7

MEASURES OF CENTRAL TENDENCY

Raw data are difficult to comprehend. Classification facilitates, many a time, quick and easy understanding of diversified nature of data. A single representative value serves the purpose in a better manner.

Quantitative data in a mass exhibit certain general characteristics. They show a tendency to concentrate at certain values, usually somewhere in the centre of the distribution. Measures of this tendency are called measures of central tendency or averages. This tendency toward centralization, though not universal, has established the expression "measure of central tendency" to describe an average. The term is imbedded in statistical language, but it is not always pertinent.

(Simpson and Kafka in Basic Statistics - Page 127)

An average is a value which is typical or representative of a set of data. (Murray R. Spiegel in Theory and Problems of Statistics-Page 45)

A measure of central tendency gives a single representative value for a set of usually unequal values. The single value is the point of location around which the individual values of the set cluster. The measures of central tendency are hence known as 'measures of location'. They are popularly called averages. Various measures of central tendency are the following:

1. Arithmetic Mean
2. Median
3. Mode
4. Geometric Mean
5. Harmonic Mean

Weighted Arithmetic Mean and positional values, viz., Quartiles, Deciles and Percentiles are discussed later. Moving Averages are considered in the chapter titled 'Analysis of Time Series'.

Besides them other averages are also there as seen overleaf.

ARITHMETIC MEAN

Definition : *Arithmetic Mean is the total of the values of the items divided by their number.*

A.M. is the abbreviation and $\overline{X}$ (read 'X bar') is the symbol for arithmetic mean. Arithmetic mean is also called mean and average (singular).

Methods of Finding Arithmetic Mean:

All the possible seven types of data are considered. Mean can be calculated by

(i) Direct Method

(ii) Short - cut Method and

(iii) Step Deviation Method.

All the methods give the same result for a problem. All the methods are illustrated.

Data - Type I (Individual Observations or Raw Data)

When the observed values are given individually such as X_1, X_2, X_3, ... X_N the methods of calculation of arithmetic mean are as follows.

Direct Method:

$$\text{Arithmetic Mean} = \frac{\text{Total of the observations}}{\text{Number of the observations}}$$

$$= \frac{X_1 + X_2 + X_3 + \dots\dots + X_N}{N}$$

$$= \frac{\sum X}{N}$$

The symbol $\sum$ (read 'sigma') denotes 'sum' or 'total'. $\sum X$ denotes the total of all the 'X' values. The calculation consists of the following two steps:

Step 1. Denote the given observations by X and find their total, $\sum X$.

Step 2. Identify N, the number of observations and divide $\sum X$ by N.

Example 1 : The expenditure of 10 families in Rupees are given below.

	A	B	C	D	E	F	G	H	I	J
Family Expenditure	30	70	10	75	500	8	42	250	40	36

Calculate the arithmetic mean.

Solution : X - Expenditure; N = 10.

Family	Expenditure (Rs.) X
A	30
B	70
C	10
D	75
E	500
F	8
G	42
H	250
I	40
J	36
Total	$\Sigma X = 1061$

$$\overline{X} = \frac{\Sigma X}{N}$$

$$= \frac{1061}{10}$$

$$= 106.1$$

The arithmetic mean is Rs. 106.10.

Short cut Method. Arithmetic mean may be obtained also as $\overline{X} = A + \left(\frac{\Sigma d}{N}\right)$

where A - assumed mean or arbitrary origin and

d (=X–A) are the deviations of the observations (X) from assumed mean (A) and

Σd is the total of the deviations (differences, X–A).

Any value between the minimum and the maximum of the values of X is assumed as A if it is not specified in a problem. A may or may not be one of the given X values. $\overline{X}$ is the same whatever value is assumed for A.

The following four steps are involved in the calculation.

Step 1: Choose certain value for A if it is not specified in a problem.

Step 2. Find the deviations of X from A. That is, calculate d = X–A corresponding to each X.

Step 3. Find Σd.

Step 4. Identify N and find $\overline{X} = A + \left(\frac{\Sigma d}{N}\right)$

Illustration-1: *Calculate mean $\bar{x}$ from the following data:*

Table 6.1

Roll Nos.	1	2	3	4	5	6	7	8	9	10
Marks	33	35	44	34	41	45	39	46	36	47

1. Direct Method

Steps: 1. Add up all the values of the variables x (marks) and find out Σx.

2. Divide Σx by their number of observation (N).

Solution

Table 6.2

Roll Nos.	Marks x
1	33
2	35
3	44
4	34
5	41
6	45
7	39
8	46
9	36
10	47
N = 10	$\Sigma x = 400$

Formula

$$\bar{x} = \frac{\Sigma x}{N}$$

where

$\bar{x}$ = Arithmetic mean

Σx = Sum of variables

N = Number of observations

The methods of calculation of arithmetic mean are illustrated below.

Direct Method: X_1 occurs f_1 times. The total of those f_1 values $= f_1 X_1$. X_2 occurs f_2 times. The total of those f_2 values $= f_2 X_2$. X_3 occurs f_3 times. The total of those f_3 values $= f_3 X_3$. In this manner the total of all the $N(=f_1+f_2+f_3+\ldots)$ values$=f_1X_1+f_2X_2+f_3X_3+\ldots\ldots = \sum fX$

Thus, arithmetic mean $= \frac{\sum fX}{N}$

The calculation consists of the following four steps:

Step 1. Form a table with columns X and f.

Step 2. Form the next column with title fX. Multiply the values of X and f in pairs and enter the products in that column.

Step 3. Find $N(=\sum f)$ and $\sum fX$.

Step 4. Divide $\sum fx$ by N to get the value of $\overline{X}$.

Example 5: Calculate the mean number of persons per house. Given

No. of persons per house	2	3	4	5	6	Total
No. of houses	10	25	30	25	10	100

Solution : X-No. of persons per house; f-No.of houses.

No. of persons per house X	No.of houses f	fx
2	10	20
3	25	75
4	30	120
5	25	125
6	10	60
Total	N=100	$\sum fx$=400

Mean

$$\overline{X} = \frac{\sum fx}{N} = \frac{400}{100} = 4$$

Short cut Method.

The formula for arithmetic mean:

B. For Discrete Series

1. Direct Method

In this method, the values of the variable are multiplied by their respective frequencies and the products so obtained and totalled. This total is divided by the total number of frequencies.

Steps : 1. Multiply each variable by its frequency (fx)

2. Add all the fx (Σfx)

3. Divide Σfx by the total of frequency (N) or Σf

The ***formula is***

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

where

$\bar{x}$ = Arithmetic mean

Σfx = the sum of products

Σf = total of frequency

Illustration-3 : *Calculate the mean for the following data:*

Table 6.5

No. of children born per family (x)	0	1	2	3	4	5	6
No. of families (f)	7	7	10	5	3	2	1

Solution

Table 6.6: Calculation of Mean.

x	f	fx
0	7	0
1	7	7
2	10	20
3	5	15
4	3	12
5	2	10
6	1	6
	$\Sigma f = 35$	$\Sigma fx = 70$

Formula

$$\bar{x} = \frac{\Sigma fx}{\Sigma f}$$

$$\bar{x} = \frac{70}{35} = 2$$

$$\bar{x} = 2$$

2. Short-cut Method

Steps : 1. Take any value from the variables (X) as assumed mean (A).

2. Find out deviations of each variable from the assumed mean (d = x-A) (where d = deviation, x = variable A = Assumed mean)

3. Multiply the deviation with the respective frequencies (d x f = fd)

4. Add all the products = Σfd

$$\bar{x} = 2 \pm \left[\frac{0}{35}\right]$$

$$\bar{x} = 2 \pm 0$$

$$\bar{x} = 2$$

C. For Continuous Series

In continuous frequency distribution, the value of each individual frequency distribution is unknown. Therefore on the assumption that the frequency of the class intervals is concentrated at the centre that the mid point of each class interval has to be found out.

1. Direct Method

Steps : 1. Find out the mid value of each class. The mid value is obtained by adding the lower limit and upper limit of the class and dividing the total by two. For e.g., in a class interval say 10-20, the mid value is 15.

i.e., $\left[\frac{10+20}{2} = \frac{30}{2} = 15\right] = (\text{mid } x)$

2. Multiply the mid value of each class by the frequency of the class. In other words ***mid x*** will be multiplied by ***f***.

3. Add all the products (Σf ***mid x***)

4. Σf ***mid x*** is divided by Σf

5. Apply the ***formula***

$$\bar{x} = \frac{\Sigma f \text{ mid } x}{\Sigma f}$$

where

$\bar{x}$	=	Arithmetic mean
Σf mid x	=	the sum of products
Σf	=	total of frequency

Illustration-5 : *From the following, find out the mean:*

Table 6.9

Marks (x)	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students (f)	10	18	20	26	30	28	18

Solution

Table 6.10

Marks (x)	Midpoint (mid x)	No. of Students (f)	f mid x
10-20	15	10	150
20-30	25	18	450
30-40	35	20	700
40-50	45	26	1170
50-60	55	30	1650
60-70	65	28	1820
70-80	75	18	1350
		$\Sigma f = 150$	$\Sigma f \text{ mid } x = 7290$

Formula

$$\bar{x} = \frac{\Sigma f \text{ mid } x}{\Sigma f}$$

$$\bar{x} = \left[\frac{7290}{150}\right]$$

$$= 48.6$$

$$\bar{x} = 48.6$$

The average mark is 48.6

2. Short-cut Method

Steps : 1. Find the mid-value of each class (mid x).

2. Assume anyone of the mid value as an assumed mean (A).

Solution :

Step 1. Form a table with columns Marks and Number of Students.

Step 2. Form m, the mid values of the class intervals column. $m = \frac{\text{Lower boundary} + \text{Upper boundary}}{2}$.

For the first class interval 20–30, $m = \frac{20+30}{2} = 25.$

Step 3. Find the products of f and m pair wise and enter them in the next column.

Step 4. Find $N(=\Sigma f)$ and Σfm.

Step 5. Divide Σfm by N to find $\overline{X}$.

Marks	No.of Students f	Mid Value m	fm
20-30	5	25	125
30-40	8	35	280
40-50	12	45	540
50-60	15	55	825
60-70	6	65	390
70-80	4	75	300
Total	N=50	—	Σfm =2460

Arithmetic Mean

$$\overline{X} = \frac{\Sigma fm}{N}$$

$$= \frac{2460}{50}$$

$$= 49.20$$

Example 9: From the following data, compute arithmetic mean by short cut method.

Marks Obtained:	0-10	10-20	20-30	30-40	40-50	50-60
No. of Students :	5	10	25	30	20	10

(B.B.M. Bharathiar, N 01)

Solution :

Step 1. Form a table with columns Marks Obtained and No. of Students.

MEDIAN

Definition: *Median is the value of the middle most item when all the items are in the order of magnitude.*

M or Me denotes median.

Arithmetic mean is calculated on the basis of **magnitudes** or **values** of all the items. But median is concerned with the **position** or **place** of the items in a series. 'Which is the middle most item?' is the question.

Median divides the series into two equal parts. Half of the items will be equal to or less than the median; half of the items will be equal to or more than the median.

II. Median

***Median** is an average which divides a distribution into 2 equal halves.* When the given values are arranged in an ascending order or descending order, that value which is in the centre is the median (middle value). In otherwords, there will be an equal number of items both above or below the median. Median is represented by the letter 'Md'. Like mean, median can also be calculated for

a. ungrouped data
b. discrete series
c. continuous series ⟶ grouped data

Median

***Median** is the **middle value** of a data when the values are arranged in the ascending or descending order.*

Median is an ***average.*** It is a ***measure of central value.***

Median divides a distribution into two equal halves. There will be equal number of items above and below the items.

Median is represented by the symbol ***md.***

Median can be calculated for ungrouped data and grouped data.

The formula for the calculation of median for ungrouped data is

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

md = median
N = Number of items

If there are ***odd number*** of items, then median is calculated as follows:

$$Md = \frac{11+1}{2} = \frac{12}{2} = 6$$

Median = Value of the 6^{th} item, when the items are arranged in an ascending order.

When there are ***even number*** of items, the median falls between two items. The values of these two items are added and divided by 2 to get median.

If there are 12 items, the median is calculated as follows:

$$Md = \frac{12+1}{2} = \frac{13}{2} = 6.5$$

Median = Value of the 6.5^{th} item. It is obtained by adding the values of 6^{th} item and 7^{th} item and dividing the sum by 2.

Median of a grouped data with class interval can be calculated by the following formula:

$$Median = L + \left(\frac{\frac{N}{2} - cf}{f} \right) \times C$$

L = Lower limit of the median class.

N = Total frequency.

cf = Cumulative frequency prior to the median class.

C = Class interval of the median class.

f = Frequency of the median class.

Merits of Median

1. Simple to calculate.
2. It can be calculated without knowing the values of all the items.
3. It is unaffected by extreme values.
4. It can be calculated graphically.

Demerits

1. It is not based on all the items.
2. It is not used as a common average.
3. It is not used for further statistical calculation

A. Calculation of Median
Individual Series or Ungrouped Data or Raw Data or Individual Observation

Steps : 1. Arrange the data in ascending or descending order.

2. Apply the ***formula***

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

Illustration -13 : *The following are the marks scored by 11 students; find out the median marks.*

Table 6.19

15	18	10	14	20	9	21	30	6	10	13

Solution

First rearranging the given values in ascending order

6, 9, 10, 10, 13, 14, 15, 18, 20, 21, 30.

Apply the ***formula***

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

where

$$N = \text{no.of items}$$
$$N = 11$$

$$\therefore \quad Md = \frac{11+1}{2} = \frac{12}{2} = 6^{th} \text{ value}$$

6^{th} value is = 14

$\therefore$ Md = 14

Under even numbers

Illustration -14 : *Find out median from the following:*

Table 6.20

5	11	6	10	14	21

Solution

First rearranging the given values in ascending order.

5, 6, 10, 11, 14, 21

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

$$N = 6$$

$$\therefore \quad Md = \frac{6+1}{2} = \frac{7}{2} = 3.5^{th} \text{ value}$$

In the data, 3^{rd} value is 10

4^{th} value is 11

$$\therefore \quad 3.5^{th} \text{ value} = \frac{10+11}{2} = \frac{21}{2} = 10.5$$

$$\therefore \quad \text{Median} = 10.5$$

B. Calculation of Median - Discrete Series

Steps : 1. Arrange the data in ascending or descending order.

2. Find the cumulative frequencies.

3. Apply the ***formula***

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

Illustration -15: *Find the median size of shoes.*

Table 6.21 from the following data

Size of shoes in inches	Frequency
4	10
5	15
6	22
7	16
8	12
9	5

Solution

First find out less than cumulative frequency.

Table 6.22

Size of shoes in inches (x)	Frequency (f)	Less than cumulative frequency (cf)
4	10	10
5	15	25
6 ←	22	47
7	16	63
8	12	75
9	5	
	N = 80	

Apply the ***formula***

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

$$= \frac{80+1}{2} = \frac{81}{2}$$

$= 40.5^{th}$ value

Here 40.5th value is in between 25 and 47 of cumulative frequency. So we take higher cumulative frequency is 47. So we take the corresponding x value of cf 47. Here the corresponding x value of cf 47 is 6.

$\therefore$ Median = 6

So median size of shoes is 6 inches.

C. Calculation of Median - Continuous Series

Steps : 1. Find the cumulative frequencies.

2. Find out the median class by using $\frac{N}{2}$.

3. Apply the ***formula***

$$\text{Median} = L + \left[\frac{N/2 - cf}{f}\right] \times C$$

where

L = lower limit of the median class
N = total number of items = Sf
cf = cumulative frequency prior to the median class
f = actual frequency of the median class
C = class interval of the median class

Illustration - 16: *Calculate the median from the following table:*

Table 6.23

Marks	Frequency
0-10	22
10-20	38
20-30	46
30-40	34
40-50	20

Solution

Table 6.24 : First find out less than cumulative frequency.

Marks (x)	Frequency (f)	Less than cumulative frequency (cf)
0-10	22	22
10-20	38	60
20-30 ←	46	106
30-40	34	140
40-50	20	
	N = 160	

N = 160

Next, find out the median class by using $\frac{N}{2}$

$$= \frac{N}{2} = \frac{160}{2} = 80$$

The value 80 is in between ***60*** and ***106*** of cumulative frequency. So we take higher cumulative frequency, namely 106.

Now we take the corresponding class of cumulative frequency 106.

Here the corresponding class of cf 106 is ***20 - 30***

So 20 - 30 is the median class.

20 is the lower limit of the median class.

46 is the actual frequency of the median class.

60 is cumulative frequency (cf) prior to the median class.

Apply the ***formula***

$$\text{Median} = L + \left[\frac{N/2 - cf}{f}\right] \times C$$

$$= 20 + \left[\frac{160/2 - 60}{46}\right] \times 10$$

C = class interval

$$= 20 + \left[\frac{80 - 60}{46}\right] \times 10$$

$$= 20 + \left[\frac{20}{46}\right] \times 10$$

$$= 20 + (0.434 \times 10)$$

$$= 20 + 4.34$$

$$= 24.34$$

Graphic Location of Median

There are two methods of locating median graphically.

1. By an Ogive

An ogive is a cumulative frequency curve. It may be more than cumulative frequency curve or less than cumulative frequency curve. Variables on the X-axis and the cumulative frequencies are on the Y-axis. The middle item is then marked on vertical scale by the formula N/2. A line parallel to the base is drawn, cutting the ogive at any point draw a line perpendicular to the base and the median is read off.

Illustration - 17

Table 6.25

Marks	No.of students	Marks	No.of students
0-10	5	50-60	25
10-20	6	60-70	10
20-30	8	70-80	8
30-40	12	80-90	6
40-50	16	90-100	4

Example 35: Calculate the median from the following

Marks	10-25	25-40	40-55	55-70	70-85	85-100
Frequency	6	20	44	26	3	1

(B.B.M. MK, N 01)

Solution :

Marks	Frequency f	Cumulative Frequency cf
10- 25	6	6
25- 40	20	26
40- 55	44	70 ←
55- 70	26	96
70- 85	3	99
85-100	1	100
Total	N=100	---

$\frac{N}{2} = \frac{100}{2} = 50$

∴ Median class interval: 40–55

∴ L = 40; f = 44; cf = 26; i = 55–40 = 15

As $$M = L + \left[\frac{i(N/2 - cf)}{f}\right],$$

$$M = 40 + \left[\frac{15(50-26)}{44}\right]$$

$$= 40 + \left[\frac{15 \times 24}{44}\right]$$

$$= 40 + 8.18$$

$$= 48.18 \simeq 48$$

Note: When the class intervals are not continuous, find the difference between the lower limit of an interval and the upper limit of the preceding interval when they are in ascending order. Find half of the difference and add it to each upper limit and subtract it from each lower limit. Resulting class intervals are called 'True Class Intervals'. They are continuous.

III. Mode

Mode is the most common item of a series. It is defined as *the value of the variable which occurs most frequently in a distribution*. It is repeated the highest number of times in the series. Like median, it is also a positional average which can be located by inspection.

When a distribution has one concentration of frequency, it is often called ***"unimodal"*** on the otherhand, when it has 2 concentrations it is termed as ***"bimodal"*** similarly if 3 concentrations it is termed as ***"trimodal"***. etc.

Merits of Mode

1. It is easy to understand.
2. It is simple to calculate.
3. It is unaffected by extreme values.
4. It is a positional average and can be located easily by inspection.
5. It can be determined by the graphic method.

Demerits

1. It is an average, which is ill-defined and indeterminate.
2. It is not further used in algebraic calculations.
3. In the case of bimodal class, the calculation is difficult as it involves grouping and analysis tables.
4. It is not based on all observations.

Mode

***Mode** is the value of the variable which occurs most frequently in a distribution.*

The value which occurs many times in the table is the mode.

It is represented by the letter ***Mo.***

Mode is an ***average.*** It is a ***positional average.*** It is a ***measure of central value.***

When a data has one concentration of frequency, it is called ***unimodal.*** When it has two concentrations, it is called ***bimodal.*** When it has 3 concentrations of frequency, it is called ***trimodal.***

Mode can be calculated for ***ungrouped data*** and ***grouped data.***

To find out mode of an ungrouped data, the values are arranged in an ascending order. The value which occurs maximum number of times is the mode.

18, 21, 23, 23, ***25,25,25,*** 27, 29, 29.

In the above data, 25 occurs maximum number of times. So ***25*** is the ***mode.***

The mode of a ***discrete distribution*** is the value of the variable which shows maximum frequency.

No.of count trees	10	11	12	13	14	15	16
No. of coconuts (Frequency)	8	4	12	24	26	7	11

In the above table, ***14*** is the mode because the values are maximum here.

Merits of Mode

1. Mode can be easily found out.
2. No calculation is needed.
3. It is not affected by extreme values.
4. It can be calculated graphically.

Demerits of Mode

1. It is not clearly defined.
2. It is not based on all observations.
3. It is not reliable.
4. It is not used for further statistical calculation.

A. Calculation of Mode : Individual Observation or Ungrouped Data or Raw Data

Step 1 : The data have to be arranged in the form of an array so that the value which has the highest frequency can be known

or

Rearranging the data into a discrete series. Find out the highest frequency.

Illustration - 18: *Determine mode from the following data:*

Table 6.27

50	62	48	50	63	65	50	48	43	62	50	50

Solution

Table 6.28 : First the data is arranged in the form of an array.

43	48	48	***50***	***50***	***50***	***50***	***50***	62	62	63	65

In this data, 50 is repeated 5 times

So Mode is ***50***

or

Rearranging the data into discrete series. It is comparatively easy.

Table 6.29

x	43	48	50	62	63	65	
f	1	2	5	2	2	2	$\Sigma f = 14$

Here the value 50 is repeated 5 times

$\therefore$ Mode = **50**

B. Calculation of Mode : Discrete Series

Step - 1 : It can be find out even by inspection i.e, which variable (x) has the highest frequency is the Mode.

Illustration - 19 : *Determine mode from the following data:*

Table 6.30

x	20	25	30	35	40	45	50
f	1	2	1	5	1	2	1

Solution

Here the value 35 is repeated 5 times

So mode is **35**

Sometimes we cannot depend on the method of inspection to find out the mode. In such situations, it is suggested to prepare a ***grouping table*** and an ***analysis table*** to find out the mode. First prepare grouping table and then an analysis table.

Illustration - 20 : *From the following data of the height of 100 plants in a garden determine the modal height.*

Table 6.31

Height (x) (cm)	58	60	61	62	63	64	65	66	68	70
Plants (No.) (f)	4	6	5	10	20	22	24	6	2	1

Solution

By inspection we can clearly say that the modal height is 65cm, because the value 65 is repeated 24 times. But in this problem, the difference between the maximum frequency and the next frequency is very small is 24 - 22 = 2. So prepare grouping table and analysis table.

(In the case of ***bimodal*** series or ***trimodal*** series, we must prepare first grouping table and analysis table).

Steps for Preparing the Grouping and Analysis Table

1. Prepare a grouping table with 6 columns.
2. Write the size of item in the margin.
3. In column 1, write the frequencies against the respective items.

4. In column 2, the frequencies are grouped in *twos* (1 and 2; 3 and 4; 5 and 6 and so on).

5. In column 3, the frequencies are grouped in *twos*; leaving the first frequency (2 and 3; 4 and 5; 6 and 7and so on).

6. In the column 4, the frequencies are grouped in *threes* (1, 2 and 3; 4,5 and 6; 7,8 and 9 and so on).

7. In the column 5, the frequencies are `grouped in *threes* leaving the first frequency (2,3 and 4; 5,6 and 7; 8,9 and 10 and so on).

8. In the column 6, the frequencies are grouped in *threes* leaving the first two frequencies (3,4 and 5; 6,7 and 8;9 ,10 and 11 and so on).

In all the processes mark down, the maximum frequencies by a circle.

9. Then an analysis table is prepared to show the exact size, which has the highest frequency.

Table 6.32: Grouping Table.

Height in cm	Frequencies					
	Column1	Column2	Column3	Column4	Column5	Column6
58	4		—	15	—	—
60	6	10	11		21	—
61	5					
62	10	15				35
63	20		30			
64	22	42		52		
65	24		46			
66	6	30		32	66	
68	2		8			52
70	1	3			9	

Table 6.33: Analysis Table.

Col. No.	Height in cm.									
	58	60	61	62	63	64	65	66	68	70
1							1			
2					1	1				
3						1	1			
4				1	1	1				
5					1	1	1			
6						1	1	1		
Total				1	3	5	4	1		

Since the value 64 has occurred the maximum number of times. i.e., 5 times. ∴ The modal height is 64 cm. But by inspection in the data, one is likely to say that the modal height is 65, since it occurs the maximum number of times i.e., 24 which is incorrect as revealed by grouping and analysis table.

C. Calculation of Mode : Continuous Series

Step : 1. The highest frequency can be find out by inspection.

In the case of bimodal series or trimodal series we prepare. Grouping and analysis table and then find out highest frequency.

2. Apply the ***formula***

$$\text{Mode} = L + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] \times C$$

where

L = lower limit of the modal class

Δ_1 = the difference between the frequency of the modal class and the preceding modal class $(f_1 - f_0)$

(iii) the sizes or lengths of the class intervals are equal.

The data are to be rearranged, if necessary. For the fulfilment of the third condition, the class intervals may be revised such that their sizes become equal. The frequencies are to be adjusted without changing the contents of the original data.

Or else, mode can be estimated from the following **empirical relation** between mean, median and mode.

Mode = 3 Median − 2 Mean

Such a relation is found to exist in **moderately skewed** series.

2. Identify the modal class interval. It is the class interval which has the greatest frequency density. In many cases, it is the one which has the greatest frequency. If necessary form the grouping table and the analysis table as explained for discrete series to identify the modal class interval. Instead of X consider the class intervals while forming the two tables.

3. **Case I.** If the class interval with the greatest frequency is identified as the modal class interval, apply the formula

$$Z = L + \left[\frac{i(f_1 - f_0)}{2f_1 - f_0 - f_2}\right]$$

Z = Mode
L = Lower boundary of the modal class interval
f_1 = Frequency of the modal class
f_0 = Frequency of the class preceding the modal class
f_2 = Frequency of the class succeeding the modal class
i = Size or length of the modal class interval
= Its upper boundary − its lower boundary.

The same is found by the following form of the formula which is more convenient:

$$Z = L + \left[\frac{iD_1}{(D_1 + D_2)}\right]$$

Z, L and i are defined as above.

$D_1 = f_1 - f_0$ = Frequency of the modal class − Frequency of the class preceding the modal class.

$D_2 = f_1 - f_2$ = Frequency of the modal class − Frequency of the class succeeding the modal class.

Murray R. Spiegel uses L_1, Δ_1, Δ_2 and C instead of l_1, D_1, D_2 and i.

or

Case II. If the class interval other than the one with greatest frequency is identified as the modal class interval by the grouping table and the analysis table, apply the formula

$$Z = L + \left[\frac{i f_2}{(f_1 + f_2)}\right]$$

Z = Mode
L = Lower boundary of the modal class interval
f_1 = Frequency of the class preceding the modal class
f_2 = Frequency of the class succeeding the modal class
i = Size or length of the modal class interval
= Its upper boundary − its lower boundary.

Note: The value of mode lies within the modal class interval in either case.

Example 45: Calculate the mode.

Daily Wage in Rs.	50-60	60-70	70-80	80-90	90-100
No. of Labourers	40	62	75	100	65

Solution : Greatest frequency = 100 and the modal class interval is 80-90.

(80-90 may be ascertained to have greatest frequency density by grouping table and analysis table)

L = 80, the lower boundary of the modal class interval,
f_1 = 100, the frequency of the modal class,
f_0 = 75, the frequency of the class preceding the modal class and
f_2 = 65, the frequency of the class succeeding the modal class and
i = 90 − 80 = 10, the size of the modal class interval.

$$D_1 = f_1 - f_0 = 100 - 75 = 25$$

$$D_2 = f_1 - f_2 = 100 - 65 = 35$$

$$Z = L + \left[\frac{(D_1)}{(D_1 + D_2)}\right]$$

$$= 80 + \left[\frac{10 \times 25}{(25 + 35)}\right]$$

$$= 80 + \left[\frac{250}{60}\right]$$

$= 80 + 4.17$

$=$ Rs. 84.17

Example 46 : Find out mode for the following data using grouping and analysis table.

Class Interval:	0-5	5-10	10-15	15-20
Frequency :	9	12	15	16
Class Interval:	20-25	25-30	30-35	35-40
Frequency :	17	15	10	13

(B.Com. Bharathiar, A 01)

Solution: A grouping table and an analysis table are formed as explained earlier.

Grouping Table

Class Interval	Frequency (1)	(2)	(3)	(4)	(5)	(6)
0- 5	9					
		21				
5-10	12			36		
			27			
10-15	15				**43**	
		31				
15-20	16					**48**
			33			
20-25	**17**			**48**		
		32				
25-30	15				42	
			25			
30-35	10					38
		23				
35-40	13					

Analysis Table

Class Interval	(1)	(2)	(3)	(4)	(5)	(6)	Total
0- 5							–
5-10					1		1
10-15					1	1	2
15-20			1	1	1	1	4
20-25	1	1	1	1		1	5
25-30		1		1			2
30-35							–
35-40							–

Modal class interval : 20-25

$L = 20$; $i = 25-20 = 5$; $D_1 = 17-16 = 1$; $D_2 = 17-15 = 2$

$$\text{Mode, } Z = L + \left[\frac{iD_1}{(D_1 + D_2)}\right]$$

$$= 20 + \left[\frac{5 \times 1}{(1+2)}\right]$$

$$= 20 + \left[\frac{5}{3}\right]$$

$$= 20 + 1.67$$

$$= 21.67$$

Example 47 : Calculate the mode.

Interval :	0-2	2-4	4-6	6-8	8-10	10-12	12-14	14-16
Frequency:	45	50	65	70	30	25	20	18

Solution : For finding the interval which has the greatest frequency density in this example, grouping table and analysis table are formed.

Grouping Table

Interval	Frequency (1)	(2)	(3)	(4)	(5)	(6)
0- 2	45					
		95				
2- 4	50			(160)		
			(115)			
4- 6	65				(185)	
		(135)				
6- 8	(70)					(165)
			100			
8-10	30			125		
		55				
10-12	25				75	
			45			
12-14	20					63
		38				
14-16	18					

Analysis Table

Interval	(1)	(2)	(3)	(4)	(5)	(6)	Total
0- 2				1			1
2- 4			1	1	1		3
4- 6		1	1	1	1	1	5
6- 8	1	1			1	1	4
8-10						1	1
10-12							–
12-14							–
14-16							–

The interval 4-6 does not have greatest frequency. But it has greatest frequency density. Consider

$$Z = L + \left[\frac{i\, f_2}{(f_1 + f_2)}\right]$$

L = 4, the lower boundary of the modal class,

i = 6–4 = 2, the size of the modal class interval,

f_1 = 50, the frequency of the class preceding the modal class

f_2 = 70, the frequency of the class succeeding the modal class.

GEOMETRIC MEAN

Definition: *Geometric mean of N values is the Nth root of the product of the N values.*

If $X_1, X_2, X_3, \ldots\ldots X_N$ are the values, their geometric mean is $\sqrt[N]{X_1 . X_2 . X_3 \ldots\ldots\ldots\ldots X_N}$

G.M. is the abbreviation.

To avoid the difficulty in multiplying all the values and finding the appropriate root of the product, log. is used.

Formulae:

$$G.M. = \text{Antilog}\left(\frac{\Sigma \log X}{N}\right) \text{ for individual observations}$$

$$= \text{Antilog}\left(\frac{\Sigma f \log X}{N}\right) \text{ for discrete series}$$

$$= \text{Antilog}\left(\frac{\Sigma f \log m}{N}\right) \text{ for continuous series.}$$

Similarly, weighted geometric mean

$$= \text{Antilog}\left(\frac{\Sigma W \log X}{\Sigma W}\right)$$

Calculation of G.M. in individual observations has the following five steps.

Step 1. Form a table with the given values of X in the first column.

Step 2. Use a calculator or refer to a logarithm table and make note of the logarithm of each X in the next column under the title log X.

Step 3. Find $\Sigma \log X$

Step 4. Identify N and divide $\Sigma \log X$ by N.

Step 5. Using a calculator or an antilogarithm table, find Antilog $\left(\frac{\Sigma \log X}{N}\right)$ which is the G.M.

Example 70: Find the geometric mean of 3 6 24 48
(I.C.W.A. Foundation, J 99 and J 01)

Solution:

X	log X
3	0.4771
6	0.7782
24	1.3802
48	1.6812

$$GM = \text{Antilog}\left(\frac{\sum \log X}{N}\right)$$

$$= \text{Antilog}\left(\frac{4.3167}{4}\right)$$

$$= \text{Antilog}(1.0792)$$

$$= 12.00$$

Note: It is seen that

$$G.M. = \sqrt[4]{3 \times 6 \times 24 \times 48}$$

$$= \sqrt[4]{3 \times 6 \times 2 \times 12 \times 4 \times 12}$$

$$= \sqrt[4]{12 \times 12 \times 12 \times 12}$$

$$= 12.$$

For calculating G.M. from a discrete series, the following six steps are used.

Step1. Form a table with the given values of X and f in the first two columns.

Step 2. Use a calculator or refer to a logarithms table and make note of the logarithm of each X in the next column under the title log X.

Step 3. Multiply f and log X in pairs and enter the products in the next column under the title f log X.

Step 4. Find $N(=\sum f)$ and $\sum f \log X$.

Step 5. Divide $\sum f \log X$ by N.

Step 6. Using a calculator or an antilogarithms table find $\text{Antilog}\left(\frac{\sum f \log X}{N}\right)$ which is the G.M.

Example 71: Calculate Geometric mean for the data given below:

X	10	15	25	40	50
f	4	6	10	7	3

(B.Com. Bharathidasan, A [illegible])

Solution:

X	f	logX	f logX
10	4	1.0000	4.0000
15	6	1.1761	7.0566
25	10	1.3979	13.9790
40	7	1.6021	11.2147
50	3	1.6990	5.0970
Total	N=30	--------	$\Sigma f \log X =$ 41.3473

$$G.M. = \text{Antilog}\left(\frac{\Sigma f \log X}{N}\right)$$

$$= \text{Antilog}\left(\frac{41.3473}{30}\right)$$

$$= \text{Antilog}\,(1.3782)$$

$$= 23.89$$

When X is replaced by m in the formula for discrete series, the formula for continuous series is obtained.

The steps in the calculation of G.M. are similar in discrete series and continuous series.

Whatever be the form in which a continuous series is given, identify the mid values (m) of the class intervals and the class frequencies (f) first. Then proceed as in a discrete series.

Example 72: Compute the Geometric mean of the following series:

Marks :	0-10	10-20	20-30	30-40	40-50
No.of students :	5	7	15	25	8

(B.Com. Madras, O 01; B.Com.(C.A.) Bharathiar, N 01)

Solution:

Marks	No.of Students f	Mid Values m	$\log m$	$f \log m$
0-10	5	5	0.6990	3.4950
10-20	7	15	1.1761	8.2327
20-30	15	25	1.3979	20.9685
30-40	25	35	1.5441	38.6025
40-50	8	45	1.6532	13.2256
Total	N=60	----	------	$\Sigma f \log m$ = 84.5243

$$G.M. = \text{Antilog}\left(\frac{\Sigma f \log m}{N}\right)$$

$$= \text{Antilog}\left(\frac{84.5243}{60}\right)$$

$$= \text{Antilog}(1.4087)$$

$$= 25.63.$$

When f are replaced by W in the formula for G.M. of a discrete series, the formula for weighted G.M. is obtained. Hence, the steps for calculation of weighted G.M. are similar to those of a discrete series.

Example 73 : Calculate weighted G.M.

Commodity	A	B	C	D
Weight	1	6	3	2
Price (Rs.)	5	17	30	42

Solution :

Commodity	Weight W	Price (Rs.) X	log X	W log X
A	1	5	0.6990	0.6990
B	6	17	1.2304	7.3824
C	3	30	1.4771	4.4313
D	2	42	1.6232	3.2464
Total	ΣW =12	---	-----	$\Sigma W \log x$ = 15.7591

HARMONIC MEAN

Definition : *Harmonic Mean is the reciprocal of the mean of the reciprocals of the values.*

If $X_1, X_2, X_3 \ldots . X_N$ are the values, their reciprocals are $\frac{1}{X_1}, \frac{1}{X_2}, \frac{1}{X_3} \ldots . \frac{1}{X_N}$. The total of the reciprocals is $\Sigma\left(\frac{1}{X}\right)$

The mean of the reciprocals is $\frac{\Sigma\left(\frac{1}{X}\right)}{N}$

$\therefore$ the reciprocal of the mean of the reciprocals is $\frac{N}{\Sigma\left(\frac{1}{X}\right)}$.

$$\text{i.e. H.M.} = \frac{N}{\frac{1}{X_1} + \frac{1}{X_2} + \ldots . . \frac{1}{X_N}}$$

H.M. is the abbreviation.

Formulae:

$$\text{H.M} = \frac{N}{\Sigma\left(\frac{1}{X}\right)} \text{ for individual observations.}$$

$$= \frac{N}{\Sigma\left(\frac{f}{X}\right)} \text{ for a discrete series.}$$

$$= \frac{N}{\Sigma\left(\frac{f}{m}\right)} \text{ for a continuous series.}$$

Similarly, weighted harmonic mean $= \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$

Calculation of H.M. in individual observations has the following four steps :

Step 1. Form a table with the values of X in the first column.

Step 2. Use a calculator or a table and find the value of $\frac{1}{X}$ corresponding to each X. Write those values in the next column under the title $\frac{1}{X}$.

Step 3. Identify N and find $\Sigma\left(\frac{1}{X}\right)$.

Step 4. Divide N by $\Sigma\left(\frac{1}{X}\right)$ to get the H.M.

Note : In other calculations, certain values are divided by N. In H.M., N is divided by $\Sigma\left(\frac{1}{X}\right)$.

Example 74: Find the Harmonic Mean for the following individual data :

6, 15, 35, 40, 900, 520, 300, 400, 1800, 2000.

(B.Com. Madras, A 99)

Solution:

Value X	$\frac{1}{X}$
6	0.1667
15	0.0667
35	0.0286
40	0.0250
900	0.0011
520	0.0019
300	0.0033
400	0.0025
1800	0.0006
2000	0.0005
Total	$\Sigma\left(\frac{1}{X}\right)$ =0.2969

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{1}{X}\right)}$$

$$= \frac{10}{0.2969}$$

$$= 33.68$$

The following are the steps to calculate the H.M. of a discrete series:

Step 1 : Form a table with the values of X and f in the two columns.

Step 2: Divide f by X and enter the quotients in the next column under the title $\left(\frac{f}{X}\right)$

Step 3: Find N (= Σf) and $\Sigma\left(\frac{f}{X}\right)$

Step 4: Divide N by $\Sigma\left(\frac{f}{X}\right)$ to get the H.M.

Example 75 : Calculate the harmonic mean from the following data.

X:	10	12	14	16	18	20
f:	5	18	20	10	6	1

(B.Com. MK, N 2K; B.B.A. Bharathidasan, N 98)

Solution:

X	f	f/X
10	5	0.5000
12	18	1.5000
14	20	1.4286
16	10	0.6250
18	6	0.3333
20	1	0.0500
Total	N=60	4.4369

$$H.M. = \frac{N}{\Sigma\left(\frac{f}{X}\right)}$$

$$= \frac{60}{4.4369}$$

$$= 13.52$$

When X is replaced by m in the formula for H.M. of a discrete series, the formula for H.M. of a continuous series is obtained. Identify the mid values of the class intervals (m) and the class frequencies (f). Then the steps for the calculation of H.M. are similar to those of a discrete series.

Example 76: Calculate the Harmonic Mean for the following data.

Value :	0-10	10-20	20-30	30-40	40-50
Frequency :	8	12	20	6	4

(B.Com.(C.A) Bharathiar, A O1)

Solution:

Value	Frequency(f)	Mid Value(m)	f/m
0-10	8	5	1.6000
10-20	12	15	0.8000
20-30	20	25	0.8000
30-40	6	35	0.1714
40-50	4	45	0.0889
Total	N=50	----	3.4603

$$H.M. = \frac{N}{\Sigma\left(\frac{f}{m}\right)} = \frac{50}{3.4603} = 14.45$$

The steps in the calculation of weighted Harmonic Mean are similar to those of Harmonic Mean of a discrete series. W replaces f.

Example 77 : Calculate weighted Harmonic Mean.

Value :	1	2	5	10	20
Weight :	2	5	10	5	2

Solution:

Value X	Weight W	$\frac{W}{X}$
1	2	2.0
2	5	2.5
5	10	2.0
10	5	0.5
20	2	0.1
Total	W= 24	$\Sigma\left(\frac{W}{X}\right)$ = 7.1

Weighted Harmonic Mean

$$= \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$$

$$= \frac{24}{7.1}$$

$$= 3.38$$

Special Use of Harmonic Mean

Average speed is the distance travelled per hour. ∴ 'total distance ÷ total time in hours' gives the average speed.

Without those calculations, average speed can be found out as follows:

8

MEASURES OF DISPERSION

In a series, all the items are not equal. There is difference variation among the values. The degree of variation is luated by various measures of dispersion.

Averages are central values. They enable comparison of or more sets of data. They are not sufficient to depict the nature of the sets. For example, consider the following rks of two students.

Student I	Student II
68	85
75	90
65	80
67	25
70	65

th have got a total of 345 and an average of 69 each. The ct is that the second student has failed in one paper. When he averages alone are considered, the two students are equal.

Less variation is a desirable characteristic. First student as less variation. That is, he is almost equally good in all the subjects. To quote Simpson and Kafka, "*An average does not ll the full story. It is hardly fully representative of a mass, nless we know the manner in which the individual items scatter around it. A further description of the series is necessary if we are to guage how representative the average is*".

Consider the following example.

Group I

	Student C		Student D		Student E	
	Mark	Deviation from Mean	Mark	Deviation from Mean	Mark	Deviation from Mean
	X	X−$\overline{X}$	X	X−$\overline{X}$	X	X−$\overline{X}$
	60	0	50	−10	20	−40
	60	0	55	− 5	40	−20
	60	0	60	0	60	0
	60	0	65	5	80	20
	60	0	70	10	100	40
Total:	300		300		300	
Mean($\overline{X}$):	60		60		60	

RANGE

Definition : *Range is the difference between the greatest (largest) and the smallest of the values.*

In symbols, Range = L–S

L - Largest Value

S - Smallest Value

In individual observations and discrete series, L and S are easily identified. In continuous series, the following two methods are followed.

Method 1 :

L - Upper boundary of the highest class

S - Lower boundary of the lowest class

Method 2 :

L - Mid value of the highest class

S - Mid value of the lowest class

Coefficient of Range = $\frac{L-S}{L+S}$

Example 1 : Find the value of range and its coefficient for the following data :

8 10 5 9 12 11

(B.Com. Bharathiar, A 02)

Solution : L = 12 S = 5

∴ Range = L – S

= 12 – 5

= 7

Coefficient of Range = $\frac{L-S}{L+S}$

= $\frac{12-5}{12+5}$

= $\frac{7}{17}$

= 0.4118

Example 2 : Calculate range and its coefficient from the following distribution:

Size :	60-62	63-65	66-68	69-71	72-74
Number :	5	18	42	27	8

(B.Com. Bharathiar, A 01)

Solution :

Method 1 : After rewriting the class intervals continuously, the lower boundary of the lowest class, S = 59.5 and the upper boundary of the highest class, L = 74.5.

$$\text{Range} = L - S = 74.5 - 59.5 = 15$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{74.5-59.5}{74.5+59.5} = \frac{15}{134} = 0.1119$$

Method 2 :

Mid value of the lowest class, S = 61

Mid value of the highest class, L = 73

$$\text{Range} = L - S = 73 - 61 = 12$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{73-61}{73+61} = \frac{12}{134} = 0.0896.$$

Uses of Range

Range is not a popular measure of dispersion. It is used in the following few situations.

1. Range is used in finding the control limits of Mean chart and Range chart in S.Q.C.

2. While quoting the prices of shares, bonds, gold, etc. on daily basis or yearly basis, the minimum and the maximum prices are mentioned.

3. The minimum and the maximum temperature likely to prevail on each day are forecasted.

6. It cannot be calculated ...

7. It is a very rarely used measure. Its scope is limited to three situations indicated earlier.

Definition 2M. & 5M

QUARTILE DEVIATION (Q.D.)

Definition : *Quartile Deviation is half of the difference between the first and the third quartiles.* Hence it is called semi **Inter Quartile Range.**

In symbols, Q.D. = $\frac{Q_3 - Q_1}{2}$ Q.D. is the abbreviation. Among the quartiles Q_1, Q_2 and Q_3, the range is $Q_3 - Q_1$. Hence, $Q_3 - Q_1$ is called inter quartile range and $\frac{Q_3 - Q_1}{2}$, semi inter quartile range.

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

As mentioned in the previous chapter, 25% of the values are below or equal to Q_1 ; 25% above or equal to Q_3. $Q_3 - Q_1$ is the distance between Q_1 and Q_3. Central 50% of the items lie between Q_1 and Q_3. It is customery to consider $\frac{Q_3 - Q_1}{2}$ as an absolute measure of dispersion.

Definitions and calculations of Q_1 and Q_3 for all types of data were considered in the previous chapter.

Example 3 : What do you mean by Quartile Deviation? Find the Quartile Deviation for the following:-

391, 384, 591, 407, 672, 522, 777, 733, 1490, 2488.

(B.Com. Madras, S 74)

Solution: The given values in ascending order :

384, 391, 407, 522, 591, 672, 733, 777, 1490, 2488.

Position of Q_1 is $\frac{N+1}{4} = \frac{10+1}{4} = 2.75.$

$$\therefore Q_1 = 2^{nd} \text{ value} + 0.75 \,(3^{rd} \text{ value} - 2^{nd} \text{ value})$$
$$= 391 + 0.75\,(407-391)$$
$$= 391 + 0.75 \times 16$$
$$= 391 + 12$$
$$= 403$$

Position of Q_3 is $3\left(\frac{N+1}{4}\right) = 3 \times 2.75 = 8.25$

$$\therefore Q_3 = 8^{th} \text{ value} + 0.25\,(9^{th} \text{ value} - 8^{th} \text{ value})$$
$$= 777 + 0.25\,(1490 - 777)$$
$$= 777 + 0.25 \times 713$$
$$= 777 + 178.25$$
$$= 955.25$$

$$\therefore Q.D. = \frac{Q_3 - Q_1}{2}$$
$$= \frac{955.25 - 403.00}{2}$$
$$= \frac{552.25}{2}$$
$$= 276.13$$

Example 4: Weekly wages of a labourer are given below. Calculate Q.D. and coefficient of Q.D.

Weekly Wage (Rs.)	100	200	400	500	600	Total
No. of Weeks	5	8	21	12	6	52

...lution:

...ly (Rs)	No.of Weeks	Cum. Freq.
	5	5
	8	13
	21	34
	12	46
	6	52
	N=52	--

Position of Q_1 is $\frac{N+1}{4} = \frac{52+1}{4}$

$= 13.25.$

$\therefore Q_1 =$ 13th value + 0.25 (14th value – 13th value)

$= 200 + 0.25\,(400-200)$

$= 200 + 0.25 \times 200$

$= 200 + 50$

$= 250$

...ion of Q_3 is $3\left(\frac{N+1}{4}\right) = 3 \times 13.25 = 39.75$

... = 39th value + 0.75 (40th value – 39th value)

$= 500 + 0.75\,(500 - 500)$

$= 500 + 0.75 \times 0$

$= 500 + 0$

$= 500$

$Q.D. = \frac{Q_3 - Q_1}{2} = \frac{500-250}{2} = \frac{250}{2} = 125$

...oefficient of Q.D.= $\frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{500-250}{500+250} = \frac{250}{750} = 0.3333$

Example 5: For the data given here, give the quartile ...ation.

X	351-500	501-650	651-800	801-950	951-1100
f	48	189	88	47	28

(B.Com. Bharathidasan, A 01)

Solution :

X	f	True class intervals	Cumulative Frequency
351- 500	48	350.5 – 500.5	48
501- 650	189	500.5 – 650.5	237 ←
651- 800	88	650.5 – 800.5	325 ←
801- 950	47	800.5 – 950.5	372
951-1100	28	950.5 –1100.5	400
Total	N = 400	------	-----

$\frac{N}{4} = \frac{400}{4} = 100$; Q_1 class is 500.5 – 650.5

$\therefore L_1 = 500.5$; $f_1 = 189$; $i_1 = 650.5 - 500.5 = 150$; $cf_1 = 48$

$$\therefore Q_1 = L_1 + \left[\frac{i_1\left(\frac{N}{4} - cf_1\right)}{f_1}\right]$$

$$= 500.5 + \left[\frac{150(100 - 48)}{189}\right]$$

$$= 500.5 + \left[\frac{150 \times 52}{189}\right]$$

$$= 500.5 + 41.27$$

$$= 541.77$$

$\frac{3N}{4} = 3 \times 100 = 300$; Q_3 class is 650.5–800.5

$\therefore L_3 = 650.5$; $f_3 = 88$; $i_3 = 800.5 - 650.5 = 150$; $cf_3 = 237$

$$\therefore Q_3 = L_3 + \left[\frac{i_3\left(\frac{3N}{4} - cf_3\right)}{f_3}\right]$$

$$= 650.5 + \left[\frac{150(300 - 237)}{88}\right]$$

$$= 650.5 + \left[\frac{150 \times 63}{88}\right]$$

$$= 650.5 + 107.39$$

$$= 757.89$$

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

$$= \frac{757.89 - 541.77}{2}$$

$$= \frac{216.12}{2}$$

$$= 108.06$$

Merits

1. It is simple to understand and easy to calculate.
2. It is not affected by extreme items.
3. It can be calculated for data with open end classes also.

EAN DEVIATION OR AVERAGE DEVIATION

Definition: *Mean deviation is the arithmetic mean ie absolute deviations of the values about their metic mean or median or mode.*

M.D. is the abbreviation for Mean Deviation. There are kinds of mean deviations, viz.,

(i) mean deviation or mean deviation about mean
(ii) mean deviation about median
(iii) mean deviation about mode.

Mean deviation about median is the least. It could sily verified in individual observations and discrete series re the actual values are considered.

relative measures are the following:

Coefficient of Mean Deviation (about Mean)

$$= \frac{\text{Mean Deviation about Mean}}{\text{Mean}}$$

Coefficient of Mean Deviation about Median

$$= \frac{\text{Mean Deviation about Median}}{\text{Median}}$$

Coefficient of Mean Deviation about Mode

$$= \frac{\text{Mean Deviation about Mode}}{\text{Mode}}$$

ividual Observations

$$\text{Mean Deviation (about Mean)} = \frac{\Sigma\left|X - \overline{X}\right|}{N}$$

The mean, $\overline{X} = \frac{\Sigma X}{N}$ is calculated first. From each X,

Example 7 : Calculate mean deviation about median the items

7 4 10 9 15 12 7 9 7

(B.Com.Bharathidasan,A 01)

Solution :

X	$\|X-M\|$
4	5
7	2
7	2
7	2
9	0
9	0
10	1
12	3
15	6
Total	$\Sigma\|X-M\|=21$

Items are considered in ascending order.

Position of Median (M) is $\frac{N+1}{2} = \frac{9+1}{2} = 5$

$\therefore$ Median (Item at 5^{th} position) $= 9$

$$\text{M.D. about Median} = \frac{\Sigma|X-M|}{N}$$

$$= \frac{21}{9}$$

$$= 2.33$$

Note : If the Coefficient of M.D. about Median is required,

$$\text{Coefficient of M.D. about Median} = \frac{\text{M.D. about Median}}{\text{Median}}$$

$$= \frac{2.33}{9} = 0.2589$$

Example 8: Daily earnings in Rs.(X) of 10 coolies are given. Calculate all the three mean deviations and the corresponding relative measures.

X: 32 51 23 46 20 78 57 56 57 30

Solution :

X (Rs.)	$\|X-\bar{X}\|$ $\bar{X}=45$	$\|X-M\|$ M=48.5	$\|X-Z\|$ Z=57
20	25	28.5	37
23	22	25.5	34
30	15	18.5	27
32	13	16.5	25
46	1	2.5	11
51	6	2.5	6
56	11	7.5	1
57	12	8.5	0
57	12	8.5	0
78	33	29.5	21
$\Sigma X=$	$\Sigma\|X-\bar{X}\|=$ 150	$\Sigma\|X-M\|=$ 148.0	$\Sigma\|X-Z\|=$ 162

$$\overline{X} = \frac{\Sigma X}{N} = \frac{450}{10} = \text{Rs.}45$$

M.D. about Mean = $\frac{\Sigma|X - \overline{X}|}{N} = \frac{150}{10}$ = Rs.15

Coefficient of M.D about Mean = $\frac{\text{M.D.}}{\text{Mean}} = \frac{15}{45} = 0.3333$

Position of Median, M is $\frac{N+1}{2} = \frac{10+1}{2} = 5.5$

Median = $\frac{5^{th}\text{ item} + 6^{th}\text{ item}}{2} = \frac{46+51}{2}$ = Rs.48.50

M.D. about Median = $\frac{\Sigma|X - M|}{N} = \frac{148}{10}$ = Rs.14.80

Coefficient of M.D. about Median $= \frac{\text{M.D. about Median}}{\text{Median}}$

$= \frac{14.80}{48.50}$

$= 0.3052$

Mode, Z = Rs.57

Mean deviation about Mode $= \frac{\Sigma|X - Z|}{N}$

$= \frac{162}{10}$

= Rs. 16.20

Coefficient of M.D. about Mode $= \frac{\text{M.D. about Mode}}{\text{Mode}}$

$= \frac{16.20}{57}$

$= 0.2842$

Note: Mean Deviation about Median is the least.

Discrete Series

The measure of central tendency Mean or Median or Mode is calculated first. The following formulae are used later.

Mean Deviation (about Mean) $= \frac{\Sigma f|X - \overline{X}|}{N}$

Mean Deviation about Median $= \frac{\Sigma f|X - M|}{N}$

Mean Deviation about Mode $= \frac{\Sigma f|X - Z|}{N}$

Example 9 :

X	2	4	6	8	10
f	1	4	6	4	1

...mean deviation for the above data.

(B.Com. Bharathidasan, N 01)

Solution :

X	f	fX	$\|X-\bar{X}\|$ $\bar{X}=6$	$f\|X-\bar{X}\|$
2	1	2	4	4
4	4	16	2	8
6	6	36	0	0
8	4	32	2	8
10	1	10	4	4
Total	N = 16	ΣfX = 96	---	$\Sigma f\|X-\bar{X}\|$ = 24

Mean, $\bar{X} = \frac{\Sigma fX}{N}$

$= \frac{96}{16}$

$= 6$

Mean Deviation about Mean $= \frac{\Sigma f|X-\bar{X}|}{N}$

$= \frac{24}{16}$

$= 1.50$

Note : If the coefficient of M.D. about Mean is required,

Coefficient of M.D. about Mean = $\frac{\text{M.D. about Mean}}{\text{Mean}}$

$= \frac{1.50}{6}$

$= 0.2500$

Example 10 : Calculate all the three mean deviations and the corresponding coefficients of mean deviations :

Age (years) :	21	25	27	32	41	46	50	55
No.of Workers:	2	3	10	20	15	10	8	2

Solution :

Age X	No. of Workers f	fX	$\lvert X-\bar{X}\rvert$ $\bar{X}=37.31$	$f\lvert X-\bar{X}\rvert$	cf	$\lvert X-M\rvert$ M=36.5	$f\lvert X-M\rvert$	$\lvert X-Z\rvert$ Z=32	$f\lvert X-Z\rvert$
21	2	42	16.31	32.62	2	15.5	31.0	11	22
25	3	75	12.31	36.93	5	11.5	34.5	7	21
27	10	270	10.31	103.10	15	9.5	95.0	5	50
32	20	640	5.31	106.20	35	4.5	90.0	0	0
41	15	615	3.69	55.35	50	4.5	67.5	9	135
46	10	460	8.69	86.90	60	9.5	95.0	14	140
50	8	400	12.69	101.52	68	13.5	108.0	18	144
55	2	110	17.69	35.38	70	18.5	37.0	23	46
Total	N = 70	Σfx = 2612	--	$\Sigma f\lvert X-\bar{X}\rvert$ = 558.00	--	--	$\Sigma f\lvert X-M\rvert$ = 558.0	---	$\Sigma f\lvert X-Z\rvert$ = 558

Mean, $\bar{X} = \frac{\Sigma fX}{N} = \frac{2612}{70} = 37.31$

Mean Deviation about Mean $= \frac{\Sigma f\lvert X-\bar{X}\rvert}{N} = \frac{558.00}{70} = 7.97$

Coefficient of M.D. about Mean $= \frac{\text{M.D. about Mean}}{\text{Mean}}$

$= \frac{7.97}{37.31}$

$= 0.2136$

Position of Median is $\frac{N+1}{2} = \frac{70+1}{2} = 35.5$

Median (M) $= \frac{35^{th}\text{item} + 36^{th}\text{item}}{2}$

$= \frac{32+41}{2}$

$= 36.5$

Mean Deviation about Median $= \frac{\Sigma f\lvert X-M\rvert}{N}$

$= \frac{558}{70}$

$= 7.97$

Coefficient of M.D. about Median = $\frac{\text{M.D. about Median}}{\text{Median}}$

$= \frac{7.97}{36.5} = 0.2184$

$Z = 32$

Mode,

Mean Deviation about Mode $= \frac{\Sigma f|X - Z|}{N} = \frac{558}{70} = 7.97$

Coefficient of M.D. about Mode $= \frac{\text{M.D. about Mode}}{\text{Mode}}$

$= \frac{7.97}{32} = 0.2491$

Continuous Series

The measure of central tendency, Mean or Median or Mode, is calculated first using the appropriate formula. The formulae considered in discrete series are used to find the necessary mean deviations by introducing m in the place of X.

Example 11: Calculate the mean deviation from the mean for the following data:

Marks :	0-10	10-20	20-30	30-40	40-50	50-60	60-70
No.of Workers :	6	5	8	15	7	6	3

(C.A. Foundation, M 99)

Solution :

Marks	No.of Students f	Mid Value m	fm	$\|m - \bar{X}\|$ $\bar{X} = 33.40$	$f\|m - \bar{X}\|$
0-10	6	5	30	28.40	170.40
10-20	5	15	75	18.40	92.00
20-30	8	25	200	8.40	67.20
30-40	15	35	525	1.60	24.00
40-50	7	45	315	11.60	81.20
50-60	6	55	330	21.60	129.60
60-70	3	65	195	31.60	94.80
Total	N = 50	---	$\Sigma fm =$ 1670	---	$\Sigma f\|m - \bar{X}\| =$ 659.20

Mean, $\bar{X} = \frac{\Sigma fm}{N} = \frac{1670}{50} = 33.40$

Mean deviation about Mean $= \frac{\Sigma f|m - \bar{X}|}{N}$

$= \frac{659.20}{50}$

$= 13.18$

Example 12 : The following is the age distribution of 80 policy holders insured through an agent:

Age Group	Number of policy holders	Age Group	Number of policy holders
16-20	8	41-45	7
21-25	15	46-50	3
26-30	13	51-55	2
31-35	20	56-60	1
36-40	11		

Calculate mean deviation from the median.

(B.B.A. Bharathidasan, N 98)

Solution :

Age Group	No. of Policy Holders f	True Class Intervals	Cum. Freq. cf	Mid Value m	$\|m - M\|$ M = 31.5	$f\|m - M\|$
16-20	8	15.5-20.5	8	18	13.5	108.0
21-25	15	20.5-25.5	23	23	8.5	127.5
26-30	13	25.5-30.5	36	28	3.5	45.5
31-35	20	30.5-35.5	56 ←	33	1.5	30.0
36-40	11	35.5-40.5	67	38	6.5	71.5
41-45	7	40.5-45.5	74	43	11.5	80.5
46-50	3	45.5-50.5	77	48	16.5	49.5
51-55	2	50.5-55.5	79	53	21.5	43.0
56-60	1	55.5-60.5	80	58	26.5	26.5
Total	N = 80	-------	--	--	----	$\Sigma f\|m - M\| =$ 582.0

N/2 = 80/2 = 40, 30.5 - 35.5 is the median class

L = 30.5, f = 20, i = 35.5 - 30.5 = 5, cf = 36.

Median,

$$M = L + \left[\frac{i\left(\frac{N}{2} - cf\right)}{f}\right]$$

$$= 30.5 + \left[\frac{5\,(40 - 36)}{20}\right]$$

$$= 30.5 + \left[\frac{5 \times 4}{20}\right]$$

$$= 30.5 + 1$$

$$= 31.50$$

Mean deviation about median $= \frac{\sum f|m - M|}{N}$

$$= \frac{582}{80}$$

$$= 7.28$$

Example 13 : Calculate

(i) Mean Deviation about Mode and

(ii) Coefficient of Mean Deviation about Mode

Mid Value :	2.5	7.5	12.5	17.5	22.5	27.5
Frequency :	19	28	50	22	10	7

Solution :

Mid Value m	Frequency f	Class Interval	$\|m - Z\|$ Z = 12.2	$f\|m - Z\|$
2.5	19	0- 5	9.7	184.3
7.5	28	5-10	4.7	131.6
12.5	50	10-15	0.3	15.0
17.5	22	15-20	5.3	116.6
22.5	10	20-25	10.3	103.0
27.5	7	25-30	15.3	107.1
Total	N = 136	--------	----	$\sum f\|m - Z\|$ = 657.6

STANDARD DEVIATION

Definition : *Standard Deviation is the root mean square deviation of the values from their arithmetic mean.*

S.D. is the abbreviation and σ (read, sigma) is the symbol. Mean square deviation of the values from their A.M. is variance and is denoted by σ^2. S.D. is the positive square root of variance. Karl Pearson introduced the concept of standard deviation in 1893. S.D. is also called ***root mean square deviation.*** It is a mathematical deficiency of mean deviation to ignore negative sign. Standard deviation possesses most of the desirable properties of a good measure of dispersion. It is the most widely used absolute measure of dispersion. The corresponding relative measure is **Coefficient of Variation.** It is very popular and so extensively used as to raise a doubt whether there is any other relative measure of dispersion.

$$\text{Coefficient of Variation} = \frac{\text{Standard Deviation}}{\text{Arithmetic Mean}} \times 100$$

Example 17: 10 students of B.Com. class of a College have obtained the following marks in Statistics out of 100 marks.

Calculate the standard deviation.

S.No.	1	2	3	4	5	6	7	8	9	10
Marks	5	10	20	25	40	42	45	48	70	80

(B.Com. (C.A.) Bharathiar, A 01)

Solution:

S. No.	Marks X	X^2
1	5	25
2	10	100
3	20	400
4	25	625
5	40	1600
6	42	1764
7	45	2025
8	48	2304
9	70	4900
10	80	6400
Total	$\Sigma X=$ 385	$\Sigma X^2=$ 20143

Standard Deviation,

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

$$= \sqrt{\frac{20143}{10} - \left(\frac{385}{10}\right)^2}$$

$$= \sqrt{2014.3 - (38.5)^2}$$

$$= \sqrt{2014.30 - 1482.25}$$

$$= \sqrt{532.05}$$

$$= 23.07$$

Method 3 : Deviations taken from Assumed Mean.

This is same as the one followed in the calculation of arithmetic mean. But the formula is as follows:

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

$d = X - A$ is preferred when $X - \overline{X}$ are fractions.

Steps :

1. Form a table with the given values, X, in the first column.

2. Assume any value as A if it is not specified in a problem. It is preferable to assume a value in between the minimum value and the maximum value of X as A.

3. Find out the deviation of each value from the assumed mean A and call it d. i.e., find $d = X - A$ and write them in the next column.

3. Multiply each fX by the corresponding X to find fX^2 (It is not $(fX)^2$. That is, fX should not be squared) Write all such fX^2 values in the next column.

4. Find N (= Σf), ΣfX and ΣfX^2

5. Substitute in the above formula and simplify.

Example 21 : Calculate the standard deviation.

No.of Goals Scored in a Match	(X)	0	1	2	3	4	5
No.of Matches	(f)	1	2	4	3	0	2

Solution :

X	f	fX	fX^2
0	1	0	0
1	2	2	2
2	4	8	16
3	3	9	27
4	0	0	0
5	2	10	50
Total	N =12	ΣfX =29	ΣfX^2 =95

Standard Deviation,

$$\sigma = \sqrt{\frac{\Sigma fX^2}{N} - \left[\frac{\Sigma fX}{N}\right]^2}$$

$$= \sqrt{\frac{95}{12} - \left(\frac{29}{12}\right)^2}$$

$$= \sqrt{7.9167 - (2.4167)^2}$$

$$= \sqrt{7.9167 - 5.8404}$$

$$= \sqrt{2.0763}$$

$$= 1.44$$

Method 3 : Deviations taken from Assumed Mean. The formula is as follows:

$$\text{Standard Deviation, } \sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left[\frac{\Sigma fd}{N}\right]^2}$$

d = x – A A - Assumed Mean N = Σf

Steps:

1. Form a table with the given values, X and the frequencies, f in the first two columns.

2. Choose the value for A, assumed mean, if it is not specified.

3. Subtract A from each X and form the next column with d = X – A values.

4. Multiply each d by the corresponding f and enter all such products in the next column under the title fd.

5. Multiply each fd by the corresponding d and enter all such products in the next column under the title fd^2. (These are not the squares of fd values)

Combined S.D., $\sigma_{12} = \sqrt{\dfrac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1 + N_2}}$

$$= \sqrt{\frac{100 \times 2^2 + 200 \times (2.5)^2 + 100 \times (-0.67)^2 + 200 \times (0.33)^2}{100 + 200}}$$

$\because d_1 = \overline{X}_1 - \overline{X}_{12} = 7.00 - 7.67 = -0.67;$

$d_2 = \overline{X}_2 - \overline{X}_{12} = 8.00 - 7.67 = 0.33$

$$= \sqrt{\frac{400.00 + 1250.00 + 44.89 + 21.78}{300}}$$

$$= \sqrt{\frac{1716.67}{300}}$$

$$= \sqrt{5.7222}$$

$= 2.39$

Example 44: From the following price of gold in a week, find the city in which the price was more stable.

Day	Mon	Tues	Wed	Thu	Fri	Sat
City A	498	500	505	504	502	509
City B	500	505	502	498	496	505

Solution :

City A X_1	$X_1 - \overline{X}_1$ $\overline{X}_1 = 503$	$(X_1 - \overline{X}_1)^2$	City B X_2	$X_2 - \overline{X}_2$ $\overline{X}_2 = 501$	$(X_2 - \overline{X}_2)^2$
498	−5	25	500	−1	1
500	−3	9	505	4	16
505	2	4	502	1	1
504	1	1	498	−3	9
502	−1	1	496	−5	25
509	6	36	505	4	16
$\Sigma X_1 = 3018$	-	$\Sigma(X_1 - \overline{X}_1)^2 = 76$	$\Sigma X_2 = 3006$	-	$\Sigma(X_2 - \overline{X}_2)^2 = 68$

City A

$\bar{X}_1 = \frac{\Sigma X_1}{N_1}$

$= \frac{3018}{6}$

$=$ Rs. 503

$\sigma_1 = \sqrt{\frac{\Sigma(X_1 - \bar{X}_1)^2}{N_1}}$

$= \sqrt{\frac{76}{6}}$

$= \sqrt{12.6667}$

$=$ Rs.3.56

$C.V._1 = \frac{\sigma_1}{\bar{X}_1} \times 100$

$= \frac{3.56}{503} \times 100$

$= 0.71$

City B

$\bar{X}_2 = \frac{\Sigma X_2}{N_2}$

$= \frac{3006}{6}$

$=$ Rs. 501

$\sigma_2 = \sqrt{\frac{\Sigma(X_2 - \bar{X}_2)^2}{N_2}}$

$= \sqrt{\frac{68}{6}}$

$= \sqrt{11.3333}$

$=$ Rs. 3.37

$C.V._2 = \frac{\sigma_2}{\bar{X}_2} \times 100$

$= \frac{3.37}{501} \times 100$

$= 0.67$

Coefficient of Variation of price in City B is less. Hence, the price was more stable in City B.

Example 45: Goals scored by two teams A and B in a series of football matches were observed as follows:

No.of Goals Scored in a Match	No.of Matches	
	Team A	Team B
0	5	4
1	7	5
2	5	5
3	3	4
4	2	3
5	3	3

Which team, A or B, may be considered as a more consistent team?

Solution: Goals(X) are common. No.of matches(f) differ between the teams.

Goals	Matches		Team A		Team B	
	Team A	Team B				
X	f_1	f_2	f_1X	f_1X^2	f_2X	f_2X^2
0	5	4	0	0	0	0
1	7	5	7	7	5	5
2	5	5	10	20	10	20
3	3	4	9	27	12	36
4	2	3	8	32	12	48
5	3	3	15	75	15	75
Total	$N_1=$ 25	$N_2=$ 24	Σf_1X =49	Σf_1X^2 =161	Σf_2X =54	Σf_2X^2 =184

Team A

Mean, $\bar{X}_1 = \dfrac{\Sigma f_1X}{N_1} = \dfrac{49}{25} = 1.96$

S.D., $\sigma_2 = \sqrt{\dfrac{\Sigma f_1X^2}{N_1} - \left(\dfrac{\Sigma f_1X}{N_1}\right)^2}$

$= \sqrt{\dfrac{161}{25} - \left(\dfrac{49}{25}\right)^2}$

$= \sqrt{6.44 - (1.96)^2}$

$= \sqrt{6.4400 - 3.8416}$

$= \sqrt{2.5984}$

$= 1.61$

C.V.$_1$ $= \dfrac{\sigma_1}{\bar{X}_1} \times 100$

$= \dfrac{1.61}{1.96} \times 100$

$= 82.14$

Team B

Mean, $\bar{X}_2 = \dfrac{\Sigma f_2X}{N_2} = \dfrac{54}{24} = 2.25$

S.D., $\sigma_2 = \sqrt{\dfrac{\Sigma f_2X^2}{N_2} - \left(\dfrac{\Sigma f_2X}{N_2}\right)^2}$

$= \sqrt{\dfrac{184}{24} - \left(\dfrac{54}{24}\right)^2}$

$= \sqrt{7.6667 - (2.25)^2}$

$= \sqrt{7.6667 - 5.0625}$

$= \sqrt{2.6042}$

$= 1.61$

C.V.$_2$ $= \dfrac{\sigma_2}{\bar{X}_2} \times 100$

$= \dfrac{1.61}{2.25} \times 100$

$= 71.56$

Coefficient of variation of Team B is less. Hence, Team B is the more consistent team.