15}{18} \times 10$$

$$= 35 + \frac{11}{18} \times 10 = 35 + \frac{110}{18}$$

$$= 35 + 6.11 = 41.11$$

Required median is 41.11 and median class is (35 – 45).

2.3.6. Various Cases in the Calculation of Median

The following are some typical examples on the calculation of Median:

1) **Median in Case of Inclusive Series**

Example 20: Calculate median from following data:

Value	Frequency	Value	Frequency
0-4	328	30-39	599
5-9	350	40-49	525
10-19	720	50-59	378
20-29	664	60-69	246

Solution: **Computation of Median**

Class-Interval	Frequency	Cumulative Frequency
0-4	328	328
5-9	350	678
10-19	720	1398
20-29	664	2062
30-39	599	2661
40-49	525	3186
50-59	378	3564
60-69	246	3810

Median is the value of $\left(\frac{3810}{2}\right)^{th}$ or 1905^{th} item which lies in 20-29 group. This is a case of inclusive class-intervals. Which should be made exclusive and hence the median group should be deemed to be in 19.5-29.5 group. Thus the value of L_1 would be 19.5 and not 20.

$$\text{Median}(M) = L_1 + \frac{m-c}{f} \times (L_2 - L_1)$$

$$= 19.5 + \frac{1905-1398}{664}(29.5-19.5)$$

$$= 19.5 + \left(\frac{507}{664} \times 10\right) = 27.1$$

2) **Median in Case of Exclusive Series**

Example 21: Find the median of the following distribution:

Class Intervals (₹)	Frequencies	Class-Intervals (₹)	Frequencies
1-3	6	11-13	16
3-5	54	13-15	4
5-7	85	15-17	5
7-9	56		
9-11	22		
			248

Solution: **Calculation of Median**

Class-Intervals	Frequency	Cumulative Frequency
1-3	6	6
3-5	54	60
5-7	85	145
7-9	56	201
9-11	22	223
11-13	16	239
13-15	4	243
15-17	5	248

Median = $\left(\frac{N}{2}\right)^{th}$ item, i.e., 124^{th} item,

which lies in 5-7 group.

$$M = L_1 + \frac{m-c}{f} \times (L_2 - L_1) = 5 + \frac{124-60}{85}(7-5) = 6.5$$

3) **Median in Case of Cumulative Frequency Distribution:** If the data is available in the form of cumulative series then for obtaining the frequency of the median class it is required to convert the series into simple series. Once that is done, the rest of the procedure is the same as in any other continuous series. The following example would illustrate the point:

Example 22: Calculate the median from the following data:

Value	Frequency	Value	Frequency
Less than 10	4	Less than 50	98
Less than 20	15	Less than 60	114
Less than 30	40	Less than 70	123
Less than 40	77	Less than 80	128

Solution:

Computation of Median

Value	Frequency	Cumulative Frequency
0-10	4	4
10-20	11	15
20-30	25	40
30-40	37	77
40-50	21	98
50-60	16	114
60-70	9	123
70-80	5	128

Middle item is $\left(\frac{128}{2}\right)^{th}$ item or 64^{th} item which lies in 30-40 group.

$$M = L_1 + \frac{m-c}{f} \times (L_2 - L_1) = 30 + \frac{64-40}{37}(40-30)$$

$$= 30 + \left(\frac{24}{37} \times 10\right) = 30 + 6.5 = 36.5$$

4) **Determination of Missing Value**

Example 23: Find out the missing frequencies if N = 100 median = ₹ 50:

Expenditure (₹)	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
No. of Families	14	?	26	?	16

Solution: Let the frequency of the class 20 – 40 be 'α' and the frequency of the class 60 – 80 be 'β'

Then, $14 + \alpha + 26 + \beta + 16 = 100$,

Total frequency

$\Rightarrow$ $\alpha + \beta + 56 = 100$ or $\alpha + \beta = 44$ ...(1)

Computation of Median

Expenditure (₹)	Frequency (f)	Cumulative frequency (c.f.)
0 – 20	14	14
20 – 40	α	$14 + \alpha$
40 – 60	26	$40 + \alpha$
60 – 80	β	$40 + \alpha + \beta$
80 – 100	16	$56 + \alpha + \beta$
Total	$56 + \alpha + \beta$	

Median = ₹ 50, N = 100

Since, the median is 50, so median class would be (40 – 60)

Median number

$$m = \frac{N}{2} =$$

$\frac{100}{2}$, $c = 14 + \alpha$, $L_1 = 40$, $L_2 = 60$, $f = 26$

Hence Median, $M = L_1 + \frac{m-c}{f} \times (L_2 - L_1)$

$\Rightarrow$ $$50 = 40 + \frac{\frac{100}{2} - (14+\alpha)}{26} \times (60-40)$$

$\Rightarrow$ $$50 - 40 = \frac{36-\alpha}{26} \times 20$$

$\Rightarrow$ $10 \times 26 = (36 - \alpha) \times 20$

$\Rightarrow$ $260 = 720 - 20\alpha$

$\Rightarrow$ $20\alpha = 720 - 260 = 460$

$\therefore$ $$\alpha = \frac{460}{20} = 23$$

Substituting this value of α in equation (1), we get $\beta = 44 - 23 = 21$

Thus the missing frequencies are 23 and 21.

2.4. MODE

2.4.1. Introduction

Mode is the value of the variable for which the frequency is maximum and is is denoted by Z or Mo.

According to Zizek, "Mode is the value occurring most frequently in a series (or group) of items and around which other items are distributed most densely".

According to Croxton and Cowden, "The mode of a distribution is the value at the point around which the items tend to be most heavily concentrated. It may be regarded as the most typical of a series of values".

2.4.2. Advantages of Mode

1) Easy to understand and calculate.
2) Can be easily found out by using inspection method.
3) It is an actual value, which most frequently occurs in the series.

4) Not affected by extreme values.
5) It is simple and accurate and can be measured in an open end class-interval without determining the class limits.

2.4.3. Disadvantages of Mode

1) It is ill defined and in few cases it is not possible to find a definite value.
2) Not a good representative because it is not based on all observations.
3) In further algebraic calculation it is of no use.
4) It may be possible that there is no mode or more than one mode present in the sample.
5) Fluctuation in sampling effects is made to a higher degree than mean or median.
6) Before calculating the mode, data sorting (ascending or descending order) is necessary.

2.4.4. Implications of Mode

1) The value which represents the whole distribution is termed as mode.
2) Mode can be used to describe the qualitative phenomenon of the distribution.
3) The value of mode can be calculated by using graphic method.
4) Mode is generally determined in the case when the given distribution is highly skewed or there is an asymmetrical distribution.

2.4.5. Methods of Calculation of Mode

1) Calculation of Mode – Individual Series
2) Calculation of Mode – Discrete Series (Ungrouped Data)
3) Calculation of Mode – Continuous Series (Grouped Data)

2.4.5.1. Calculation of Mode – Individual Series

If the data is not arranged then it has to be arranged in increasing order. Now find the value occurring the maximum number of times. This value is called the 'mode' of the given data.

Example 24: Calculate mode from the following data of income earned by 10 employees:

S. No.	Income (in ₹)	S. No.	Income (in ₹)
1	10	6	27
2	27	7	20
3	24	8	18
4	18	9	15
5	27	10	32

Solution:
Calculation of Mode

Income (in ₹)	Number of Employees
10	1
15	1
18	2
20	1
24	1
27	3
32	1
Total	**10**

Since the item 27 occurs the maximum number of times, i.e. 3, hence the modal income is 27.

Note: The process of determining mode in case of individual observations essentially involves grouping of data. When there are two or more values having the same maximum frequency, one cannot say which the modal value is and hence mode is said to be **ill-defined**. Such a series is also known as **bi-modal or multi-modal**.

2.4.5.2. Calculation of Mode – Discrete Series (Ungrouped Data)

In the discrete series the mode is the value which has highest frequency. If two or more values have the same frequency or distribution of frequency is irregular, then mode cannot be determined by the method of inspection. In this situation, a grouping table and an analysis table is to be prepared to find out the value of mode. First we prepare grouping table and then an analysis table. The following steps are taken for calculating the mode:

1) A grouping table is prepared with 6 columns.
2) The size of the item is written in the margin.
3) Frequencies of the respective items are written in column 1 and highest frequency is marked.
4) In column 2, the frequencies are grouped in twos by adding 1st and 2nd, 3rd and 4th, 5th and 6th and so on and highest total is marked.
5) In column 3, leaving the first frequency in column 1and the rest frequencies are grouped in twos by adding 2nd and 3rd, 4th and 5th, 6th and 7th and so on and highest total is marked.
6) In column 4, starting from the first frequency of column 1, the frequencies are grouped in threes by adding 1st, 2nd and 3rd; 4th, 5th and 6th; 7th, 8th, and 9th; and so on and highest total is marked.
7) In column 5, starting from the second frequency (leaving the first frequency) of column 1, frequencies are grouped in threes by adding 2nd, 3rd and 4th; 5th, 6th and 7th; 8th, 9th and 10th and so on and highest total is marked.
8) In column 6, starting from the second frequency (leaving the first two frequencies) of column 1, frequencies are grouped in threes by adding 3rd, 4th and 5th; 6th, 7th and 8th so on and highest total is marked.

9) After completing the grouping of frequencies table, an analysis table is prepared to find out the item (group) which appears as the highest frequency. The same steps are applied in the case of continuous series.

Example 25: Compute the mode from the following data:

Size	2	3	4	5	6	7	8	9	10	11	12	13
Frequency	3	8	10	12	16	14	10	8	17	5	4	1

Solution: Due to irregular distribution of frequencies, we use the method of grouping to decide which one may be considered as maximum frequency.

Grouping Table

Value of the Items (m)	Frequency or Sum of Frequencies					
	1	2	3	4	5	6
2	3	11		21		
3	8		18		30	
4	10	22				38
5	12		28	42		
6	16	30			40	
7	14		24			32
8	10	18		35		
9	8		25		30	
10	17	22				26
11	5		9	10		
12	4	5				
13	1					

Analysis table is as follows:

Analysis Table

Column No.	Size of Items Containing Maximum Frequency											
	2	3	4	5	6	7	8	9	10	11	12	13
1									1			
2					1	1						
3				1	1							
4				1	1	1						
5					1	1	1					
6			1	1	1							
No. of Items	**0**	**0**	**1**	**3**	**5**	**3**	**1**	**0**	**0**	**0**	**0**	**0**

Since 6 occurs maximum number of times, hence the value 6 is the mode.

2.4.5.3. Calculation of Mode – Continuous Series (Grouped Data)

In a continuous series, determination of mode requires one step more than that used for discrete series. Firstly, one has to find the modal class by observation, or in case of 'irregular distribution' the modal class is determined by preparing grouping table and analysis tables. Finally, mode is obtained by following formula:

$$\text{Mode or } Z = L_1 + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$$

Where,

L_1 = lower limit of the modal class,

f_m = frequency of the modal class

f_1 = frequency of the class preceding the modal class.

f_2 = frequency of the class succeeding the modal class

$i = (L_2 - L_1)$ = Width of modal Class (class size)

Example 26: Following is the frequency distribution of yearly expenditure of 60 families.

Yearly Expenditure (in thousand)	50-55	55-60	60-65	65-70	70-75	75-80
No. of Families	3	8	14	20	16	2

Find the modal value.

Solution: Here the class-interval (65 – 70) has the maximum frequency, i.e., 20, therefore, modal class is (65 – 70).

Now we can calculate the modal value of using the formula, $\text{Mode} = L_1 + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times i$

Here, $L_1 = 65$, $f_m = 20$, $f_1 = 14$, $f_2 = 16$ and $i = 5$;

$$\therefore \quad \text{Mode} = 65 + \frac{20-14}{2 \times 20 - 14 - 16} \times 5 = 65 + \frac{30}{10}$$

or Modal value = 65 + 3 = ₹68 thousand

Note: Generally, modes are used for nominal scores, medians for ordinal scores, and means for interval scores.

2.5. QUARTILES

2.5.1. Introduction

The measure of central tendency which divides a group of data into four subgroups or parts then it is called quartiles. The three quartiles are denoted as Q_1, Q_2, and Q_3.

The first quartile, Q_1, divides a frequency distribution in such a way that one-fourth (25%) of the distribution has a value less than Q_1 and three-fourth (75%) have a value more than Q_1.

The second quartile, Q_2 divides a frequency distribution in such a way that it has equal number of observations above and below it. Hence, it is equal to the median of the data.

The third quartile, Q_3 divides a frequency distribution in such a way that three-fourth (75%) of the observations have a value less than Q_3 and one-fourth (25%) have a value more than Q_3.

Uses of Quartiles

Quartiles often are used in sales and survey data to divide populations into groups. For example, one can use QUARTILE to find the top 25 per cent of incomes in a population.

Quartiles are used to summarise a group of numbers. Instead of looking a big list of numbers, we are looking at just a few numbers that give you a picture of what's going on in the big list.

Quartiles are used to:

1) Helps to find
2) Measures of dispersion (Quartile Deviation)

Measures of Dispersion

3.1. MEASURES OF DISPERSION

3.1.1. Meaning and Definition

Distribution cannot be clearly depicted by measuring the averages or central tendency. Averages provide the observations of only the central part of the distribution. So, study of scatteredness of observations is very important and this study is known as **measure of dispersion.**

The word 'Dispersion' literally means 'fluctuation', 'scatter', 'variation', 'deviation', or 'spread'. So, measure of dispersion shows the variation of an individual item from its average.

It is also known as **measure of variation** as it explains the extent of scatterness in an observation and it measures the mean deviation about some central value.

Dispersion is based on values from first order (the average or mean). So, it is known as **second order.**

Deviation is calculated by measuring the difference between mean and actual value.

According to A.L. Bowley, "Dispersion is the measure of the variation of the items."

According to Reigleman, "Dispersion is the extent to which the magnitudes or qualities of the items differ; that is the degree of diversity".

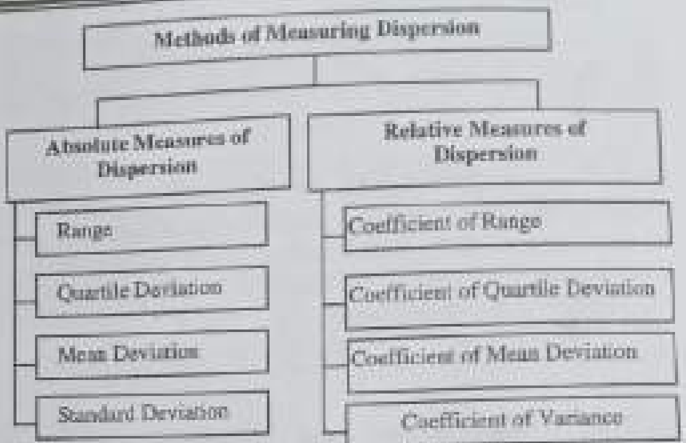
According to Connor, "Dispersion is a measure of the extent to which the individual vary."

According to Kafka, "The measurement of a scatteredness of the mass of figures in a series about an average is called measure of dispersion or measure of variation".

3.1.2. Methods of Measuring Dispersion

The two types of methods for measuring dispersion are as follows:

1) Absolute Measure of Dispersion
2) Relative Measure of Dispersion



3.1.2.1. Absolute Measures of Dispersion

The measure of dispersion which is expressed in terms of the units of the observations (e.g., Rupees, Meter, Years, etc.) is called absolute measure of dispersion.

1) Range
2) Quartile Deviation
3) Mean Deviation
4) Standard Deviation

3.1.2.2. Relative Measures of Dispersion

The measure of dispersion which is independent of unit or may involve the point about which the deviations are taken, is known as relative measure of dispersion or coefficient of dispersion. There is only one relative measure corresponding to an absolute measure of dispersion.

1) Coefficient of Range
2) Coefficient of Quartile Deviation
3) Coefficient of Mean Deviation
4) Coefficient of Variance

3.2. RANGE

3.2.1. Introduction

Range is the simplest absolute measure of dispersion which shows the difference between the highest and the lowest value in a series.

Mathematically: $R = L - S$

Where,

R = Range,
L = Maximum (largest) value,
S = Minimum (smallest) value.

3.2.2. Advantages of Range

1) It is very easy to calculate and simple to understand.
2) It is rigidly defined.
3) It provides the limit (which is measured from set of data) within which all the data items occur.

3.2.3. Disadvantages of Range

1) This method is highly affected by the extreme items as it gives importance to the two extreme values.
2) This method does not provide any information about the structure of the series.
3) It is affected by the sampling fluctuations.
4) Range is not a reliable method to measure the dispersion.

3.2.4. Coefficient of Range

Relative measure of variation (coefficient of range) is used to compare the series. Coefficient of range is defined as:

Coefficient of Range (C.R.) =

$$\frac{\text{Largest Value} - \text{Smallest Value}}{\text{Largest Value} + \text{Smallest Value}} = \frac{L-S}{L+S}$$

3.2.5. Methods of Calculation of Range

1) Calculation of Range - Individual Series
2) Calculation of Range - Discrete Series (Ungrouped Data)
3) Calculation of Range - Grouped Series (Grouped Data)

3.2.5.1. Calculation of Range - Individual Series

To find out the value of range in the individual series, one needs to calculate the difference between highest and lowest value of the series.

Example 1: Calculate Range and its Coefficient from the following data:

54, 45, 16, 15, 75, 85, 28

Solution: Here, Largest value (L) = 85, Smallest value (S) = 15

$$\text{Range} = L - S = 85 - 15 = 70$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{85-15}{85+15} = \frac{70}{100} = 0.7 \text{ or } 70\%$$

3.2.5.2. Calculation of Range - Discrete Series (Ungrouped Data)

In case of individual series, it can also be calculated by taking the difference of highest and lowest value. In this case frequency of the series is not taken into consideration for the calculation.

Example 2: Find range and coefficient of range from the following data:

X	5	15	25	35	45	55
f	7	14	18	10	4	1

Solution: Here,

Largest value (L) = 55, Smallest value (S) = 5

$$\therefore \quad \text{Range} = L - S = 55 - 5 = 50$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{55-5}{55+5} = \frac{50}{60} = 0.83 \text{ or } 83$$

3.2.5.3. Calculation of Range - Grouped Series (Grouped Data)

In case of grouped series it can be calculated by taking the difference of lower limit of the lowest class interval and upper limit of the highest class interval. It can also be measured by taking the difference of mid value of the lowest class interval and mid value of the highest class interval.

Example 3: Calculate Range and the Coefficient of range from the following data:

X	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60	60 – 70	70 – 80	80 – 90	90 – 100
f	6	4	15	24	11	3	10	16	20

Solution: Largest value (L) = 100; Smallest value (S) = 10

$\therefore$ Range (L – S) = 100 – 10 = 90

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{(100-10)}{(100+10)} = \frac{90}{110} = 0.82 \text{ or } 82\%$$

3.3. INTERQUARTILE RANGE

In a data set 'Interquartile range' is the spread of values between third quartile (Q_3) and first quartile (Q_1). In other words, it can be defined by the range of middle 50% of the data. Interquartile range is expressed as follows:

Interquartile Range = $Q_3 - Q_1$

Quartiles are the points which divide the data in four equal parts (Q_1, Q_2, Q_3 and Q_4).

Where,

$$Q_1 = \text{Value of the item} \left(\frac{1}{4}\right)^{th}$$

$$Q_3 = \text{Value of the item} \left(\frac{3}{4}\right)^{th}$$

Half the total numbers of items are included between Q_1 and Q_3.

The difference between third quartile and first quartile ($Q_3 - Q_1$) only includes the central items and excludes the extreme values. The interquartile range is a good and convenient indicator of the absolute variability. If the interquartile range is large, then the variability in the data set will also be large.

Example 4: Calculate Inter-quartile range from the following data:

Years	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
No. of Persons	4	10	15	20	11

Solution:

Years	No. of Persons f	Cumulative Frequency c.f.
0 – 20	4	4
20 – 40	10	14
40 – 60	15	29
60 – 80	20	49
80 – 100	11	60
	N = 60	

$Q_1 = \text{Size of} \left(\frac{N}{4}\right)^{th} \text{item} = \left(\frac{60}{4}\right)^{th} = 15^{th}$ item which lies in

Class interval (40 – 60).

$L_1 = 40, c = 14, f = 15, i = 20$

$$\text{Thus, } Q_1 = L_1 + \frac{\frac{N}{4} - c}{f} \times i = 40 + \frac{15-14}{15} \times 20$$

$$= 40 + \frac{1}{15} \times 20 = 41.33$$

Third Quartile, $Q_3 = \text{Size of} \left(\frac{3N}{4}\right)^{th} \text{item} = \left(\frac{3 \times 60}{4}\right)^{th}$ item

$= 45^{th}$ item which lies in Class interval (60 – 80).

Thus,

$$Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i = 60 + \frac{45-29}{20} \times 20$$

$$= 60 + \frac{16}{20} \times 20 = 60 + 16 = 76$$

Inter-Quartile Range = $Q_3 - Q = 76 - 41.33 = 34.67$ years.

2.5. QUARTILES

2.5.1. Introduction

The measure of central tendency which divides a group of data into four subgroups or parts then it is called quartiles. The three quartiles are denoted as Q_1, Q_2, and Q_3.

The first quartile, Q_1, divides a frequency distribution in such a way that one-fourth (25%) of the distribution has a value less than Q_1 and three-fourth (75%) have a value more than Q_1.

The second quartile, Q_2 divides a frequency distribution in such a way that it has equal number of observations above and below it. Hence, it is equal to the median of the data.

The third quartile, Q_3 divides a frequency distribution in such a way that three-fourth (75%) of the observations have a value less than Q_3 and one-fourth (25%) have a value more than Q_3.

Uses of Quartiles

Quartiles often are used in sales and survey data to divide populations into groups. For example, one can use QUARTILE to find the top 25 per cent of incomes in a population.

Quartiles are used to summarise a group of numbers. Instead of looking a big list of numbers, we are looking at just a few numbers that give you a picture of what's going on in the big list.

Quartiles are used to:

1) Helps to find
2) Measures of dispersion (Quartile Deviation)

3) Coefficient of dispersion (Coefficient of Quartile deviation)
4) Measures of skewness (coefficient of Bowley's coefficient skewness)

2.5.2. Advantages of Quartiles

Salient advantages of quartiles are as follows:
1) It is simple to understand and easy to calculate.
2) It is less affected by the extreme values.
3) It can be calculated exactly in case of open-end classes.

2.5.3. Disadvantages of Quartiles

Quartiles have some disadvantages too, which are as follows:
1) It is not based on all observations. So it may not be a good representative. In fact, 50% of the items in any series are ignored. In other words, it does not cover the first 25% and the last 25% items of a series.
2) It is not capable of further algebraic treatment.
3) It is affected to a greater extent by fluctuations of sampling.

2.5.4. Calculation of Quartiles

1) Calculation of Quartiles - Individual Series
2) Calculation of Quartiles - Discrete Series (Ungrouped Data)
3) Calculation of Quartile - Grouped Series (Grouped Data)

2.5.4.1. Calculation of Quartiles - Individual Series

Let the series be in ascending order. Let N be the number of observations. Then

First quartile $(Q_1) = \left(\frac{N+1}{4}\right)^{th}$ item ;

Second quartile $(Q_2) = \left[\frac{2(N+1)}{4}\right]^{th}$ item or $\left(\frac{N+1}{2}\right)^{th}$ item

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

Example 27: Calculate quartiles from the following data:
Marks obtained: 06, 30, 37, 18, 14, 42, 34, 11, 09, 26, 22, 03, 28, 52, 48

Solution: Arranging the series in ascending order, we get

Serial No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Marks Obtained	03	06	09	11	14	18	22	26	28	30	34	37	42	48	52

Number of observations, N = 15

$$Q_1 = \left(\frac{N+1}{4}\right)^{th} \text{ item} = \left(\frac{15+1}{4}\right)^{th} \text{ item} = 4^{th} \text{ item} = 11;$$

$$Q_2 = \left[\frac{2(N+1)}{4}\right]^{th} \text{ item} = \left[\frac{2(15+1)}{4}\right]^{th} \text{ item} = 8^{th} \text{ item} = 26$$

$$Q_3 = \left[\frac{3(N+1)}{4}\right]^{th} \text{ item} = \left[\frac{3(15+1)}{4}\right]^{th} \text{ item} = 12^{th} \text{ item} = 37$$

2.5.4.2. Calculation of Quartiles - Discrete Series

The computation of quartiles from discrete series involves the following steps:
1) Arrange the data in ascending order of magnitude (if not arranged).
2) Find the less than type cumulative frequencies.
3) Calculate quartiles using the formulae:

$$Q_1 = \left(\frac{N+1}{4}\right)^{th} \text{ item ;}$$

$$Q_2 = \left[\frac{2(N+1)}{4}\right]^{th} \text{ item ;}$$

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

Example 28: Calculate quartile from the following data.

X	3	5	8	12	18	21	26
f	6	11	24	21	16	13	9

Solution:

Variable (X)	Frequency (f)	Cumulative frequency (c.f.)
3	6	6
5	11	17
8	24	41
12	21	62
18	16	78
21	13	91
26	9	100
Total	N = 100	–

$Q_1 = \left(\frac{N+1}{4}\right)^{th}$ item $= \left(\frac{100+1}{4}\right)^{th}$ item $= 25.25^{th}$ item $= 25^{th}$ item $+ 0.25(26^{th}$ item $- 25^{th}$ item$) = 8 + 0.25(8 - 8) = 8$

$Q_2 = \left[\frac{2(N+1)}{4}\right]^{th}$ item $= 50.50^{th}$ item $= 50^{th}$ item $+ 0.5(51^{st}$ item $- 50^{th}$ item$) = 12 + 0.5(12 - 12) = 12$

$Q_3 = \left[\frac{3(N+1)}{4}\right]^{th}$ item $= 75.75^{th}$ item $= 75^{th}$ item $+ 0.75(76^{th}$ item $- 75^{th}$ item$) = 18 + 0.75(18 - 18) = 18$

2.5.4.3. Calculation of Quartile - Grouped Series

The process of computing quartiles in case of a grouped frequency distribution involves the following steps:
1) Arrange the data in ascending order.
2) Obtain less than type cumulative frequencies.

3) Convert the classes into exclusive form if given otherwise.
4) Use the following formula to locate quartiles.

$$\textbf{Formula: } Q_k = L_1 + \frac{\frac{kN}{4} - c}{f}(L_2 - L_1)$$

Where, k = 1, 2, 3; i = $(L_2 - L_1)$ = k^{th} quartile class.
N = total frequency; f = frequency of k^{th} quartile class.
c = cumulative frequency of the class preceding the k^{th} quartile class.

Obviously:

$$Q_1 = L_1 + \frac{\frac{N}{4} - c}{f} \times i$$

$$Q_2 = L_1 + \frac{\frac{2N}{4} - c}{f} \times i$$

$$Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i$$

The values of L_1, L_2, c, i, f, are taken according to respective quartiles

Example 29: Calculate quartiles from the following data:

Age (Years)	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
Number of Persons	3	8	20	12	7

Solution:

Age	Number of Persons (f)	Cumulative frequency
0 – 10	3	3
10 – 20	8	11
20 – 30	20	31
30 – 40	12	43
40 – 50	7	50
Total	N = 50	–

$Q_1 = \left(\frac{N}{4}\right)^{th} = \left(\frac{50}{4}\right)^{th} = 12.5^{th}$ item which is lies in the class-interval (20 – 30).

Hence, $L_1 = 20$, $L_2 = 30$, $f = 20$, $c = 11$, $i = 10$

$$Q_1 = L_1 + \frac{\frac{N}{4} - c}{f} \times i = 20 + \frac{12.5 - 11}{20} \times 10 = 20.75$$

$Q_2 = \left(\frac{2N}{4}\right)^{th}$ item $= \left(\frac{2 \times 50}{4}\right)^{th}$ item $= 25^{th}$ item which is lies in the class-interval (20 – 30).

Hence, $L_1 = 20$, $L_2 = 30$, $f = 20$, $c = 11$, $i = 10$

$$Q_2 = L_1 + \frac{\frac{2N}{4} - c}{f} \times i = 20 + \frac{25 - 11}{20} \times 10 = 27$$

Example 4: Calculate Inter-quartile range from the following data:

Years	0 – 20	20 – 40	40 – 60	60 – 80	80 – 100
No. of Persons	4	10	15	20	11

Solution:

Years	No. of Persons f	Cumulative Frequency c.f.
0 – 20	4	4
20 – 40	10	14
40 – 60	15	29
60 – 80	20	49
80 – 100	11	60
	N = 60	

Q_1 = Size of $\left(\frac{N}{4}\right)^{th}$ item $=\left(\frac{60}{4}\right)^{th}$ = 15^{th} item which lies in Class interval (40 – 60).

$L_1 = 40, c = 14, f = 15, i = 20$

Thus, $Q_1 = L_1 + \frac{\frac{N}{4}-c}{f} \times i = 40 + \frac{15-14}{15} \times 20$

$= 40 + \frac{1}{15} \times 20 = 41.33$

Third Quartile, Q_3 = Size of $\left(\frac{3N}{4}\right)^{th}$ item $=\left(\frac{3\times60}{4}\right)^{th}$ item

= 45^{th} item which lies in Class interval (60 – 80).

Thus,

$$Q_3 = L_1 + \frac{\frac{3N}{4}-c}{f} \times i = 60 + \frac{45-29}{20} \times 20$$

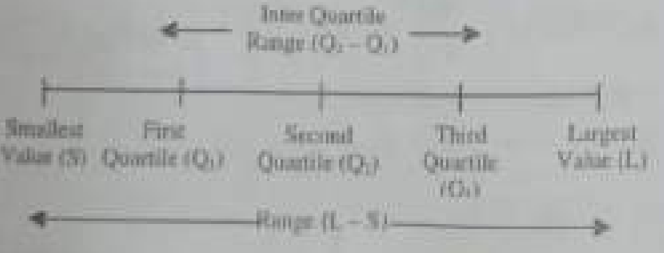
$$= 60 + \frac{16}{20} \times 20 = 60 + 16 = 76$$

Inter-Quartile Range = $Q_3 - Q = 76 - 41.33 = 34.67$ years.

3.4. QUARTILE DEVIATION

3.4.1. Introduction

Quartile deviation is another measure of variation which gives the solution to overcome the limitation of range. In a data set, it calculates the spread over the middle half of the values. This measure of variation minimises the effect of extreme values (also known as outliers). In this method, the study of **Interquartile range** is necessary because a large amount of values in the data set lie in the central part of the frequency distribution. To calculate this value, all the data set is divided into four parts. Every part of the data set contains 25% of the observed value. In these values the highest one is known as quartile.



The half of the difference between third quartile and first quartile ($Q_3 - Q_1$) is known as 'semi-interquartile or quartile deviation'.

Mathematically,

$$\text{Quartile Deviation, Q.D.} = \frac{Q_3 - Q_1}{2}$$

Where, $(Q_3 - Q_1)$ = Interquartile range

3.4.2. Coefficient of Quartile Deviation

If the difference of the third and first quartiles is divided by the sum of the third and first quartiles then it is known as the 'coefficient of quartile deviation'. In case of open-ended distribution (when the frequency distribution has the extreme class limits and no specific class limits), it is a very useful measure.

Mathematically, coefficient of quartile deviation is defined as follows:

$$\text{Coefficient of Quartile Deviation (CQD)} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

Where,

CQD = Coefficient of Quartile Deviation
Q_3 = Third quartile
Q_1 = First quartile

3.4.3. Methods of Calculation of Quartile Deviation

1) Computation of Quartile Deviation - Individual Series
2) Computation of Quartile Deviation - Discrete Series
3) Computation of Quartile Deviation - Continuous Series

3.4.3.1. Computation of Quartile Deviation - Individual Series

Following are the formula for calculating Q_1 and Q_3 in individual series.

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

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

Where, N = number of observation.

Example 5: Compute Quartile Deviation and Coefficient of Quartile Deviation for the following data:
24, 14, 21, 100, 104, 68, 34, 106, 100, 72, 16, 21, 14, 21, 72, 100, 106, 34, 100

Solution: First we arrange the following data in ascending order:

X	f	c.f.
14	2	2
16	1	3
21	3	6
24	1	7
34	2	9

68	1	10
72	2	12
100	4	16
104	1	17
106	2	19

$Q_1 = \text{Size of } \left(\frac{N+1}{4}\right)^{th} \text{ item or } \left(\frac{19+1}{4}\right)^{th} \text{ item}$

$= 5^{th} \text{ item} = 21$

$Q_3 = \text{Size of } \left[\frac{3(N+1)}{4}\right]^{th} \text{ item or } \left[\frac{3(19+1)}{4}\right]^{th} \text{ item}$

$= 15^{th} \text{ item} = 100$

$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{100-21}{2} = 39.5$

$\text{Coefficient of Quartile Deviation} = \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$= \frac{100-21}{100+21} = \frac{79}{121} = 0.653$

3.4.3.2. Computation of Quartile Deviation- Discrete Series

Following are the formula for calculating the Q_1 and Q_3 in discrete series:

$Q_1 = \text{Size of} \left(\frac{N+1}{4}\right)^{th} \text{item}$ and

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

Where, N = number of observation.

Example 6: Find the quartile deviation and coefficient of quartile deviation from the given data:

No. of Persons in a Family	1	2	3	4	5	6	7	8	9	10
No. of Families	2	4	16	12	28	31	8	22	9	7

Solution:

No. of Persons in a Family	No. of Families (f)	Cumulative Frequency
1	2	2
2	4	6
3	16	22
4	12	34
5	28	62
6	31	93
7	8	101
8	22	123
9	9	132
10	7	139
	N = 139	

$Q_1 = \text{Size of} \left(\frac{N+1}{4}\right)^{th} \text{item} = \left(\frac{139+1}{4}\right)^{th} \text{item}$

$= 35^{th} \text{ item} = 5$

$Q_3 = \text{Size of} \left[\frac{3(N+1)}{4}\right]^{th} \text{item} = \left[\frac{3(139+1)}{4}\right]^{th} \text{item}$

$= 105^{th} \text{ item} = 8$

$\text{Quartile Deviation} = \frac{(Q_3 - Q_1)}{2} = \frac{8-5}{2} = 1.5$

Coefficient of Quartile Deviation (CQD)

$= \frac{(Q_3 - Q_1)}{(Q_3 + Q_1)} = \frac{8-5}{8+5} = \frac{3}{13} = 0.23$

3.4.3.3. Computation of Quartile Deviation - Continuous Series (Exclusive)

Following are the formulae for calculating the Q_1 and Q_3 in continuous series:

$$Q_1 = L_1 + \frac{\frac{N}{4} - c}{f} \times i \quad \text{and} \quad Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i$$

Where, L_1 = Lower limit;
N = Number of observation,
c = cumulative frequency of class preceding to quartile class;
i = class interval

Following example explains the calculation of quartile deviation in Continuous series.

Example 7: Calculate Quartile Deviation from the following data:

Population (in thousand)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of Localities	4	7	15	12	3	9	6

Solution:

Population	No. of Localities (f)	Cumulative Frequency
10-20	4	4
20-30	7	11
30-40	15	26
40-50	12	38
50-60	3	41
60-70	9	50
70-80	6	56

$Q_1 = \text{Size of} \left(\frac{N}{4}\right)^{th} \text{item} = \left(\frac{56}{4}\right)^{th} \text{item} = 14^{th} \text{ item}$

which lies in class interval 30-40.

Here, $L_1 = 30$, $c = 11$, $f = 15$ and $i = 10$.

So,

$$Q_1 = L_1 + \frac{\frac{N}{4} - c}{f} \times i = 30 + \frac{14-11}{15} \times 10$$

$= 30 + \frac{30}{15} = 32$

$Q_3 = \text{Size of} \left(\frac{3N}{4}\right)^{th} \text{item} = \left(\frac{3 \times 56}{4}\right)^{th} \text{item} = 42^{th} \text{ item}$

which lies in class interval 60-70.

Here, $L_1 = 60$, $c = 41$, $f = 9$ and $i = 10$.

So. $Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i = 60 + \frac{42-41}{9} \times 10$

$= 60 + 1.1 = 61.1$

Quartile Deviation $= \frac{Q_3 - Q_1}{2} = \frac{61.1-32}{2} = \frac{29.1}{2} = 14.55$

B.com First Semester

Subject Business Statistics

Solved Problems on Quartiles: (Individual Series)

Problem 1: Find Quartile 1 for the given data: 10, 30, 5, 12, 20, 40, 25, 15, 18.

Solution:

Step 1: Sort the given data in order (ascending order / descending order)
5, 10, 12, 15, 18, 20, 25, 30, 40

Step 2: Find 1st Quartile
First Quartile =(n+1/4) th term

Here n = 9 because there are total 9 numbers in the given data.
⇒ First Quartile = ((9 + 1)/4)th term
⇒ First Quartile = (10/4)th term
⇒ First Quartile = 2.5th term

Now, 2.5th term = 2nd term + (0.5) (3rd term - 2nd term)
⇒ 2.5th term = (10) + (0.5) (12 - 10)
⇒ 2.5th term = 10+1
⇒ 2.5th term = 11

The First Quartile value is 11.

Problem 2: Find the Second Quartile for the data 10, 30, 5, 12, 20, 40, 25, 15, 18.

Step 1: Sort the given data in the ascending order

5, 10, 12, 15, 18, 20, 25, 30, 40

Step 2: Find 2nd Quartile

Second Quartile $= [2(\frac{n+1}{4})^{th}]$ *term*

Here n = 9 because there are total 9 numbers in the given data.
⇒ Second Quartile $= (\frac{9+1}{2})^{th}$ *term*
⇒ Second Quartile = (10/2)th term
⇒ Second Quartile = 5th term
5th term is 18

So the Second Quartile value is 18.

Problem 3: Find the third Quartile for the data 10, 30, 5, 12, 20, 40, 25, 15, 18.

Step 1: Sort the given data in the ascending order
5, 10, 12, 15, 18, 20, 25, 30, 40

Step 2: Find 3rd Quartile

Third Quartile $= \frac{3(n+1)^{th}}{4}$ *term*

Here n = 9 because there is total 9 numbers in the given data.

⇒ *Third Quartile* $= \frac{3(n+1)}{4}^{th}$ *term*

⇒ *Third Quartile*= $\frac{3\times(10)}{4}^{th}$ *term*

⇒ *Third Quartile= 7.5th term*

7.5th term is average result of 7th and 8th term = (25 + 30)/2 = 27.5
Remember: 7.5th term = 7th term + (0.5) (8th term - 7th term)

The most recommended method to find value is mentioned above
Because the term not always N.5 something it may vary from N.1 to N.9
Here, N be any natural number.

So, the third Quartile value is 27.5.

Problem 4: Find the first, second, and third quartiles for the data 8, 5,15, 20, 18, 30, 40, 25

Step 1: Sort the given data in the ascending order

5, 8, 15, 18, 20, 25, 30, 40.

Step 2: Find all Quartiles step by step

First Quartile= {(n + 1)/4}th term

Here n = 8 because there is total 8 numbers in the given data.

⇒ First Quartile = {(8 + 1)/4}th term
⇒ First Quartile= {9/4})th term
⇒ First Quartile= 2.25th term

Thus, 2.25th Term = 2nd term + (0.25) (3rd term - 2nd term)

⇒ 2.25th Term = 8+(0.25) (15-8) = 9.75

First Quartile value is 9.75

Second Quartile = {(n + 1)/2}th term

⇒ Second Quartile = (9 + 1)/2}th term
⇒ Second Quartile = {10/2}th term
⇒ Second Quartile = 5th term

5th term is 20

So, the second Quartile value is 20.

Third Quartile = 3(n + 1)/4th term

⇒ Third Quartile = (3(8 + 1)/4) th term
⇒ Third Quartile = (27/4)th term
⇒ Third Quartile = 6.75th term

Thus, 6.75th = 6th term +(0.75) (7th -6th)

⇒ 6.75th = 25+ (0.75) (5) = 28.75

So, the third Quartile value is 28.75

Problem 5: What is the Interquartile Range for the data if the first quartile is 10 and the third quartile is 30cm?

Solution:

Given,

- *Q1 = 10*
- *Q3 = 30*

Interquartile range = Q3 - Q1
⇒ Interquartile range = 30 – 10 = 20

Examples of Quartiles in Discrete Series

Example 1:

Height (in cm)	153	155	157	159	161	163	165	167	169
No. of Students	6	2	4	6	3	4	7	1	4

Height (in cm)	No. of Students (f)	Cumulative Frequency (c.f.)
153	6	6
155	2	10
157	4	14
159	6	20
161	3	23
163	4	27
165	7	34
167	1	35
169	4	39
	$\sum = 39$	

Q1=Size of 4*N*+1*th item=Size of* 439+1*th item=Size of* 10*th item*

Q1 = 155 centimetres

Q3=Size of 3[N+14]th item=Size of 3[39+14]th item=Size of 30th item*Q*3=*Size of* 3[4*N*+1]*th item=Size of* 3[439+1]*th item=Size of* 30*th item*

Q3 = 163 centimetres

Interquartile Range = Q3 - Q1 = 163- 155 = 8

Quartile Deviation = Q3−Q12=163−1552=42*Q*3−*Q*1=2163−155=4

Example 2:

Marks (X)	2	4	6	8	10	12
No. of Students (f)	4	1	2	3	4	5

Marks (X)	No. of Students (f)	Cumulative Frequency (c.f.)
2	4	4
4	1	5
6	2	7
8	3	10
10	4	14
12	5	19
	∑ = 19	

*Q*1=*Size of* 4*N*+1*th item*=*Size of* 419+1*th item*=*Size of* 5*th item*

Q1 = 4

Q3=Size of 3[N+14]th item=Size of 3[19+14]th item=Size of 15th item*Q*3=*Size of* 3[4*N*+1
]*th item*=*Size of* 3[419+1]*th item*=*Size of* 15*th item*

Q3 = 12

Interquartile Range = Q3 - Q1 = 12 - 4 = 8

Quartile Deviation = Q3−Q12=12−42=42*Q*3−*Q*1=212−4=4

Coefficient of Quartile Deviation = Q3−Q1Q3+Q1=12−812+8=0.2*Q*3+*Q*1*Q*3−*Q*1=12+812−8=0.2

Examples of Quartiles in Grouped (Continuous) Series:

1. Calculate Quartile-3 from the following grouped data

X	Frequency
0	1
1	5
2	10
3	6
4	3

Solution:

x	Frequency *f*	*cf*
0	1	1
1	5	6
2	10	16
3	6	22
4	3	25
---	---	---
	n = 25	--

Here, n=25

$Q3$=(3(n+1)4)th value of the observation

=(3⋅264)th value of the observation

=(19.5)th value of the observation

=3

2. Calculate Quartile-1 from the following grouped data

X	Frequency
10	3
11	12
12	18
13	12
14	3

Solution:

x	**Frequency** ***f***	***cf***
10	3	3
11	12	15
12	18	33
13	12	45
14	3	48
---	---	---
	n = 48	--

Here, $n=48$

$Q1=(n+14)th$ value of the observation

$=(494)th$ value of the observation

$=(12.25)th$ value of the observation

$=11$

3. Calculate Quartile-3 from the following grouped data

Class	Frequency
2 - 4	3
4 - 6	4
6 - 8	2
8 - 10	1

Solution:

Class	**Frequency** f	cf
2 - 4	3	3
4 - 6	4	7
6 - 8	2	9
8 - 10	1	10
---	---	---
	n = 10	--

Here, n=10

$Q3$ class :

Class with $(3n4)th$ value of the observation in cf column

$=(3\cdot 104)th$ value of the observation in cf column

$=(7.5)th$ value of the observation in cf column

and it lies in the class 6-8.

$\therefore Q3$ class : 6-8

The lower boundary point of 6-8 is 6.

$\therefore L=6$

$Q3=L+3n4-cff\cdot c$

$=6+7.5-72\cdot 2$

$=6+0.52\cdot 2$

$=6+0.5$

$=6.5$

---

4. Calculate Quartile-1 from the following grouped data

Class	Frequency
0 - 2	5
2 - 4	16
4 - 6	13
6 - 8	7
8 - 10	5
10 - 12	4

Solution:

Class	**Frequency** ***f***	***cf***
0 - 2	5	5
2 - 4	16	21
4 - 6	13	34
6 - 8	7	41
8 - 10	5	46
10 - 12	4	50
---	---	---
	n = 50	--

Here, *n*=50

*Q*1 class :

Class with (*n*4)*th* value of the observation in *cf* column

=(504)*th* value of the observation in *cf* column

=(12.5)*th* value of the observation in *cf* column

and it lies in the class 2-4.

∴*Q*1 class : 2-4

The lower boundary point of 2-4 is 2.

∴*L*=2

*Q*1=*L*+*n*4-*cff*·*c*

=2+12.5-516·2

=2+7.516·2

=2+0.9375

=2.9375

5. Calculate Quartile-3 from the following grouped data

Class	Frequency
10 - 20	15
20 - 30	25
30 - 40	20
40 - 50	12
50 - 60	8
60 - 70	5
70 - 80	3

Solution:

Class	**Frequency** f	cf
10 - 20	15	15
20 - 30	25	40
30 - 40	20	60
40 - 50	12	72
50 - 60	8	80
60 - 70	5	85
70 - 80	3	88
---	---	---
	n = 88	--

Here, n=88

*Q*3 class :

Class with (3*n*4)*th* value of the observation in *cf* column

=(3·884)*th* value of the observation in *cf* column

=(66)*th* value of the observation in *cf* column

and it lies in the class 40-50.

∴*Q*3 class : 40-50

The lower boundary point of 40-50 is 40.

∴*L*=40

*Q*3=*L*+3*n*4-*cff*·*c*

=40+66-6012·10

=40+612·10

=40+5

=45

---

6. Calculate Quartile-1 from the following grouped data

Class	Frequency
20 - 25	110
25 - 30	170
30 - 35	80
35 - 40	45
40 - 45	40
45 - 50	35

Solution:

Class	**Frequency** ***f***	***cf***
20 - 25	110	110
25 - 30	170	280
30 - 35	80	360
35 - 40	45	405
40 - 45	40	445
45 - 50	35	480
---	---	---
	n = 480	--

Here, *n*=480
*Q*1 class :
Class with (*n*4)*th* value of the observation in *cf* column

=(4804)*th* value of the observation in *cf* column

=(120)*th* value of the observation in *cf* column

and it lies in the class 25-30.

∴*Q*1 class : 25-30
The lower boundary point of 25-30 is 25.
∴*L*=25

*Q*1=*L*+*n*4-*cff*·*c*

=25+120-110170·5

=25+10170·5

=25+0.2941

=25.2941

Calculation of Quartile Deviation in Different Series

1. Individual Series:

Example:

With the help of the data given below, find the interquartile range, quartile deviation, and coefficient of quartile deviation.

150	100	268	280	195	140	200

S.No.	Items arranged in ascending order
1	100
2	140
3	150
4	195
5	200
6	268
7	280
N = 7	

$Q_1 = Size\ of\ [\frac{N+1}{4}]^{th}\ item = Size\ of\ [\frac{7+1}{4}]^{th}\ item = Size\ of\ 2^{nd}\ item$

$Q_1 = 140$

$Q_3 = Size\ of\ 3[\frac{N+1}{4}]^{th}\ item = Size\ of\ 3[\frac{7+1}{4}]^{th}\ item = Size\ of\ 6^{th}\ item$

$Q_3 = 268$

Interquartile Range = $Q_3 - Q_1 = 268 - 140 = 128$

Quartile Deviation = $\frac{Q_3 - Q_1}{2} = \frac{268-140}{2} = 64$

Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{268-140}{268+140} = \frac{128}{408} = 0.31$

Interquartile Range = 128

Quartile Deviation = 64

Coefficient of Quartile Deviation = 0.31

2. Discrete Series:

Example:

From the following table giving marks of students, calculate the interquartile range, quartile deviation, and coefficient of quartile deviation.

Marks of Students	60	62	68	70	75	80	88	90	97
No. of Students	25	21	28	18	24	20	24	17	22

Marks of Students (X)	No. of Students (f)	Cumulative Frequency (c.f.)
60	25	25
62	21	46
68	28	74
70	18	92
75	24	116
80	20	136
88	24	160
90	17	177
97	22	199

$Q_1 = Size\ of\ [\frac{N+1}{4}]^{th}\ item = Size\ of\ [\frac{199+1}{4}]^{th}\ item = Size\ of\ 50^{th}\ item$

$Q_1 = 68$

$Q_3 = Size\ of\ 3[\frac{N+1}{4}]^{th}\ item = Size\ of\ 3[\frac{199+1}{4}]^{th}\ item = Size\ of\ 150^{th}\ item$

$Q_3 = 88$

Interquartile Range = $Q_3 - Q_1 = 88 - 68 = 20$

Quartile Deviation = $\frac{Q_3 - Q_1}{2} = \frac{88-68}{2} = 10$

Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{88-68}{88+68} = \frac{20}{156} = 0.12$

Interquartile Range = 20

Quartile Deviation = 10

Coefficient of Quartile Deviation = 0.12

3. Continuous Series:

Example:

Calculate interquartile range, quartile deviation, and coefficient of quartile deviation from the following figures:

Size (X)	0-10	10-20	20-30	30-40	40-50	50-60
Frequency (f)	9	15	29	24	3	20

Size (X)	Frequency (f)	Cumulative Frequency (c.f.)
0-10	9	9
10-20	15	24
20-30	29	53
30-40	24	77
40-50	3	80
50-60	20	100
	N = ∑f = 100	

$Q_1 = Size\ of\ [\frac{N}{4}]^{th}\ item = Size\ of\ [\frac{100}{4}]^{th}\ item = Size\ of\ 25^{th}\ item$

Q_1 lies in the group 20-30

$l_1 = 20, c.f. = 24, f = 29, i = 10$

$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 20 + \frac{25-24}{29} \times 10 = 20 + 0.34$

$Q_1 = 20.34$

$Q_3 = Size\ of\ [\frac{3N}{4}]^{th}\ item = Size\ of\ [\frac{3 \times 100}{4}]^{th}\ item = Size\ of\ 75^{th}\ item$

Q_3 lies in the group 30-40

$l_1 = 30, c.f. = 53, f = 24, i = 10$

$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 30 + \frac{75-53}{24} \times 10 = 30 + 9.16$

$Q_3 = 39.16$

Interquartile Range = $Q_3 - Q_1 = 39.16 - 20.34 = 18.82$

Quartile Deviation = $\frac{Q_3 - Q_1}{2} = \frac{39.16 - 20.34}{2} = 9.41$

Coefficient of Quartile Deviation = $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{39.16 - 20.34}{39.16 + 20.34} = \frac{18.82}{59.5} = 0.31$

Interquartile Range = 18.82

Quartile Deviation = 9.41

Coefficient of Quartile Deviation = 0.31

2.7. PERCENTILE

2.7.1. Introduction

If the values of the observation in a data set are arranged in ascending or descending order and divided it into hundred equal parts by using ninety nine points on the scale of observations then it is called **percentile**. If the data is arranged in ascending order, the first point is known as 'first percentile', denoted by P_1, the second point is called 'second percentile', denoted by P_2, and so on, the last point is called 'ninety-ninth percentile' and is denoted by P_{99}.

2.7.2. Advantages of Percentile

Percentile has the following advantages:

1) It is simple to understand and easy to calculate.
2) Decile is less affected by the extreme values.
3) It can be calculated exactly in case of open-end classes.

2.7.3. Disadvantages of Percentile

Disadvantages of percentile are as follows:

1) Percentiles are not based on all observations. So it may not be a good representative.

2) It is not capable of further algebraic treatment
3) Its value is affected to a greater extent by fluctuations of sampling

2.7.4. Calculation of Percentiles

1) Calculation of Deciles – Individual and Discrete Series
2) Calculation of Deciles – Grouped Series

2.7.4.1. Calculation of Deciles – Individual and Discrete Series

Formula:

$$\left.\begin{array}{l} P_1 = \left(\frac{N+1}{100}\right)^{th} \text{item} \\ P_2 = \left[\frac{2N+1}{100}\right]^{th} \text{item} \\ \ldots\ldots \\ P_{99} = \left[\frac{99(N+1)}{100}\right]^{th} \text{item} \end{array}\right\} \text{For Individual and Discrete Series}$$

Example 32: Find P_{60} and P_{70} from the following data:

80	48	26	42	65	16	58	75	98

Solution: Arranging the data in ascending order, we get:

16	26	42	48	58	65	75	80	98

$$P_{60} = \text{Size of} \left[\frac{60(N+1)}{100}\right]^{th} \text{item} = \left[\frac{60(9+1)}{100}\right]^{th} \text{item} = 6^{th}$$

item = 65

$$P_{70} = \text{Size of} \left[\frac{70(N+1)}{100}\right]^{th} \text{item} = \left[\frac{70(9+1)}{100}\right]^{th} \text{item} = 7^{th}$$

item = 75

2.7.4.2. Calculation of Deciles – Grouped Series

Formula: $$P_k = L_1 + \frac{\frac{kN}{100} - c}{f} \times (L_2 - L_1)$$

Where, k = 1, 2, 3, 4, 5 .... 100;
i = $(L_2 - L_1)$ = k^{th} quartile class
N = total frequency;
f = frequency of k^{th} quartile class
c = frequency of the class preceding the k^{th} quartile class.

Obviously:

$$\left.\begin{array}{l} P_1 = L_1 + \frac{\frac{N}{100} - c}{f} \times i \\ P_2 = L_1 + \frac{\frac{2N}{100} - c}{f} \times i \\ \ldots\ldots \\ P_{99} = L_1 + \frac{\frac{99N}{100} - c}{f} \times i \end{array}\right\} \text{(For Grouped Series)}$$

4) Use the following formula to locate quartiles:

Formula: $$Q_k = L_1 + \frac{\frac{kN}{4} - c}{f}(L_2 - L_1)$$

Where, $k = 1, 2, 3$; $i = (L_2 - L_1) = k^{th}$ quartile class,
N = total frequency; f = frequency of k^{th} quartile class,
c = cumulative frequency of the class preceding the k^{th} quartile class.

Obviously:

$$Q_1 = L_1 + \frac{\frac{N}{4} - c}{f} \times i$$

$$Q_2 = L_1 + \frac{\frac{2N}{4} - c}{f} \times i$$

$$Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i$$

The values of L_1, L_2, c, i, f, are taken according to respective quartiles

1. Calculate Quartile-3 from the following grouped data

X	Frequency
0	1
1	5
2	10
3	6
4	3

2. Calculate Quartile-1 from the following grouped data

X	Frequency
10	3
11	12
12	18
13	12
14	3

3. Calculate Quartile-3 from the following grouped data

Class	Frequency
2 - 4	3
4 - 6	4
6 - 8	2
8 - 10	1

4. Calculate Quartile-1 from the following grouped data

Class	Frequency
0 - 2	5
2 - 4	16
4 - 6	13
6 - 8	7
8 - 10	5
10 - 12	4

5. Calculate Quartile-3 from the following grouped data

Class	Frequency
10 - 20	15
20 - 30	25
30 - 40	20
40 - 50	12
50 - 60	8
60 - 70	5
70 - 80	3

6. Calculate Quartile-1 from the following grouped data

Class	Frequency
20 - 25	110
25 - 30	170
30 - 35	80
35 - 40	45
40 - 45	40
45 - 50	35

* Range :-

① $L - S$ ② Coeff $= \frac{L-S}{L+S}$ [illegible]

* Standard Deviation

① Individual series $\sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2}{N}}$

Σ = Summation

$\bar{x}$ = [illegible], $\bar{x} = \frac{\Sigma x}{N}$, $x - \bar{x}$ [illegible]

Q. 11, 14, 15, 17, 18

x	$x-\bar{x}$	$(x-\bar{x})^2$
11	[illegible]	[illegible]
14	[illegible]	[illegible]
15	[illegible]	[illegible]
17	[illegible]	[illegible]
18	[illegible]	[illegible]
		[illegible]

$\sigma = \sqrt{\frac{\Sigma (x-\bar{x})^2}{N}} = \sqrt{\frac{30}{5}} = \sqrt{6} = 2.74$

Coeff of S.D. $= \frac{S.D.}{\bar{x}}$ $= 0.163$ (ii) Coeff of variation $= \frac{S.D.}{\bar{x}} \times 100$ $= \frac{2.74}{15} \times 100 = 16.32$

Q. The scores of Batsmen [illegible] in 10 innings were 38, 70, 48, 34, 42, 55, 63, 46, 54, 44. Find the [illegible] of the scores [illegible]

x	d	d^2
38	[illegible]	[illegible]
70	[illegible]	
48	[illegible]	
34		
42		
55		
63		
46		
54		
44		

$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$ = [illegible]

= [illegible]

= [illegible]

$\bar{x}$ = [illegible]

= [illegible]

σ = [illegible]

If mean not a whole no.

$\sigma = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$

$d = x - A$

[illegible] $\frac{\Sigma d^2}{\Sigma d / N}$

Individual series

Mean whole no. S.D. [illegible]

$\sqrt{\frac{\Sigma (x-\bar{x})^2}{N}}$ Variance $= (S.D.)^2$ $\sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$

$d = x - A$

Coeff of S.D. $= \frac{S.D.}{\bar{x}}$

Coeff of variance $= \frac{S.D.}{\bar{x}} \times 100$ [illegible]

② Discrete series

[illegible]

$\sqrt{\frac{\Sigma f (x-\bar{x})^2}{N}}$, $\sqrt{\frac{\Sigma f d^2}{N} - \left(\frac{\Sigma f d}{N}\right)^2}$ Variance $= (S.D.)^2$, Coeff [illegible] $= \frac{S.D.}{\bar{x}}$, [illegible] $= \frac{S.D.}{\bar{x}} \times 100$

③ Continuous series

$\sigma = \sqrt{\frac{\Sigma f d'^2}{N} - \left(\frac{\Sigma f d'}{N}\right)^2} \times i$ ← Step deviation method

d [illegible] (mid)

$d' = \frac{x - A}{i}$

C.I. [illegible]

Measures of Dispersion

① Discrete Series of S.D

Q. find S.D

x	1	2	3	4	5	6	7
f	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8

Q. [illegible]

Mean = $\frac{\Sigma fx}{N}$ [illegible]

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

[illegible]

③ Continuous Series of S.D

Q. find the S.D

Marks: [illegible]

Step deviation [illegible]

Assumed mean [illegible]

$\bar{X} = \frac{\Sigma fx}{N}$ [illegible]

$d' = \frac{x - A}{i}$ [illegible]

$$\text{Mean}(\bar{X}) = \frac{\sum fX}{\sum f} = \frac{2800}{70} = 40$$

$$\text{Mean Deviation (M.D.)} = \frac{\sum f|d_x|}{\sum f} = \frac{1040}{70} = 14.86$$

3.6. STANDARD DEVIATION (S.D.)

3.6.1. Introduction

Standard deviation is **used to measure the spread of** items in a set of observation. If all the observation values are identical or distribution of items of a set is uniform, then deviation of every value from mean is zero. When elements of the set are more dispersed, then standard deviation becomes larger. It is denoted by the small Greek letter σ (Sigma). If $\bar{X}$ is the mean of $X_1, X_2, \text{---}, X_N$, then,

$$\sigma = \sqrt{\frac{1}{N}\left[(X_1 - \bar{X})^2 + \text{\_\_\_\_\_\_} + (X_n - \bar{X})^2\right]} = \sqrt{\sum_{i=1}^{n}\frac{(X_i - \bar{X})^2}{N}}$$

According to Yule and Kenndall, "The standard deviation is the square-root of the arithmetic mean of the square of all the deviations, deviations being measured from the arithmetic mean of the observation."

3.6.2. Properties of Standard Deviation

1) Standard deviation is always positive.
2) Standard deviation will not change if all the variables are summated with the same number.
3) If researcher multiplies the variable with the same number, then the output received will be multiplied by the square of that same number.
4) In case of multiple distribution (with same mean and known standard deviation), SD can be calculated as:
   i) If samples size are same
   $$\sigma = \sqrt{\frac{\sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2}{N}}$$
   ii) If samples size are different
   $$\sigma = \sqrt{\frac{k_1\sigma_1^2 + k_2\sigma_2^2 + ... + k_N\sigma_N^2}{k_1 + k_2 + ... + k_N}}$$

3.6.3. Advantages of Standard Deviation

1) S.D. is rigidly defined and considers all the observations for calculation.
2) It contains various mathematical properties and is used in algebric calculation.
3) Combined S.D. can be calculated for two or more groups.
4) Compared to other measures of dispersion, standard deviation is less affected by sampling fluctuation.
5) It is used to calculate coefficient of variation.

3.6.4. Disadvantages of Standard Deviation

1) More complex to calculate and understand, compared to other measures of dispersion.
2) Distance from the mean decides the weightage of the element. If distance (from the mean) is more, then it is given more weight and if distance is less (from mean), then that element is given less weight.

3.6.5. Methods of Calculation of Standard Deviation

1) Calculation of Standard Deviation – Individual Series
2) Calculation of Standard Deviation – Discrete series
3) Calculation of Standard Deviation – Continuous Series

3.6.5.1. Calculation of Standard Deviation – Individual Series

There are two methods of calculating standard deviation in an individual observation or series:

1) **Deviation Taken from Actual Mean:** This method is adopted when the mean is a whole number. The following are the steps:
   i) Find out the actual mean of the series.
   ii) Find out the deviation of each value from the mean (d_x or $x = X - \bar{X}$).
   iii) Square the deviations and take the total of squared deviations $\sum x^2$.
   iv) Divide the total ($\sum x^2$) by the number of observations. The square root of the quotient is standard deviation. Thus apply the following formula:
   $$\sigma = \sqrt{\frac{\sum x^2}{N}} \text{ or } \sqrt{\frac{\sum(X - \bar{X})^2}{N}}$$
2) **Deviation Taken from Assumed Mean (Short-cut Method):** This method is adopted when the arithmetic average is a fractional value. Taking deviations from fractional value would be a very difficult and tedious task. To save time and labour, we apply short cut method; deviations are taken **from an assumed mean.** The following are the steps:
   i) Assume any one of the item in the series as an average (A).
   ii) Find out the deviations from the assumed mean; i.e., (X – A) denoted by d_x.
   iii) Find out the total of the deviations; i.e., $\sum d_x$.
   iv) Square the deviations; i.e., d_x^2 and add up the squares of deviations, i.e., $\sum d_x^2$.
   v) Then substitute the values in the following formula:
   $$\sigma = \sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2}$$

   Where,

   d_x stands for the deviation from assumed mean = (X – A).

Example 11: Find S.D of (₹) 8, 10, 15, 24, 28.

Solution: Let the assumed mean (A) = 16.

Calculation from Arithmetic Mean

X	$x = (X - \bar{X})$	x^2
8	-9	81
10	-7	49
15	-2	4
24	7	49
28	11	121
ΣX = 85		$\Sigma x^2 = 304$

Calculation from Assumed Mean

X	$d_x = (X - A)$	d_x^2
8	-8	64
10	-6	36
15	-1	1
24	8	64
28	12	144
	5	$\Sigma d_x^2 = 309$

From Arithmetic Mean

$$\bar{X} = \frac{\sum X}{N} = \frac{85}{5} = 17$$

$$\sigma = \sqrt{\frac{\sum x^2}{N}} \text{ or } \frac{\sum (X-\bar{X})^2}{N}$$

$$= \sqrt{\frac{1}{5} \times 304} = \sqrt{60.8} = 7.8$$

From Assumed Mean

Let A (assumed mean) = 16

$$\sigma = \sqrt{\frac{\sum d_x^2}{N} - \left(\frac{\sum d_x}{N}\right)^2} = \sqrt{\frac{309}{5} - \left(\frac{5}{5}\right)^2}$$

$$= \sqrt{61.8 - (1)^2} = \sqrt{60.8} = 7.8$$

Note: If the actual mean is in fraction, then it is better to take deviations from an assumed mean for avoiding too much calculation.

3.6.5.2. Calculation of Standard Deviation – Discrete Series (Ungrouped Data)

There are three methods for calculating standard deviation in discrete series:

1) **Actual Mean Method:** It includes the following steps:
   i) Calculate the mean of the series.
   ii) Find deviations for various items from the mean i.e., $(X - \bar{X}) = d_x$.
   iii) Square the deviations d_x^2 and multiply by the respective frequencies (f). We get fd_x^2.
   iv) Total the product (Σfd_x^2). Then apply the formula.

$$\sigma = \sqrt{\frac{\Sigma f d_x^2}{\Sigma f}}$$

Example 12: Calculate standard deviation from the following series:

Age (in years)	15	25	35	45	55	65
No. of Persons	7	25	20	16	11	6

Solution: Calculation of Standard Deviations

Age (X)	No. of Persons (f)	fX	$d_x = (X - \bar{X})$	fd_x^2
15	7	105	-22	3388
25	25	625	-12	3600
35	20	700	-2	80
45	16	720	8	1024
55	11	605	18	3564
65	6	390	28	4704
	Σf = 85	ΣfX = 3145		$\Sigma fd_x^2 = 16360$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{3145}{85} = 37$$

Standard Deviation (σ) =

$$\sqrt{\frac{\sum f(d_x)^2}{\sum f}} = \sqrt{\frac{16360}{85}} = \sqrt{192.5} = 13.87$$

2) **Assumed Mean Method:** We applied short-cut method in individual observation. Similarly we can apply it in discrete series also. Here deviations are taken, not from the actual mean, but from an assumed mean. It involves the following steps:
   i) Assume any one of the items in the series as an average and this is called assumed average and denoted by A.
   ii) Find out the deviations from assumed mean, i.e, (X – A), and denote it by d_x.
   iii) Multiply these deviations by the respective frequencies and get the $\sum fd_x$.
   iv) Square the deviations (d_x^2).
   v) Multiply the squared deviations (d_x^2) by the respective frequencies (f) and get $\sum fd_x^2$.
   vi) Substitute the values in the following formula:

$$\sigma = \sqrt{\frac{\Sigma f d_x^2}{\sum f} - \left(\frac{\Sigma f d_x}{\sum f}\right)^2} \quad \text{Where } d_x = X - A$$

Example 13: Find the Standard Deviation of the following series:

X	10	11	12	13	14	Total
f	4	16	22	14	6	62

Solution: Let the assumed mean (A) = 12.

Calculation of Standard Deviations

X	f	$d_x = (X - A)$	d_x^2	fd_x	fd_x^2
10	4	-2	4	-8	16
11	16	-1	1	-16	16
12	22	0	0	0	0
13	14	1	1	14	14
14	6	2	4	12	24
Total	Σf = 62			$\Sigma fd_x = 2$	$\Sigma fd_x^2 = 70$

$$\sigma = \sqrt{\frac{\sum f(d_x)^2}{\sum f} - \left(\frac{\sum fd_x}{\sum f}\right)^2} = \sqrt{\frac{70}{62} - \left(\frac{2}{62}\right)^2}$$

$$= \sqrt{1.13 - 0.001} = 1.06$$

3) **Step Deviation Method:** Here we take a common factor for all the items of the series. In this method the

calculation becomes easy and simple. The formula for this is

$$\sigma = \sqrt{\frac{\sum fd_x'^2}{\sum f} - \left(\frac{\sum fd_x'}{\sum f}\right)^2} \times i$$

Where, $d_x' = \frac{X-A}{i}$, i = Common Factor.

Example 14: Find the Standard Deviation for the following distribution:

X	4.5	14.5	24.5	34.5	44.5	54.5	64.5
f	5	3	7	18	14	9	4

Solution: Let the assumed mean (A) = 34.5.

Calculation of Standard Deviation

X	f	$d_x = (X-A)$	$d_x' = \left(\frac{d_x}{10}\right)$	fd_x'	$f(d_x')^2$
4.5	5	−30	−3	−15	45
14.5	3	−20	−2	−6	12
24.5	7	−10	−1	−7	7
34.5	18	0	0	0	0
44.5	14	10	1	14	14
54.5	9	20	2	18	36
64.5	4	30	3	12	36
	$\sum f = 60$			$\sum fd_x' = 16$	$\sum f(d_x')^2 = 150$

$$\sigma = \sqrt{\left[\frac{\sum f(d_x')^2}{\sum f} - \left(\frac{\sum fd_x'}{\sum f}\right)^2\right]} \times i$$

$$= \sqrt{\frac{150}{60} - \left(\frac{16}{60}\right)^2} \times 10 = \sqrt{2.5 - 0.071} \times 10 = 1.56 \times 10$$

$= 15.6$

3.6.5.3. Calculation of Standard Deviation – Continuous Series (Grouped Data)

In the continuous series the method of calculating standard deviation is almost the same as in a discrete frequency distribution. But in a continuous series, mid-values of the class intervals are to be found out. The step deviation method is widely used. It involves the following steps:

1) Find out the mid-value of each group or class.
2) Assume one of the mid-values as an average and denote it by A.
3) Find out deviation of each mid-value from the assumed average A and denote these deviations by d_x.
4) If the class-intervals are equal, then take a common factor. Divide each deviation by the common factor and denote this column by d_x'.
5) Multiply these deviations d_x' by the respective frequencies and get $\sum fd_x'$.
6) Square the deviations and get $d_x'^2$.
7) Multiply the squared deviation ($d_x'^2$) by the respective frequencies (f). Then obtain the total $\sum fd_x'^2$.
8) Substitute the values in the following formula to get the standard deviation.

$$\sigma = \sqrt{\frac{\sum f d_x'^2}{\sum f} - \left(\frac{\sum f d_x'}{\sum f}\right)^2} \times i$$

Where, $d_x' = \frac{X-A}{i}$, i = Common factor

Example 15: Find the S.D. from the following figures:

Height (Inches)	44-46	46-48	48-50	50-52	52-54	Total
No. of Children	5	25	28	22	5	85

Solution: Let A (assumed mean) = 49

Height (Inches)	Mid-point (X)	No. of Children (f)	d_x = X - 49	d_x' = d/2	fd_x'	$fd_x'^2$
44-46	45	5	−4	−2	−10	20
46-48	47	25	−2	−1	−25	25
48-50	49	28	0	0	0	0
50-52	51	22	2	1	22	22
52-54	53	5	4	2	10	20
Total		$\sum f = 85$			$\sum fd_x' = -3$	$\sum fd_x'^2 = 87$

$$\sigma = \sqrt{\left[\frac{\sum f(d_x')^2}{\sum f} - \left(\frac{\sum d_x'}{\sum f}\right)^2\right]} \times i = \sqrt{\frac{87}{85} - \left(\frac{-3}{85}\right)^2} \times 2$$

$$= \sqrt{\left(\frac{7395-9}{7225}\right)} \times 2 = \sqrt{\frac{7386}{7225}} \times 2 = 1.01 \times 2 = 2.02$$

3.7. VARIANCE

Absolute values are not conducive to easy manipulation; due to this reason mathematicians developed an alternative mechanism for overcoming the zero-sum property of deviations from the mean. This approach utilizes the square of the deviations from the mean. The result is the variance, an important measure of variability. The variance is the average of the squared deviations about the arithmetic mean for a set of numbers. The population variance is denoted by σ^2.

3.7.1. Sample Variance

The sample variance is denoted by s^2. The main use for sample variances is as estimator of population variances. Because of this, computation of the sample variance differs slightly from computation of the population variance. The sample variance uses N – 1 in the denominator instead of N because using n in the denominator of a sample variance results in a statistic that tends to underestimate the population variance. Using N – 1 allows it to be an unbiased estimator, which is a desirable property in inferential statistics.

Sample variance, $s^2 = \frac{\sum(x - \bar{x})^2}{N-1}$

For example, sample of six of the largest accounting firms in the United States and the number of partners associated with each firm as reported by the Public Accounting Report.

Firm	Number of Partners
Price Waterhouse	1062
McGladrey & Pullen	381
Deloitte & Touche	1719
Andersen Worldwide	1673
Coopers & Lybrand	1277
BDO Seidman	217

The sample variance and sample standard deviation can be computed by:

X	$(X-\overline{X})^2$
1062	51.41
381	454,046.87
1719	441,121.79
1673	382,134.15
1277	49,359.51
217	701,959.11
$\sum X = 6329$	$\Sigma(X-\overline{X})^2 = 2,028,672.84$

$$\overline{X} = \frac{6329}{6} = 1054.83$$

$$s^2 = \frac{\sum(X-\overline{X})^2}{N-1} = \frac{2,028,672.84}{5} = 405,734.57$$

The sample variance is 405,734.57.

3.7.2. Population Variance

Because absolute values are not conducive to easy manipulation, mathematicians developed an alternative mechanism for overcoming the zero-sum property of deviations from the mean. This approach utilizes the square of the deviations from the mean. The result is the variance, an important measure of variability. The variance is the average of the squared deviations about the arithmetic mean for a set of numbers. The population variance is denoted by σ^2.

Population variance,

$$\sigma^2 = \frac{\sum(X-\mu)^2}{N}$$

For example, table 3.1 shows the original production numbers for the computer company, the deviations from the mean, and the squared deviations from the mean.

The sum of the squared deviations about the mean of a set of values – called the sum of squares of x and sometimes abbreviated as SS_X – is used throughout statistics. For the computer company, this value is 130. Dividing it by the number of data values (5 weeks) yields the variance for computer production.

$$\sigma^2 = \frac{130}{5} = 26.0$$

Table 3.1: Variance for Computer Production Data

X	X − μ	|X − μ|
5	-8	+8
9	-4	+4
16	+3	+3
17	+4	+4
18	+5	+5
$\sum X = 65$	$\sum(X-\mu) = 0$	$\sum\|X-\mu\| = 24$

$$\sigma^2 = \frac{\sum(|X-\mu|)^2}{N} = \frac{(24)^2}{5} = \frac{576}{5} = 115.2$$

3.8. COEFFICIENT OF VARIATION (CV)

3.8.1. Introduction

It is an absolute measure of dispersion. Coefficient of variation is expressed in terms of units in which actual data is collected and stated. The standard deviation of the weights of people cannot be compared with the standard deviation of heights of the people because both are given in two different units i.e., weights are given in kilogram and heights are given in meters.

For the purpose of comparison, standard deviation is definitely converted into relative measure (known as coefficient of variation) of dispersion. The ratio of the standard deviation to the mean in percentage is known as 'coefficient of variation'.

Mathematically,

$$\text{Coefficient of variation (C.V.)} = \frac{\sigma}{\overline{X}} \times 100$$

According to Karl Pearson, "Coefficient of variation is the percentage variation in mean, standard deviation being considered as the total variation in the mean." That is, it shows the relationship between the standard deviation and the arithmetic mean expressed in terms of percentage.

3.8.2. Advantages of Coefficient of Variation

1) Since, standard deviation of data is always understood in the context of the mean of the data, hence coefficient of variation is very useful.
2) Instead of standard deviation, coefficient of variation is used to compare the data sets with different units or widely different means.

3.8.3. Disadvantages of Coefficient of Variation

1) It gives infinity when the mean value is close to zero and due to this it is very sensitive to small changes in the mean.
2) It cannot be directly used to measure the confidence interval for the mean.

3.8.4. Methods of Calculation of Coefficient of Variation

1) **Calculation of Coefficient of Variation – Individual Series**

Example 16: Calculate Coefficient of Variation (C.V.) for the following data: 56, 66, 61, 68, 54, 70, 55.

Solution: Let the assumed mean is (A) = 60.

Calculation of Coefficient of Variation

X	dx = (X – A)	dx^2
56	–4	16
66	6	36
61	1	1
68	8	64
54	–6	36
70	10	100
55	–5	25
	$\sum dX = 10$	$\sum dX^2 = 278$

$$\overline{X} = A + \frac{\sum dx}{N} = 60 + \frac{(10)}{7} = 60 + 1.43 = 61.43$$

$$S.D.\,(\sigma) = \sqrt{\frac{\sum (dx)^2}{N} - \left(\frac{\sum dx}{N}\right)^2}$$

$$= \sqrt{\frac{278}{7} - \left(\frac{10}{7}\right)^2}$$

$$= \sqrt{\frac{278}{7} - \frac{100}{49}} = \sqrt{\frac{278 \times 7 - 100}{49}}$$

$$= \sqrt{\frac{1846}{49}} = \sqrt{37.67} = 6.14$$

Coefficient of Variation

$$= \frac{\sigma}{\overline{X}} \times 100 = \frac{6.14}{61.43} \times 100 = \frac{614}{61.43} = 9.99$$

2) **Calculation of Coefficient of Variation – Discrete Series (Ungrouped Data)**

Example 17: Calculate coefficient of variation of the following data:

Weekly Rent (₹)	400	700	800	950	1000	1200	1450
No. of Persons Paying the Rent	11	13	34	39	18	8	2

Solution: **Calculation of Coefficient of Variation**

Weekly Rent (X)	No. of Persons (f)	fX	$d_x = (X - \overline{X})$	$f(d_x)$
400	11	4400	– 466	2388716
700	13	9100	–166	358228
800	34	27200	– 66	148104
950	39	37050	84	275184
1000	18	18000	134	323208
1200	8	9600	334	892448
1450	2	2900	584	682112
	$\Sigma f = 125$	$\Sigma fX = 108250$		$\Sigma fd_x^2 = 5068000$

$$\overline{X} = \frac{\sum fX}{\sum f} = \frac{108250}{125} = 866$$

Standard Deviation (σ)

$$= \sqrt{\frac{\sum f(d_x)^2}{\sum f}} = \sqrt{\frac{5068000}{125}} = \sqrt{40544} = 201.3$$

Coefficient of Variation

$$= \frac{\sigma}{\overline{X}} \times 100 = \frac{201.3}{866} \times 100 = 23.24$$

Thus, CV = 23.24%

3) **Calculation of Coefficient of Variation – Continuous Series (Grouped Data)**

Example 18: Following data represents daily wages paid to workers in two factories A and B:

Daily Wages	No. of Workers	
	Factory A	Factory B
50-100	7	6
100-150	31	21
150-200	23	29
200-250	14	15
250-300	9	13
300-350	11	5
350-400	5	1

i) Find the average wage in both factories.
ii) Which factory has a more consistent wage structure?

Solution: For Factory A ⇒

Calculation of Coefficient of Variation for Factory A

Daily Wages	Mid Value (X)	f	fX	d_x	$f(d_x)^2$
50-100	75	7	525	–120	100800
100-150	125	31	3875	–70	151900
150-200	175	23	4025	–20	9200
200-250	225	14	3150	30	12600
250-300	275	9	2475	80	57600
300-350	325	11	3575	130	185900
350-400	375	5	1875	180	162000
Total		**Σf = 100**	**ΣfX = 19500**		$\sum f(d_x)^2 = 680000$

$$\text{Mean } (\overline{X}) = \frac{\Sigma fX}{\Sigma f} = \frac{19500}{100} = 195$$

$$\sigma_x = \sqrt{\frac{\Sigma f(d_x)^2}{\Sigma f}} = \sqrt{\frac{680000}{100}} = \sqrt{6800} = 82.5$$

$$C.V. = \frac{\sigma_x}{\overline{X}} \times 100 = \frac{82.5}{195} \times 100 = 42.31$$

So, $Q_3 = L_1 + \frac{\frac{3N}{4} - c}{f} \times i = 60 + \frac{42 - 41}{9} \times 10$

$= 60 + 1.1 = 61.1$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{61.1 - 32}{2} = \frac{29.1}{2} = 14.55$$

3.5. MEAN DEVIATION/ MEAN ABSOLUTE DEVIATION (M.D.)

3.5.1. Introduction

The arithmetic average of the absolute deviation of a series is known as the mean deviation. This deviation is calculated from any one of the measures of averages such as mean, mode and median.

Mathematically,

$$\text{Mean Deviation (MD)} = \frac{\sum |d|}{N}$$

Where, $\sum |d|$ = Sum of all deviations (taken either from mean or median, ignoring ± signs.)

In a series, it can be defined as the arithmetic average of the deviation of various items from mean or median of that series.

For calculating the mean deviation, median is preferred because the addition of the deviation from the mean is greater than that from the median. So, the value of mean deviation measured from the mean is greater than the value measured from the median.

Mean deviation is also known as **First Moment of Dispersion**.

Note: It is generally measured either from median or mean.

3.5.2. Advantages of Mean Deviation

1) It is easy to understand and defined rigidly.
2) It is a better measure of dispersion in comparison to range and quartile deviation because it includes all the observations in the calculation.
3) It gives a good measure for the comparison of the formation of different distributions because it is based on the deviation about average.
4) Compared to standard deviation, it is less affected by the extreme values.

3.5.3. Disadvantages of Mean Deviation

1) For open ended classes, it is not possible to calculate mean deviation.
2) Mean deviation tends to increase with the size of the sample, though not proportionately and not as rapidly as range.
3) In sociological studies, it is rarely used.
4) The ignorance of sign is mathematically unsound and illogical.

3.5.4. Coefficient of Mean Deviation

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

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

Coefficient of mean deviation from mode = $\frac{\text{Mean Deviation (M.D.)}}{\text{Mode}}$

3.5.5. Method of Calculation of Mean deviation

1) Computation of Mean Deviation - Individual Series
2) Computation of Mean Deviation - Discrete Series
3) Computation of Mean Deviation - Continuous Series

3.5.5.1. Computation of Mean Deviation - Individual Series

Following steps are involved in the calculation of the mean deviation:

1) First, measure the average mean, median or mode of the series.
2) Find the deviation of the items from the average, while ignoring the positive (+) and negative (-) signs. The resultant deviation is denoted by $|dx|$.
3) Next, measure the total sum of these deviations. It is represented by $\sum |dx|$.
4) In the last step, divide the sum obtained by the number of items.

Symbolically,

$$\text{Mean Deviation} = \frac{\sum |d_x|}{N}$$

Where,

$|d_x|$ = deviation from mean (or median) ignoring ± signs.

N = Number of items

Example 8: The following are the monthly expenditure of six families. Calculate mean deviation from mean and mean deviation from median.

Expenditure (₹)	4,260	4,980	8,460	5,240	4,780	6,480

Solution: First we arrange the given data in ascending order:

Expenditure (₹)	4,260	4,780	4,980	5,240	6,480	8,460

Here, N = 6 which is even, so

$$\text{Median} = \frac{\left(\frac{N}{2}\right)^{th}\text{value} + \left(\frac{N}{2}+1\right)^{th}\text{value}}{2}$$

$$= \frac{\left(\frac{6}{2}\right)^{th}\text{value} + \left(\frac{6}{2}+1\right)^{th}\text{value}}{2}$$

$$= \frac{3^{rd}\text{ value} + 3.5^{th}\text{ value}}{2} = \frac{4980+5240}{2} = \frac{10220}{2} = 5110$$

$$A.M. = \frac{\sum X}{N} = \frac{1}{6} \times 34200 = 5700.$$

About Mean

Serial No.	X (₹)	Deviation from A.M. ignoring ± sign $\lvert d_x \rvert$
1	4260	1440
2	4780	920
3	4980	720
4	5240	460
5	6480	780
6	8460	2760
Total	ΣX = 34200	$\Sigma\lvert d_x \rvert$ = 7080

About Median

Serial No.	X (₹)	Deviation from Median Ignoring ± sign $\lvert d_x \rvert$
1	4260	850
2	4780	330
3	4980	130
4	5240	130
5	6480	1370
6	8460	3350
Total	ΣX = 34200	$\Sigma\lvert d_x \rvert$ = 6160

$$\text{Mean deviation (about mean)} = \frac{\sum \lvert d_x \rvert}{N} = \frac{7080}{6} = 1180$$

$$\text{Mean deviation (about median)} = \frac{\sum \lvert d_x \rvert}{N} = \frac{6160}{6} = 1026.6$$

3.5.5.2. Computation of Mean Deviation - Discrete Series

In case of discrete series following steps are involved:

1) First, calculate the averages such as mean, mode or median.
2) From the central tendency calculate the deviation of the size, ignore the positive and negative signs. The result is represented by $\lvert d_x \rvert$.
3) In this step, deviation of every size ($\lvert d_x \rvert$) is multiplied by their respective frequency (f) and summated ($\Sigma f \lvert d_x \rvert$).
4) In the last step, total sum is divided by the total frequency.

Symbolically

$$M.D. = \frac{\sum f \lvert d_x \rvert}{\sum f}$$

Where,

$\lvert d_x \rvert$ = deviation from mean (or median) ignoring ± signs.

N = total frequency

Example 9: Calculate the Mean-deviation from the following data:

Quantity Demanded (Units)	10	20	30	40	50	60	70	80	90	100
frequency	7	13	16	6	14	19	28	17	21	9

Solution:

Quantity Demanded (X)	frequency (f)	fX	$\lvert d_x \rvert = (X - \overline{X})$ ignoring ± sign	$f\lvert d_x \rvert$
10	7	70	50	350
20	13	260	40	520
30	16	480	30	480
40	6	240	20	120
50	14	700	10	140
60	19	1140	0	0
70	28	1960	10	280
80	17	1360	20	340
90	21	1890	30	630
100	9	900	40	360
Total	Σf = 150	ΣfX = 9000		3220

$$\text{Arithmetic Mean } (\overline{X}) = \frac{\sum fX}{\sum f} = \frac{9000}{150} = 60$$

$$\text{Mean Deviation} = \frac{\Sigma f \lvert d_x \rvert}{\Sigma f} = \frac{3220}{150} = 21.5$$

3.5.5.3. Computation of Mean Deviation - Continuous Series

In case of continuous series, following steps are involved:

1) First, calculate the mid value of the class intervals.
2) Measure the median or arithmetic mean.
3) Find $\lvert d_x \rvert$ and $f\lvert d_x \rvert$.
4) Divide the total sum $\sum f \lvert d_x \rvert$ by the total frequency $\sum f$.

Example 10: Calculate Mean deviation from the mean for the following data:

Sales (₹ in thousand)	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of days	5	6	8	15	17	7	9	3

Solution:

Calculation of Mean Deviation

Sales	Mid Value(X)	No. of days (f)	fX	$\lvert d_X \rvert = \lvert X - \overline{X} \rvert$	$f\lvert d_X \rvert$
0 – 10	5	5	25	35	175
10 – 20	15	6	90	25	150
20 – 30	25	8	200	15	120
30 – 40	35	15	525	5	75
40 – 50	45	17	765	5	85
50 – 60	55	7	385	15	105
60 – 70	65	9	585	25	225
70 – 80	75	3	225	35	105
		Σf = 70	ΣfX = 2800		$\Sigma f\lvert d_x \rvert$ = 1040

MEAN DEVIATION

Definition

Mean Deviation or Average Deviation is arithmetic average of deviations of all the values taken from a statistical average (Mean, Median or Mode) of the series. In taking deviations of values, algebraic sign + and – are not taken into consideration (± signs are ignored) that is, negative deviations are also treated as positive deviations.

Mean deviation is also known as *First Absolute Moment*. Mean deviation may be based on mean, median or mode. That is deviations are taken either from mean or median or mode.

Remark : If the point about which we have to calculate mean deviation is not mentioned, then use median. Mean deviation d_a is least when a = median.

Formula (For Direct Method)

(a) *Mean Deviation from Mean* :

(i) Individual Series : $$\delta_{\bar{X}} = \frac{\Sigma|X-\bar{X}|}{N} = \frac{\Sigma d_{\bar{X}}}{N}$$

(ii) Discrete Series : $$\delta_{\bar{X}} = \frac{\Sigma f|X-\bar{X}|}{\Sigma f} = \frac{\Sigma f d_{\bar{X}}}{N}$$

(iii) Grouped Series : $$\delta_{\bar{X}} = \frac{\Sigma f|X-\bar{X}|}{\Sigma f} = \frac{\Sigma f d_{\bar{X}}}{N}$$

Here X = Mid-value.

(b) *Mean Deviation from Median* :

(i) Individual Series : $$\delta_M = \frac{\Sigma|X-M|}{N} = \frac{\Sigma d_M}{N}$$

(ii) Discrete Series : $$\delta_M = \frac{\Sigma f|X-M|}{\Sigma f} = \frac{\Sigma f d_M}{N}$$

(iii) Grouped Series : $$\delta_M = \frac{\Sigma f|X-M|}{\Sigma f} = \frac{\Sigma f d_M}{N}$$

Here X = Mid-value.

(c) *Mean Deviation from Mode* :

(i) Individual Series : $$\delta_Z = \frac{\Sigma|X-Z|}{N} = \frac{\Sigma d_Z}{N}$$

(ii) Discrete Series : $$\delta_Z = \frac{\Sigma f|X-Z|}{\Sigma f} = \frac{\Sigma f d_Z}{N}$$

(iii) Grouped Series : $$\delta_Z = \frac{\Sigma f|X-Z|}{\Sigma f} = \frac{\Sigma f d_Z}{N}$$

Here X = Mid-value.

COEFFICIENT OF MEAN DEVIATION

Let d_a denote mean deviation about any point a (may be mean, median, mode.) Then coefficient of mean deviation is defined as $\frac{\delta_a}{a}$.

Thus, Coefficient of mean deviation from mean $= \frac{\delta_{\bar{X}}}{\bar{X}}$

Coefficient of mean deviation from median $= \frac{\delta_M}{M}$

Coefficient of mean deviation from mode $= \frac{\delta_Z}{Z}$

COMPUTATION OF MEAN DEVIATION AND ITS COEFFICIENT IN INDIVIDUAL SERIES

Direct Method

Steps

(i) Find mean (or median or mode as the case may be).

(ii) Find deviations about mean, median or mode, taking positive sign only. That is find $d\bar{x} = |X - \bar{X}|$.

(iii) Add all these absolute deviations. That is find $\Sigma d\bar{x}$.

(iv) Divide this total by number of observations to find the required mean deviation. That is find

$$\delta\bar{x} = \frac{\Sigma|d\bar{x}|}{N}$$

(v) Divide this mean deviation by mean to obtain the coefficient of mean deviation. That is find $\dfrac{\delta\bar{x}}{\bar{X}}$

Illustration 1

Given below are the weights of 9 students of a class. Calculate Mean Deviation and their coefficients from Mean :

Serial Number :	1	2	3	4	5	6	7	8	9
Weight (in kg) :	50	52	55	57	60	61	64	65	67

Solution

S. No.	*Weight in kgs. X*	*Deviation from Mean* $\|d\bar{x}\| = \|X - \bar{X}\|$
1	50	9
2	52	7
3	55	4
4	57	2
5	60	1
6	61	2
7	64	5
8	65	6
9	67	8
$N = 9$	$\Sigma X = 531$	$\Sigma\|d\bar{x}\| = 44$

$$\text{Mean, } \bar{X} = \frac{\Sigma X}{N} = \frac{531}{9} = 59 \text{ kg}$$

M. D. from Mean, $$\delta\bar{x} = \frac{\Sigma|d\bar{x}|}{N} = \frac{44}{9} = 4.89 \text{ kg}$$

$$\text{Coeff. of } M.D. = \frac{\delta\bar{x}}{\bar{X}} = \frac{4.89}{59} = 0.0829$$

Illustration 2

For the following data, calculate mean deviation from mode :

14, 8, 20, 18, 30, 24, 14, 18, 14

Solution

x : 14, 8, 20, 18, 30, 24, 14, 18, 14

Calculation Table of Mean Deviation

x	$dz = x - Z = x - 14$	$\|dz\| = \|x - Z\|$	
14	0	0	Mean deviation from Mean
8	−6	6	$\delta_Z = \frac{\Sigma \| dz \|}{N}$
20	6	6	
18	4	4	$= \frac{46}{9}$
30	16	16	
24	10	10	$= 5.11$
14	0	0	
18	4	4	
14	0	0	
		$\Sigma \| dz \| = 46$	

Short-cut Method (Mean Deviation from Mean)

(i) Calculate mean for N values (given).
(ii) Find ΣM_A = Total of values above the mean.
(iii) Find ΣM_B = Total of values below the mean.
(iv) Find N_A = Number of items above mean.
(v) Find N_B = Number of items below mean.
(vi) N = Total number of items
(vii) Use the formula :

$$\delta_{\bar{X}} = \frac{\Sigma M_A - \Sigma M_B - (N_A - N_B)\bar{X}}{N}$$

Illustration 3

Find the coefficient of mean deviation from mean for the following set of observations using short-cut method :

97, 85, 98, 79, 86, 89

Solution

Short-cut Method

$$\Sigma X = 97 + 85 + 98 + 79 + 86 + 89 = 534$$

$$\text{Mean} = \frac{\Sigma X}{N}$$

$$= \frac{534}{6} = 89$$

Arrange the values in ascending order :

79, 85, 86, 89, 97, 98

Total of values above the mean = 97 + 98 = 195

ΣM_A = Total of values below the mean = 79 + 85 + 86 = 250

ΣM_B = Total of values above the mean = 2

N_A = Number of observations above the mean = 3

N_B = Number of observations below the mean = 3

Formula :

$$\delta_{\bar{X}} = \frac{\Sigma M_A - \Sigma M_B - (N_A - N_B)\bar{X}}{N}$$

$$= \frac{195 - 250 - (2 - 3) \times 89}{6}$$

$$= \frac{-55 - (-1) \times 89}{6} = \frac{-55 + 89}{6} = \frac{34}{6} = 5.67$$

Coefficient of Mean Deviation

$$= \frac{\delta_{\bar{X}}}{\bar{X}} = \frac{5.67}{89} = 0.06$$

Alternative (or Second) Short-cut Method

$$\delta_{\bar{X}} = \frac{\Sigma|d_x| + \text{Adjustment of total error}}{N}$$

where d_x = deviation from assumed mean

$\Sigma|d_x|$ = Sum of the absolute deviations from assumed mean

Adjustment of total error = $(\bar{X} - A)(N_B - N_A)$

N_B = Number of items below the mean

N_A = Number of items above the mean

Let assumed mean, $A = 90$

X	$d_x = X - 90$	$\|d_x\|$
79	−11	11
85	−5	5
86	−4	4
89	−1	1
97	7	7
98	8	8
Total	—	36

$\Sigma|d_x| = 36$
$\bar{X} = 89$
$A = 90$
$N_B = 3$
$N_A = 2$
$N = 6$

Mean deviation from mean, $\delta_{\bar{X}} = \frac{36 + (89 - 90)(3 - 2)}{6}$

$$= \frac{36 + (-1 \times 1)}{6} = \frac{35}{6} = 5.83$$

Short-cut Method (Mean Deviation from Median)

The following procedure is applied to calculate Mean Deviation through Median :

1. Firstly, the average median is calculated on the basis of which mean deviation is to be computed.
2. In case of discrete series value (x) and in case of continuous series mid-values are multiply by the respective frequencies and then (fx) is obtained.
3. The total of products of values or mid-values greater than the average multiplied by their respective frequencies is called ΣfX_A and the total of products of values or mid-values less than the average multiplied by their respective frequencies is known as ΣfX_B. If there is any value or mid-value equal to the average, it is left out.
4. Then Σf_A is obtained by adding the frequencies relating to values greater than the average median and Σf_B by adding the frequencies relating to values less than the average median.
5. Lastly, the following formula is applied :

$$M.D.\text{ (Median)} = \frac{\Sigma fX_A - \Sigma fX_B - (\Sigma f_A - \Sigma f_B)M}{N}$$

Illustration 4

Calculate mean deviation about median for the following data using direct method and short-cut method :

75, 70, 68, 57, 66, 55, 50, 60, 62, 79, 63

Solution

Arranging the data in ascending order, we have

50, 55, 57, 60, 62, 63, 65, 68, 70, 75, 79

Median, $M = \frac{N+1}{2}$th value $= \frac{11+1}{2}$th value = 6th value = 63.

Short-cut Method

$$\delta_M = \frac{\Sigma M_A - \Sigma M_B - (N_A - N_B)\,M}{N}$$

Here N = Total number of observations = 11

M = Median = 63

N_A = Number of observations above median = 5

N_B = Number of observations below median = 5

ΣM_A = Sum of the values above median = 65 + 68 + 70 + 75 + 79 = 357

ΣM_B = Sum of the values below median = 50 + 55 + 57 + 60 + 62 = 284

$$\therefore \quad \delta_M = \frac{(357-284)-(5-5)\times 63}{11}$$

$$= \frac{73-0}{11} = \frac{73}{11} = 6.636 \simeq 6.64$$

Illustration 5

Calculate Mean deviations from (i) arithmetic mean, (ii) mode, (iii) median in respect of the marks obtained by nine students given below and show that mean deviation from median is minimum.

Marks (out of 25) : 7, 4, 10, 9, 15, 12, 7, 9, 7.

If the marks are doubled (converted out of 50) will the variation of marks increase? Give reasons.

Solution

$$\text{Mean, } \overline{X} = \frac{7+4+10+9+15+12+7+9+7}{9} = \frac{80}{9} = 8.9 \text{ marks}$$

To calculate median arrange the data in ascending order :

S. No. :	1	2	3	4	5	6	7	8	9
Marks :	4	7	7	7	9	9	10	12	15

Median, $M = \left(\frac{N+1}{2}\right)$th item $= \frac{9+1}{2}$th item = 5th item = 9 marks

Mode, Z = The item with maximum frequency = 7 marks

Computation of Mean Deviation

S.No. (m)	Devs. from Means $\lvert d\overline{x}\rvert = \lvert m - 8.9\rvert$	Devs. from Median $\lvert d_M\rvert = \lvert m - 9\rvert$	Devs. from Mode $\lvert dz\rvert = \lvert m - 7\rvert$
4	4.9	5	3
7	1.9	2	0
7	1.9	2	0
7	1.9	2	0
9	0.1	0	2
9	0.1	0	2
10	1.1	1	3
12	3.1	3	5
15	6.1	6	8
	$\Sigma\lvert d\overline{x}\rvert = 21.1$	$\Sigma\lvert d_M\rvert = 21$	$\Sigma\lvert dz\rvert = 23$

Mean Deviation from Mean :

$$\delta \bar{x} = \frac{\Sigma|d\bar{x}|}{N} = \frac{21.1}{9} = 2.34 \text{ marks}$$

Mean Deviation from Median :

$$\delta_M = \frac{\Sigma|d_M|}{N} = \frac{21}{9} = 2.33 \text{ marks}$$

Mean Deviation from Mode :

$$\delta_Z = \frac{\Sigma|d_Z|}{N} = \frac{23}{9} = 2.56 \text{ marks}$$

Obviously mean deviation is least about median. When the marks are doubled, then mean, median and mode will also be twice. Hence $\Sigma|d\bar{x}|$, $\Sigma|d_M|$ and $\Sigma|d_Z|$ will remain same. This implies that mean deviations will remain same. This means the variation between the marks will not increase. However, coefficients of mean deviation will decrease.

COMPUTATION OF MEAN DEVIATION AND ITS COEFFICIENT IN DISCRETE SERIES

Direct Method

Steps

1. Calculate the mean (or median or mode as the case may be).
2. Find the absolute deviations.
3. Multiply these deviations with corresponding frequencies.
4. Add these products.
5. Divide this sum by total frequency.
6. Divide the mean deviation from the point a, say by a to obtain the coefficient of mean deviation.

Illustration 6

Find out the mean deviation from Mean, Median and Mode and their coefficient from the following series :

Size of items :	4	6	8	10	12	14	16
Frequency :	2	1	3	6	4	3	1

Solution

Size (m)	*Frequency (f)*	*mf*	*c.f.*	*From Mean* $\vert d\bar{X}\vert$	*From Mean* $f\vert d\bar{X}\vert$	*From Median* $\vert d_M\vert$	*From Median* $f\vert d_M\vert$
4	2	8	2	6.2	12.4	6	12
6	1	6	3	4.2	4.2	4	4
8	3	24	6	2.2	6.6	2	6
10	6	60	12	0.2	1.2	0	0
12	4	48	16	1.8	7.2	2	8
14	3	42	19	3.8	11.4	4	12
16	1	16	20	5.8	5.8	6	6
Total	20	204			48.8		48

$$\text{Mean, } \bar{X} = \frac{\Sigma mf}{\Sigma f} = \frac{204}{20} = 10.2$$

$$\text{Median, } M = \frac{N+1}{2} \text{th value}$$

$$= \frac{20+1}{2} \text{th value} = 10.5\text{th value} = 10$$

Mode, Z = The value with maximum frequency = 10

(By inspection)

Mean Deviation from Mean :

$$\delta_{\bar{x}} = \frac{\Sigma f|d\bar{x}|}{N} = \frac{48.8}{20} = 2.44$$

$$\text{Coeff. of } M.D. = \frac{\delta_{\bar{x}}}{\bar{X}} = \frac{2.44}{10.2} = 0.2392$$

Mean Deviation from Median :

$$\delta_M = \frac{\Sigma f|d_M|}{N} = \frac{48}{20} = 2.4$$

$$\text{Coeff. of } M.D. = \frac{\delta_M}{M} = \frac{2.4}{10} = 0.24$$

Mean Deviation from Mode :

Since Median and Mode are equal, the mean deviation from median and mean deviation from mode will be same.

$$\delta_Z = d_M = 2.4$$

and Coeff. of $M.D. = 0.24$

Short-cut Method

Formula :

$$\delta_M = \frac{\Sigma mf_A - \Sigma mf_B - (\Sigma f_A - \Sigma f_B)M}{N}$$

$$\delta_{\bar{X}} = \frac{\Sigma mf_A - \Sigma mf_B - (\Sigma f_A - \Sigma f_B)\bar{X}}{N}$$

$$\delta_Z = \frac{\Sigma mf_A - \Sigma mf_B - (\Sigma f_A - \Sigma f_B)Z}{N}$$

where, Σmf_A = represents the total size $(m \times f_A)$ above the related average

Σmf_B = represents the total size $(m \times f_B)$ below the related average

Σf_A = represents the total frequencies corresponding to values above the related average

Σf_B = represents the total frequencies corresponding to values below the related average

Illustration 7

Calculate mean deviation about median and its coefficient from the following data using (i) direct method, and (ii) short-cut method :

Height (cm)	:	125	126	127	128	129	130	135
No. of Children	:	5	8	15	12	10	6	4

Solution

X	f	$c.f.$	$\|X-M\|$	$f\|X-M\|$
125	5	5	3	15
126	8	13	2	16
127	15	28	1	15
128	12	40	0	0
129	10	50	1	10
130	6	56	2	12
135	4	60	7	28
Total	60	—	—	96

$$\text{Median, } M = \frac{N}{2}\text{th item}$$

$$= \frac{60}{2}\text{th item} = 30\text{th item} = 128 \text{ cm}$$

Direct Method

$$d_M = \frac{\Sigma f|X-M|}{\Sigma f} = \frac{96}{60} = 1.6 \text{ cm}$$

Short-cut Method

$\Sigma mf_A = 129 \times 10 + 130 \times 6 + 135 \times 4 = 1,290 + 780 + 540 = 2,610$

$\Sigma mf_B = 125 \times 5 + 126 \times 8 + 127 \times 15 = 625 + 1,008 + 1,905 = 3,538$

$\Sigma f_A = 10 + 6 + 4 = 20$

$\Sigma f_B = 5 + 8 + 15 = 28$

$$\delta_M = \frac{(\Sigma mf_A - \Sigma mf_B) - (\Sigma f_A - \Sigma f_B) \times M}{N}$$

$$= \frac{(2,610 - 3,538) - (20 - 28) \times 128}{60}$$

$$= \frac{-928 - (-8) \times 128}{60}$$

$$= \frac{-928 + 1,024}{60} = \frac{96}{60} = 1.6 \text{ cm}$$

$$\text{Coefficient of } M.D. = \frac{\delta_M}{M} = \frac{1.6}{128} = 0.0125$$

COMPUTATION OF MEAN DEVIATION AND ITS COEFFICIENT IN CASE OF GROUPED SERIES

Steps

Find the mid-values. Then the grouped series becomes discrete with mid-values. Proceed as in case of discrete series.

***Illustration* 8**

Find mean deviation from mode and its coefficient for the following frequency distribution :

Class	:	0—10	10—20	20—30	30—40	40—50
Frequency	:	5	8	15	16	6

Solution

Maximum frequency = 16, Modal class is (30—40).

$$\text{Mode, } Z = L_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times (L_2 - L_1)$$

$$\Rightarrow \quad Z = 30 + \frac{16 - 15}{2 \times 16 - 15 - 6}(40 - 30)$$

$$= 30 + \frac{10}{11} = 30 + .9 = 30.9$$

Computation Table for Mean Deviation

Mid-value x	f	*Deviation from Mode* $\|dz\| = \|x - Z\|$	$f\|dz\|$
5	5	25.9	129.5
15	8	15.9	127.2
25	15	5.9	88.5
35	16	4.1	65.6
45	6	14.1	84.6
	$N = 50$		495.4

$$\text{Mean Deviation from Mode, } \delta_Z = \frac{\Sigma f|dz|}{N} = \frac{495.4}{50} = 9.908$$

$$\text{Coeff. of } M.D. \text{ from Mode} = \frac{\delta_Z}{Z} = \frac{9.908}{30.9} = 0.3206 \text{ or } 32.06\%$$

***Illustration* 9**

Calculate mean deviation from mean using (i) direct method and (ii) short-cut method of the following data :

Size	:	3—4	4—5	5—6	6—7	7—8	8—9	9—10
Frequency	:	3	7	22	60	85	32	8

Solution

Class	Mid-value (x)	d_x (x − 6.5)	f	fd_x	$\lvert x-\bar{X}\rvert = \lvert d\bar{x}\rvert$	$f\lvert d\bar{x}\rvert$
3–4	3.5	−3	3	−9	3.6	10.8
4–5	4.5	−2	7	−14	2.6	18.2
5–6	5.5	−1	22	−22	1.6	35.2
6–7	6.5	0	60	0	0.6	36.0
7–8	7.5	1	85	85	0.4	34.0
8–9	8.5	2	32	64	1.4	44.8
9–10	9.5	3	8	24	2.4	19.2
Total			217	128		198.2

$$\bar{X} = a + \frac{\Sigma fd_x}{N} = 6.5 + \frac{128}{217} = 6.5 + 0.59 = 7.09 \simeq 7.1$$

Direct Method $\delta_{\bar{X}} = \frac{\Sigma f\lvert d\bar{x}\rvert}{N} = \frac{198.2}{217} = 0.913$

Short-cut Method

Formula : $\delta_{\bar{X}} = \frac{\Sigma f\lvert d_x\rvert + (\bar{X} - A)(\Sigma f_B - \Sigma f_A)}{N}$

where d_x = deviation from any point A

$\Sigma f\lvert d_x\rvert$ = Sum of the products of d_x and f

Σf_B = Number of items below the mean

Σf_A = Number of items above the mean

From the above table for $A = 6.5$, we have

$$\Sigma f\lvert d_x\rvert = 3\times\lvert-3\rvert + 7\times\lvert-2\rvert + 22\times\lvert-1\rvert + 60\times\lvert 0\rvert + 85\times\lvert 1\rvert + 32\times\lvert 2\rvert + 8\times\lvert 3\rvert$$

$$= 9 + 14 + 22 + 0 + 85 + 64 + 24 = 218$$

$$\Sigma f_A = 85 + 32 + 8 = 125$$

$$\Sigma f_B = 3 + 7 + 22 + 60 = 92$$

$$\therefore \quad \delta_{\bar{X}} = \frac{218 + (7.1 - 6.5)(92 - 125)}{217}$$

$$= \frac{218 + 0.6\times(-33)}{217} = \frac{218 - 19.8}{217} = \frac{198.2}{217} = 0.913$$

Illustration 10

Calculate mean deviation from median by short-cut method from the following data :

Marks	:	10—20	20—30	30—40	40—50	50—60	60—70	70—80
Frequency	:	15	14	16	20	15	11	9

Solution

Marks	Mid-value m	Frequency f		c.f.	m.f.	
10–20	15	15	$\Sigma f_A = 45$	15	225	$\Sigma mf_B = 1{,}135$
20–30	25	14		29	350	
30–40	35	16		45	560	
40–50	45	20	$\Sigma f_B = 55$	65	900	$\Sigma mf_A = 3{,}115$
50–60	55	15		80	825	
60–70	65	11		91	715	
70–80	75	9		100	675	
Total	—	100		—		

$$m = \frac{N}{2}\text{th item} = \frac{100}{2}\text{th item} = 50\text{th item}$$

$$\text{Median, } M = L_1 + \frac{m - c}{f} \times i$$

$$= 40 + \frac{50 - 45}{20} \times 10 = 40 + \frac{5 \times 10}{20} = 40 + 2.5 = 42.5 \text{ Marks}$$

Formula : $$\delta_M = \frac{\Sigma m f_A - \Sigma m f_B - (\Sigma f_A - \Sigma f_B) \times M}{N}$$

$$= \frac{(3{,}115 - 1{,}135) - (55 - 45) \times 42.5}{100}$$

$$= \frac{1{,}980 - 10 \times 42.5}{100}$$

$$= \frac{1{,}980 - 425}{100} = \frac{1{,}555}{100} = 15.55 \text{ Marks}$$

Illustration 11

Calculate (i) mean deviation from median and its coefficient (ii) quartile deviation and its coefficient for the following data :

Class :	0—6	6—12	12—18	18—24	24—30
Frequency :	8	10	12	9	6

Solution

(i) *Mean Deviation from Median and its Coefficient* :

Class	*Frequency (f)*	*Cumulative Frequency*
0—6	8	8
6—12	10	18
12—18	12	30
18—24	9	39
24—30	6	45
Total	45	

$$\text{Median Number, } m = \frac{N}{2} = \frac{45}{2} = 22.5$$

Median, $M = m$th item $= 22.5$th item

This lies in the class (12 — 18).

$$\text{Median, } M = L_1 + \frac{m - c}{f} \times i$$

$$= 12 + \frac{22.5 - 18}{12} \times 6$$

$$= 12 + \frac{4.5 \times 6}{12} = 12 + 2.25 = 14.25$$

Computation Table for Mean Deviation

Class	*Mid-value (x)*	*Frequency (f)*	$\lvert x - M \rvert = \lvert d_M \rvert$	$f\lvert d_M \rvert$
0—6	3	8	11.25	90.00
6—12	9	10	5.25	52.50
12—18	15	12	0.75	9.00
18—24	21	9	6.75	60.75
24—30	27	6	12.75	76.50
Total		45		288.75

Mean Deviation, $\delta_M = \frac{\Sigma f|d_M|}{N} = \frac{288.75}{45} = 6.42$

Coefficient of mean deviation from median

$$= \frac{\delta_M}{M} = \frac{6.42}{14.25} = 0.45$$

(iii) *Computation for quartile deviation and its coefficient :*

$$q_1 = \frac{N}{4} = \frac{45}{4} = 11.25$$

$$q_3 = \frac{3N}{4} = \frac{3 \times 45}{4} = 33.75$$

First quartile, $Q_1 = L_1 + \frac{q_1 - c}{f} \times i$

$$= 6 + \frac{11.25 - 8}{10} \times 6$$

$$= 6 + \frac{3.25 \times 6}{10} = 6 + 1.95 = 7.95$$

Third quartile, $Q_3 = L_1 + \frac{q_3 - c}{f} \times i$

$$= 18 + \frac{33.75 - 30}{9} \times 6$$

$$= 18 + \frac{3.75 \times 6}{9} = 18 + 2.5 = 20.5$$

$\therefore$ Quartile Deviation, $Q.D. = \frac{Q_3 - Q_1}{2} = \frac{20.5 - 7.95}{2}$

$$= \frac{12.55}{2} = 6.275$$

Coefficient of quartile deviation $= \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{20.5 - 7.95}{20.5 + 7.95} = \frac{12.55}{28.45} = 0.44$

Illustration 12

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

Mid-value	:	35	40	45	50	55
Frequency	:	2	5	8	6	4

Solution

Make the sqares with the help of mid-values.

C.I.	*Mid-Value (x)*	(f)	fx	c.f.
32.5—37.5	35	2	70	2
37.5—42.5	40	5	200	7
42.5—47.5	45	8	360	15
47.5—52.5	50	6	300	21
52.5—57.5	55	4	220	25
Total		25	1,150	—

$$\text{Mean}, \overline{X} = \frac{\Sigma fx}{\Sigma f} = \frac{1,150}{25} = 46$$

Median, $M = L_1 + \frac{m - c}{f} \times i$, $m = \frac{N}{2} = \frac{25}{2} = 12.5$th item

it is in 42.5—47.5 class.

$$= 42.5 + \frac{12.5 - 7}{8} \times 5$$

$\therefore$

$$M = 42.5 + \frac{5.5 \times 5}{8} = 42.5 + 3.4375 = 45.9375 \simeq 46$$

$\because$ Mean = Median = 46, $\therefore \delta_{\overline{X}} = \delta_M$

x	f	|x − 46|	f |x − 46|
35	2	11	22
40	5	6	30
45	8	1	8
50	6	4	24
55	4	9	36
Total	25	—	120

$$\delta_M = \delta_{\bar{X}} = \frac{\Sigma f|x-46|}{\Sigma f} = \frac{120}{25} = 4.8$$

Illustration 13

Calculate mean deviation from median and its coefficient of the data given below:

Age (in years under)	*Number of Persons Died*
10	15
20	30
30	53
40	75
50	100
60	110
70	115
80	125

Solution

Convert the cumulative frequency distribution into ordinary frequency distribution :

Marks	*Mid-value (m)*	*Frequency (f)*		*c.f.*	*mf*	
0—10	5	15	$\Sigma f_B = 53$	15	75	$\Sigma mf_B = 875$
10—20	15	15		30	225	
20—30	25	23		53	575	
30—40	35	22	$\Sigma f_A = 72$	75	770	$\Sigma mf_A = 3{,}520$
40—50	45	25		100	1,125	
50—60	55	10		110	550	
60—70	65	5		115	325	
70—80	75	10		125	750	
Total	—	125		—		4,395

$$\text{Median, } M = L_1 + \frac{\frac{N}{2} - c}{f} \times i$$

$$= 30 + \frac{62.5 - 53}{22} \times 10$$

$$= 30 + \frac{9.5 \times 10}{22} = 30 + 4.318 = 34.318 \simeq 34.32 \text{ years}$$

Using Short-cut Method :

Formula :
$$\delta_M = \frac{(\Sigma mf_A - \Sigma mf_B) - (\Sigma f_A - \Sigma f_B) \times M}{N}$$

$$= \frac{(3{,}520 - 875) - (72 - 53) \times 34.32}{125}$$

$$= \frac{2{,}645 - (19) \times 34.32}{125}$$

$$= \frac{2{,}645 - 652.08}{125} = \frac{1{,}992.92}{125} = 15.94 \text{ years}$$

$$\text{Coefficient of } M.D. = \frac{\delta_M}{M} = \frac{15.94}{34.32} = 0.465 \text{ or } 46.5\%$$

UNIT- 1

Name: Jyoti Yadav
Designation: Assistant Professor
Department: Management And Commerce Department

Subject: Business Statistics

B. Com First Semester

Business Statistics

Meaning of Statistics (Indian Context)

Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of data. In the Indian context, statistics plays a vital role in economic planning, policymaking, agriculture, industry, and social development.

Father of Statistics in India: **Prof. Prasanta Chandra Mahalanobis (P.C. Mahalanobis)** is called the **Father of Indian Statistics**.

About P.C. Mahalanobis

Feature	Details
Full Name	Prasanta Chandra Mahalanobis
Born	29 June 1893
Famous For	Mahalanobis Distance, Large-Scale Sample Surveys
Founder of	Indian Statistical Institute (ISI), Kolkata
Key Contribution	Designed India's Second Five-Year Plan
Position	Member of the Planning Commission of India
Recognition	National Statistics Day is celebrated on **29th June** in his honour

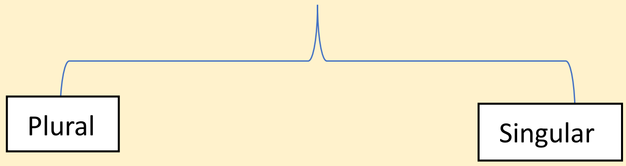
Introduction to Statistics:

The word **statistics** is derived from the Latin word *status* or the Italian *Statista*, meaning "statesman". Professor **Gottfried Achenwall** popularized the term in the 18th century to describe data related to the political and economic state of a region. Initially used for state records like census and wealth, its scope later expanded to a wide range of data-based analysis.

Meaning of Statistics

In Simple Words Statistics is the science of collecting, organizing, analysing, and interpreting numerical data to support decision-making or draw conclusions.

In broader sense the word **"Statistics"** has two meanings:



1. In Plural Sense

Meaning: *"Statistics" as data or facts expressed in numbers.* It refers to **numerical data** collected for analysis. Example: Population figures, income levels, exam scores, etc. or the literacy rate of India is 77.7% — this is a statistic.

2. In Singular Sense

Meaning: *"Statistics" as a discipline or method of study* It refers to the **science or method** of collecting, organizing, analysing, interpreting, and presenting data*. Example:* The process of analysing market trends using data is statistics as a subject.

Definitions of Statistics:

- **Prof. Horace Secrist**
  Statistics is the aggregate of facts, affected to a marked extent by a multiplicity of causes, numerically expressed, enumerated or estimated according to reasonable standards of accuracy, collected in a systematic manner for a predetermined purpose and placed in relation to each other.
- **Croxton and Cowden**
  Statistics may be defined as the collection, presentation, analysis, and interpretation of numerical data.
- **Seligman**
  Statistics is the science which deals with the methods of collecting, classifying, presenting, comparing, and interpreting numerical data to throw light on any sphere of enquiry.
- **A.L. Bowley** Statistics is the science of counting, and is used to draw inferences on the basis of numerical data."

Scope of Statistics

1. Policy Planning

Statistical data helps policymakers assess past trends and future projections to design effective and result-oriented policies.

2. Management

In organizations, statistics assists in monitoring employee performance and organizational growth for informed decision-making.

3. Social Sciences

Statistics supports the analysis of both qualitative and quantitative data to interpret social behaviour and predict trends.

4. Commerce and Accounts

It is essential in managing finances, cost-benefit analysis, and investment planning to ensure economic efficiency.

5. Industries

From production to employee welfare, statistical data helps industries optimize resources and reduce unnecessary expenses.

6. Sciences and Mathematics

Statistics provides precise tools to measure and analyse results in pure sciences, and it supports mathematical applications by quantifying variations.

7. Problem Solving

By comparing variables and identifying patterns, statistics helps individuals find optimal solutions and minimize errors.

8. Theoretical Research

Researchers use statistical tools to validate theories by establishing the relevance and significance of observed data and patterns.

Importance of Statistics

1. Decision-Making Tool

Statistics provides data-based insights to help individuals, businesses, and governments make informed decisions.

2. Planning and Forecasting

It helps in predicting future trends based on historical data—useful in business forecasting, budgeting, and policy planning.

3. Simplifies Complex Data

Through averages, percentages, graphs, and charts, statistics presents large data sets in a clear and understandable form.

4. Essential for Research

In scientific and social research, statistics helps in collecting, analysing, and interpreting data to validate hypotheses.

5. Quality Control

Industries use statistical methods to maintain and improve product quality through control charts and defect analysis.

6. Aids in Comparison

Statistics allows comparison between groups, time periods, or regions, helping identify patterns and differences.

7. Economic and Social Planning

Governments use statistics in national planning, population studies, employment policies, and development programs.

8. Risk Management

In sectors like insurance, finance, and health, statistics assesses probabilities and helps in managing risk.

Limitations of Statistics

1. Deals Only with Aggregates

Statistics applies to groups or collections of data, not to individual cases or single observations.

2. Limited to Quantifiable Data

It is most effective with numerical data. Qualitative data must be converted into numbers for statistical analysis.

3. Indirect Application to Qualitative Phenomena

Qualitative aspects like emotions or opinions need numerical scales or ratings for statistical treatment.

4. Not Always Precisely Accurate

Statistical conclusions are based on averages and probabilities, so they may not yield exact results like mathematical formulas.

Statistical investigation:

Statistical investigation involves the systematic collection, analysis, interpretation, and presentation of data to draw meaningful conclusions. It begins with defining the objective or problem, followed by designing the study and determining the population or sample. Data collection methods are then chosen, such as surveys, experiments, or observations. Once the data is gathered, it is organized and analysed using statistical techniques like measures of central tendency, dispersion, or hypothesis testing. The results are interpreted to provide insights or inform decisions Finally, the findings are presented in reports, charts, or graphs Statistical investigation is crucial in fields like economics, business, heath, and social sciences, enabling informed decision-making based on empirical evidence.

Planning and Organization Statistics:

Planning and Organization in Statistics refers to the methodical preparation and management of the various steps involved in a statistical investigation. Effective planning ensures that the study is well-structured, objective-driven, and produces reliable results

1. Defining Objectives:

The first step is to clearly define the purpose of the investigation. This involves specifying the problem or questions to be answered, which guides all subsequent steps.

2. Identifying the Population and Sample:

Decide whether the Investigation will cover the entire population or a sample. If a sample is chosen, determine the sampling technique (e.g., random, stratified) to ensure it is representative of the population.

3. Determining Data Collection Methods:

Select appropriate methods for data collection-surveys, experiments, observations, or secondary data sources. The choice depends on the nature of the study and available resources.

4. Resource Allocation:

Proper planning Involves managing resources like time, manpower, and budget. A well-organized reinvestigation ensures that these resources are used efficiently.

5. Data Collection and Organization:

Systematically collect data according to the predefined methods Organize the data for easy access and often using tools like spreadsheets, database, or specialized software.

6. Selection of Statistical Tools:

Determine which states (e.g., descriptive statistics inferential statistics) will be used for analysis based on the type of data and objectives.

7. Data Analysis:

Analyse the collected data sing appropriate statistical methods. The planning phase ensures that the data analysis is aligned with the research objectives and is carried out accurately.

8. Presentation of Finding:

Plan how the results will be parented through reports, charts, or visualisations to effectively commentate the Conclusion and insights.

9. Review and Adjustment:

Evaluate the process at various stages, making adjustments as needed. Continuous monitoring ensures that the investigation remains on track and produces reliable results.

Statistical units:

Statistical units are the basic entities or elements about which data is collected and analysed in a statistical investigation. They are classified into two main types

1. Investigation Units: The objects, individuals, or phenomena being studied, such as people, households, businesses, or events.

2. Analysis Units: The units on which data is processed or aggregated, like averages, totals, or proportions.

Statistical units can further be categorized as primary units (directly observed) and secondary units (derived or aggregated from primary data) The choice of statistical unit depends on the study's objectives and determines how data is collected, organized, and interpreted Proper identification of statistical units ensures accurate representation and meaningful analysis in statistical research.

Methods of Statistical Investigation:

1. Census Method:

This method involves collecting data from every unit in the population. It is comprehensive and provides highly accurate results but is time-consuming and costly. It is typically used in national censuses and large-scale studies.

2. Sample Survey Method:

Instead of surveying the entire population, a representative sample is selected. This method is more practical quicker, and cost-effective. Sampling techniques include random sampling, stratified sampling, and cluster sampling, ensuring that the sample accurately represents the population.

3. Observational Method:

In this method, data is collected by observing subjects in their natural environment without interference. It is commonly used in fields like sociology, psychology, and market research. Observational studies can be participant based or non-participant-based.

4. Experimental Method

This involves conducting experiments where conditions are controlled to test hypotheses. It is commonly used in scientific research to determine cause-and-effect relationships. Experimental designs include controlled trials, lab experiments, and field experiments.

5. Survey Method: Surveys involve collecting data through questionnaires, interviews, or online forms. They are widely used in market research, social sciences, and public opinion polling. Surveys can be conducted face-to-face, over the phone, or digitally.

6. Case Study Method:

A detailed analysis is conducted on a specific instance or case, often to explore unique or complex phenomena. The method is qualitative and typically used in social sciences, business, and psychology.

7. Secondary Data Analysis:

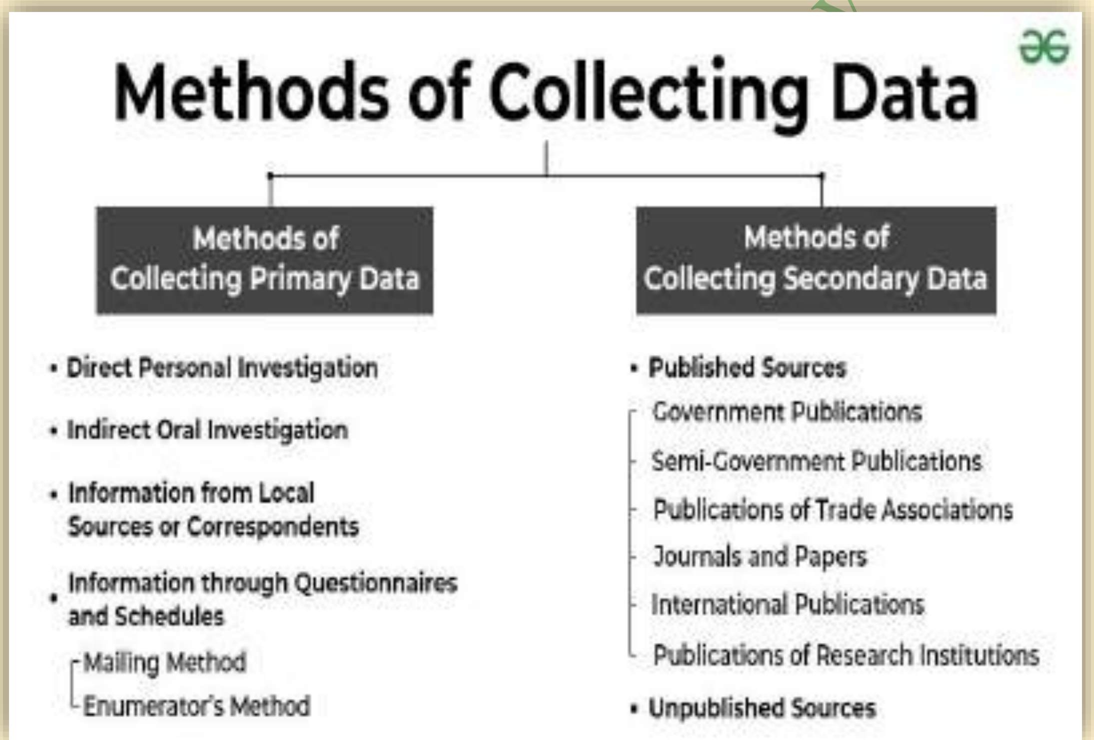
This method involves analysing existing data collected by other organizations, such as government reports, industry studies, or historical records. It is cost-effective and time-saving but may be Limited by the quality of the original data.

8. Simulation Method:

Simulations use mathematical models to replicate real-world scenarios. It is used when direct experimentation is impractical or impossible, often in operations research, economics, and engineering.

Data Collection:

Data Collection is the process of collecting information from relevant sources to find a solution to the given **statistical inquiry**. Collection of Data is the first and foremost step in a statistical investigation. It's an essential step because it helps us make informed decisions, spot trends, and measure progress.



There are two different methods of collecting data

Primary Data

Secondary

Primary Data

Primary data refers to **information collected directly from first-hand sources** specifically for a particular research purpose. This type of data is gathered through **various methods, including**

surveys, interviews, experiments, observations, and focus groups. One of the main advantages of primary data is that it **provides current, relevant, and specific information** tailored to the researcher's needs, offering a **high level of accuracy** and **control over data quality.**

Methods of Collecting Primary Data

There are a number of methods of collecting primary data, some of the common methods are as follows:

1. Interviews: Collect data through direct, one-on-one conversations with individuals. The investigator asks questions either directly from the source or from its indirect links.

1. **Direct Personal Investigation: T**he method of direct personal investigation involves collecting data personally from the source of origin. In simple words, the investigator makes direct contact with the person from whom he/she wants to obtain information. **For example,** direct contact with the household women to obtain information about their daily routine and schedule.
2. **Indirect Oral Investigation:** In the indirect oral investigation method of collecting primary data, the investigator does not make direct contact with the person from whom he/she needs information, instead they collect the data orally from some other person who has the necessary required information. **For example,** collecting data of employees from their superiors or managers.

- **Advantage:** Provides real-time, natural data; no reliance on self-reported information.
- **Disadvantage:** Observer bias; limited to what can be seen; may influence subjects' behaviour.
- **Suitable Use Case:** Behavioural studies, user experience research.

2. Questionnaires: Collect data by asking people a set of questions, either online, on paper, or face-to-face. In this method the investigator prepares a questionnaire to collect **Information through Questionnaires and Schedules,** while keeping in mind the motive of the study, . The investigator can collect data through the questionnaire in two ways:

1. **Mailing Method:** This method involves mailing the questionnaires to the informants for the collection of data. The investigator attaches a letter with the questionnaire in the mail to define the purpose of the study or research.

2. **Enumerator's Method:** This method involves the preparation of a questionnaire according to the purpose of the study or research. However, in this case, the enumerator reaches out to the informants himself with the prepared questionnaire.

- **Advantage:** Can reach a large audience quickly and cost-effectively.
- **Disadvantage:** Responses may be biased or inaccurate; low response rates.
- **Suitable Use Case:** Customer satisfaction surveys, market research.

3. Observations: The observation method involves collecting data by watching and recording behaviours, events, or conditions as they naturally occur. The observer systematically watches and notes specific aspects of a subject's behavior or the environment, either covertly or overtly.

- **Advantage:** Provides real-time, authentic data without reliance on self-reported information.
- **Disadvantage:** Observer bias can influence the results, and the presence of an observer might alter subjects' behaviour.
- **Suitable Use Case:** Studying user interactions with a product in a natural setting, monitoring wildlife behaviour, or assessing classroom dynamics.

4. Experiments: The experiment method involves manipulating one or more variables to determine their effect on another variable, within a controlled environment. Researchers create two groups (control and experimental), apply the treatment or variable to the experimental group, and compare the outcomes between the groups.

- **Advantage:** Allows for the establishment of cause-and-effect relationships with high precision.
- **Disadvantage:** Experiments can be artificial, limiting the ability to generalize findings to real-world settings, and they can be resource-intensive.
- **Suitable Use Case:** Testing the efficacy of a new drug, assessing the impact of a new teaching method, or evaluating the effect of a marketing campaign.

5. Focus Group: The focus group method involves gathering a small group of people to discuss a specific topic or product, facilitated by a moderator. A group of 6-12 participants engages in a guided discussion led by a moderator who asks open-ended questions to elicit opinions, attitudes, and perceptions.

- **Advantage:** Provides in-depth insights and diverse perspectives through interactive discussions, revealing the reasoning behind participants' thoughts and feelings.
- **Disadvantage:** Results can be influenced by dominant participants or groupthink, and the findings are not easily generalizable due to the small, non-representative sample size.
- **Suitable Use Case:** Exploring customer attitudes towards a new product, gathering feedback on a marketing campaign, or understanding public opinion on social issues.

6. Information from Local Sources or Correspondents: In this method, for the collection of data, the investigator appoints correspondents or local persons at various places, which are then furnished by them to the investigator. With the help of correspondents and local persons, the investigators can cover a wide area.

Secondary Data

Secondary data refers to **information that has already been collected, processed, and published by others**. This type of data can be sourced from **existing research papers, government reports, books, statistical databases, and company records.** The advantage of secondary data is that it is readily available and **often free or less expensive to obtain** compared to primary data. It saves time and resources since the data collection phase has already been completed.

Methods of Collecting Secondary Data

Secondary data can be collected through different published and unpublished sources. Some of them are as follows:

1. Published Sources

- **Government Publications:** Government publishes different documents which consists of different varieties of information or data published by the Ministries, Central and State Governments in India as their routine activity. As the government publishes these Statistics, they are fairly reliable to the investigator. **Examples** of Government publications on Statistics are the Annual Survey of Industries, Statistical Abstract of India, etc.
- **Semi-Government Publications:** Different Semi-Government bodies also publish data related to health, education, deaths and births. These kinds of data are also reliable and

used by different informants. Some **examples** of semi-government bodies are Metropolitan Councils, Municipalities, etc.

- **Publications of Trade Associations:** Various big trade associations collect and publish data from their research and statistical divisions of different trading activities and their aspects. **For example**, data published by Sugar Mills Association regarding different sugar mills in India.
- **Journals and Papers:** Different newspapers and magazines provide a variety of statistical data in their writings, which are used by different investigators for their studies.
- **International Publications:** Different international organizations like IMF, UNO, ILO, World Bank, etc., publish a variety of statistical information which are used as secondary data.
- **Publications of Research Institutions:** Research institutions and universities also publish their research activities and their findings, which are used by different investigators as secondary data. For example, National Council of Applied Economics, the Indian Statistical Institute, etc.

2. Unpublished Sources

Unpublished sources are another source of collecting secondary data. The data in unpublished sources is **collected by different government organizations and other organizations**. These organizations usually **collect data for their self-use and are not published anywhere**. For example, research work done by professors, professionals, teachers and records maintained by business and private enterprises.

Conclusion

Data collection is the **backbone of any research** or statistical investigation, providing the necessary information to **make informed decisions, identify trends, and measure progress**. By understanding the various **methods of data collection**—such as direct personal investigation, indirect oral investigation, questionnaires, observations, experiments, and focus groups—researchers can choose the most suitable approach to gather primary data that is current, relevant, and accurate. Similarly, using **secondary data** from published and unpublished sources like government reports, trade associations, and research institutions can save time and resources while offering valuable insights. Mastering these data collection techniques ensures the reliability and validity of the research, ultimately leading to sound and actionable conclusions.

Data Editing:

After collecting data (from surveys, interviews, forms, etc.), it may contain **errors**, **incomplete responses**, or **inconsistent entries**. Before using this data for analysis, it must be **cleaned and corrected**. This process is called **Data Editing**.

Data Editing is the process of reviewing, checking, and correcting collected data to ensure that it is **accurate, complete, consistent**, and **ready for analysis**.

It is an important step in data processing, just like coding, classification, and tabulation.

Objectives of Data Editing

- To detect and correct errors in data
- To ensure completeness of responses
- To maintain consistency in data
- To improve the quality and reliability of data
- To prepare data for analysis

Steps Involved in Data Editing

1. **Review all data entries** for readability and completeness
2. **Check for missing responses** or unanswered questions
3. **Correct obvious errors** (e.g., wrong age, date, gender, etc.)
4. **Ensure consistency** (e.g., a 10-year-old cannot have a PhD)
5. **Handle ambiguous or unclear answers** by referring back to respondents if possible
6. **Remove duplicates** and fix formatting or spelling issues

Classification of data:

Classification means grouping of related facts into different classes. The method of arranging data into homogeneous classes according to the common features present in the data is known as classification. **Classification of data is a function very similar to that of sorting letters in a post office.** categories based on common characteristics, to make it easier to understand, analyse, and interpret.

Bases of Classification of Data:

Classification of data is the process of arranging data in an orderly manner. four important bases of classification are discussed below:

1. **Qualitative Base:**

classification of data according to characteristics and attributes is called qualitative classification of data. In such a classification of data; data are categorized based on some attributes or quality such as gender, honesty, hair colour, literacy, intelligence, religion, etc.

2. **Quantitative Base:**

Quantitative classification of data refers to variables of quantities that can be either estimated or operated on. This implies in contrast to qualitative classification, quantitative classification of data enables the numerical distribution of data into classes.

3. **Geographical Base:**

The classification of data according to location is what classification is called a geographical classification of data. Ex. Eastern Region, Western Region, Southern Region, Northern Region.

4. **Chronological Base:**

In chronological classification of data, data are classified on the basis of time of existence, such as years, months, weeks, days, etc. In such a type of classification, data are arranged either in ascending or descending order with reference to time such as years, quarters, months, weeks, days etc. Chronological data classification is also known as temporal classification of data

Frequency Distribution

A frequency distribution shows the frequency of repeated items in a graphical form or tabular form. It gives a visual display of the frequency of items or shows the number of times they occurred. For examples marks scored by students, temperatures of different towns, points scored in a volleyball match, etc.

After data collection, we have to show data in a meaningful manner for better understanding. Organize the data in such a way that all its features are summarized in a table. This is known as ***frequency distribution***.

Definition of Frequency Distribution

1) According to Erricker:

"A classification according to the number possessing same value of the variable".

2) According to Croxton and Cowden:

"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".

Types of Frequency Distribution:

There are several types of frequency distributions that are commonly used to summarize and analyze data. The choice of a specific type depends on the nature of the data and the objectives of the analysis. Here are a few types of frequency distributions:

1) Discrete or Ungrouped Frequency Distribution:

This is the simplest form of frequency distribution where the individual values of a dataset are listed along with their frequencies. Each unique value has its frequency count displayed in the distribution. In this form of distribution, the frequency refers to discrete (countable) value. Here the data are presented in a way that exact measurement of units is clearly indicated. The process of preparing this type of distribution is very simple. We have just to count the number of times a particular value is repeated, which is called the frequency of that class.

Example:
Number of children in families

No. of Children	Frequency
0	2
1	5
2	7
3	3

2) Continuous or Grouped Frequency Distribution:

In cases where the dataset has a large range of values, it is often helpful to group the values into intervals or classes. The grouped frequency distribution displays the intervals or classes along with their corresponding frequencies. Continuous series is one where measurements are only approximations and are expressed in class intervals.

Example:

Marks in a test:

Marks Range	Frequency
0 – 10	3
11 – 20	5
21 – 30	8
31 – 40	4

3) Cumulative Frequency Distribution:

This type of distribution shows the cumulative frequencies up to a certain value or class. It provides information about the number or proportion of data points that fall below or equal to a particular value. A cumulative distribution of frequencies shows the number of data items with values less than or equal to the upper-class limit of each class. While a cumulative relative frequency distribution gives the proportion of the data items and a cumulative percentage frequency distribution shows the percentage of data items with values less than or equal to the upper-class limit of each class.

Example (from above):

Marks Range	Frequency	Cumulative Frequency
0 – 10	3	3
11 – 20	5	8
21 – 30	8	16
31 – 40	4	20

4) Relative Frequency Distribution:

Instead of showing the actual frequencies, the relative frequency distribution displays the proportions or percentages of values within each class. It is calculated by dividing the frequency of each class by the total number of data points.

Example:

If total students = 20

Marks Range	Frequency	Relative Frequency
0 – 10	3	0.15
11 – 20	5	0.25
21 – 30	8	0.40
31 – 40	4	0.20

5) Cumulative Relative Frequency Distribution:

Similar to the cumulative frequency distribution, this type of distribution shows the cumulative relative frequencies up to a certain value or class. It provides information about the proportion or percentage of data points that fall below or equal to a particular value.

Example:

Marks Range	Relative Frequency	Cumulative Relative Frequency
0 – 10	0.15	0.15
11 – 20	0.25	0.40
21 – 30	0.40	0.80
31 – 40	0.20	1.00

Statistical Series:

Statistical Series refers to a collection of data arranged in a specific order to show the frequency, variation, and distribution of a particular phenomenon. These series are used in various research studies to represent the data in a structured and organized manner, making it easier to interpret and analyse.

For example, let's say you want to study the consumption pattern of a specific product in a particular region. You can collect the data related to the product's sales, customer feedback, market trends, etc., and arrange them in a chronological or geographical order to form a Statistical Series. This series will help you understand the pattern of consumption, the peak season, customer preference, etc.

A **statistical series** is a more general term for any ordered presentation of statistical data. It can be based on values, time, location, etc.

Types of Statistical Series

There are mainly two types of Statistical Series:

Time Series:

Time series is a type of Statistical Series where data is collected over time at regular intervals. This type of series is used to represent the trend, seasonality, and cyclical variations in a phenomenon. The data collected in a time series can be represented using various charts and graphs such as line charts, bar graphs, histograms, etc.

Cross-sectional Series:

Cross-sectional series is a type of Statistical Series where data is collected at a particular point in time. This type of series is used to represent the variations and differences between the characteristics of different groups or entities. The data collected in a cross-sectional series can be represented using various charts and graphs such as pie charts, stacked bar charts, etc.

Importance of Statistical Series

Statistical Series plays a crucial role in research studies as they help in:

- Understanding the pattern and trend of a phenomenon over time or space
- Identifying the peak season, slow season, and cyclical variations in a phenomenon
- Analysing the frequency and distribution of a phenomenon
- Comparing and contrasting the characteristics of different groups or entities
- Visualizing and presenting the data in an organized and structured manner

Conclusion:

In conclusion, Statistical Series is a valuable tool in research studies that can help researchers represent and analyse data in an efficient and structured manner.

- ✓ A frequency distribution is a type of statistical series.
- ✓ All frequency distributions are statistical series, but not all statistical series are frequency distributions.

Tabulation of Data Diagrammatic and Graphical Presentation of Data

1.2 Tabulation

1.2.1 What is a Table

A table is a symmetric arrangement of statistical data in rows and columns.

1.2.2 Definitions

According Prof. L.R.Connor,"

"Table involves the orderly and systematic presentation of numerical data in a form designed to elucidate the problem under consideration."

According to Prof. M.M. Blaire

"Table in its broadest sense is an orderly arrangement of data in column and rows."

1.2.3 Meaning

In the light of above mentioned definitions we can say in brief, "Table is systematic organization and presentation of data in the form of rows and columns. Whereas rows are horizontal arrangements and columns are vertical arrangements.

Tabulation is the process of presenting data in a systematic and organized table with rows and columns.

➤ Purpose:

- To simplify large data.
- To make comparison easier.
- To highlight important features of data.

➤ Example:

Fruit	Number of Students
Apple	30
Banana	20
Orange	25
Grapes	15

Diagrammatic Presentation of Data

Diagrammatic presentation is the visual form of presentation of data in which facts are highlighted in the language of diagrams.

- It consists in presenting statistical material in interesting and attractive geometrical figures (Bars, Circle, Rectangle, Squares), pictures, maps and charts etc.
- It will attract the attention of a large number of persons.
- It facilitates comparison between two or more sets of data.

Diagrammatic presentation refers to the **visual representation of data** using diagrams or charts to make it easier to understand and interpret.

➤ Advantages:

- Quick and easy understanding.
- Helps in visual comparison.
- Highlights trends and patterns.

➤ Types of Diagrams:

1. **Bar Diagram** – Used for comparing categories.
2. **Pie Chart** – Shows percentage distribution.
3. **Histogram** – For continuous frequency data.
4. **Line Graph** – Shows data changes over time.
5. **Pictogram** – Uses symbols or images to represent data.

Conclusion:

Tabulation	Diagrammatic Presentation
Data shown in tables (rows & columns)	Data shown visually using diagrams/charts
Good for detailed study	Good for quick understanding
Easy to reference	Easy to interpret at a glance

Graphical Presentation of Data

Definition:

Graphical presentation of data is a method of presenting numerical information using **visual formats** like charts, graphs, or diagrams.
It helps to **easily interpret, compare, and analyse** data at a glance.

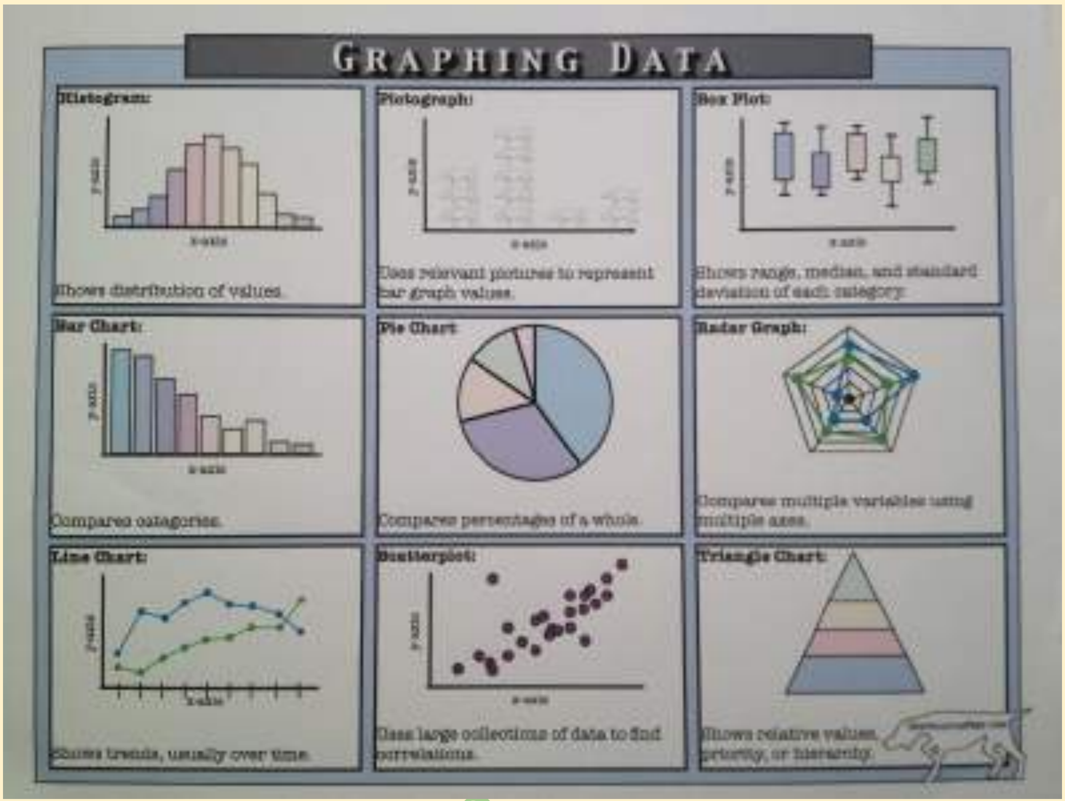
Importance

- Makes complex data simple and understandable.
- Helps identify **trends**, **patterns**, and **outliers**.

- Useful in **reports, presentations, and analysis**.
- Appeals to **visual learners** and saves time.

Types of Graphical Presentations

Here are the main types:



1. Bar Diagram

- Used for **comparing categories** (e.g., sales by product).
- Bars can be **vertical** or **horizontal**.
- Length of bar = size of data.

Example: Comparing favourite fruits of students.

2. Pie Chart

- A **circular chart** divided into sectors.
- Each sector shows **percentage or proportion** of the total.
- Best for **showing parts of a whole**.

Example: Budget distribution, population by religion.

3. Histogram

- Used for **continuous data** grouped into intervals.
- Similar to bar graphs, but **bars touch each other**.
- Shows **frequency distribution**.

Example: Age distribution of a group.

4. Line Graph

- Represents data points connected by lines.
- Best for **showing trends over time**.

Example: Temperature change across months.

5. Pictogram (Pictograph)

- Uses **symbols or pictures** to represent data.
- Often used for **children or simple presentations**.

Example: Using 🍎 to represent 10 apples sold.

Advantages of Graphical Presentation

Benefit	Explanation
Quick Understanding	Easy to grasp large data visually
Better Comparison	See differences and similarities easily
Visual Appeal	Engaging and effective in reports
Highlights Trends	Time-based data trends are easily seen

Disadvantages:

Limitation	Explanation
May Oversimplify	Important details can be lost
Requires Accuracy	Poor scaling or labelling can mislead
Not Always Precise	Better for overview than exact figures

Conclusion:

Graphical presentation is a powerful tool to **simplify, visualize, and communicate data** effectively. It should be chosen based on the **type of data** and the **message** you want to convey.