Preparation of a Frequency Distribution with Class Interval from Mid-values

Mid-value	Frequency	Class	Technique
5	6	0—10	$l_1 = 5 - \frac{10}{2} = 0,\ l_2 = 5 + \frac{10}{2} = 10$
15	5	10—20	$l_1 = 15 - \frac{10}{2} = 10,\ l_2 = 15 + \frac{10}{2} = 20$
25	11	20—30	$l_1 = 25 - \frac{10}{2} = 20,\ l_2 = 25 + \frac{10}{2} = 30$
35	9	30—40	$l_1 = 35 - \frac{10}{2} = 30,\ l_2 = 35 + \frac{10}{2} = 40$
45	8	40—50	$l_1 = 45 - \frac{10}{2} = 40,\ l_2 = 45 + \frac{10}{2} = 50$

(v) Cumulative Frequency Series: Cumulative frequency series is that series in which the frequencies are added corresponding to each class interval in the distribution. The frequencies then become cumulative frequencies. Cumulative frequency corresponding to the first 'class interval' would obviously be the same as the frequency itself. But for the second 'class interval' the cumulative frequency would be the sum total of the frequencies corresponding to both the second as well as first 'class intervals. The following table gives an example how a cumulative frequencies are obtained:

Cumulative Frequency Series

Marks	Frequency (No. of students)	Cumulative frequencies
5—10	3	3
10—15	8	8 + 3 = 11
15—20	9	11 + 9 = 20
20—25	4	20 + 4 = 24
25—30	4	24 + 4 = 28

Cumulative frequency series may be presented in two ways.

(*i*) Cumulative frequencies may be expressed on the basis of upper limits of the class intervals, *e.g.*, less than 10, less than 15, less than 20, when the class intervals are 5—10, 10—15 and 15—20.

(*ii*) Cumulative frequencies may be expressed on the basis of lower limits of the class intervals, *e.g.*, more than 5, more than 10, more than 15, when the class intervals are 5—10, 10—15 and 15—20.

Thus, when a frequency distribution is to be converted into a cumulative frequency distribution, the cumulative frequencies would correspond to either the lower class limits or the upper class limits of the class intervals in a distribution. Accordingly, the class intervals would get converted into 'less than' or 'more than' values. If it is of the less than type, it will represent the total frequency of all values less than and equal to the class value to which it is related. If it is a 'more than' type it will represent the total frequency of all values more than and equal to the class values to which it is related. Following is an example how a simple frequency distribution is converted into a cumulative frequency distribution.

Cumulative Frequency Series

Less than method		More than method	
Marks	**No. of students**	**Marks**	**No. of students**
Less than 10	3	More than 5	25 + 3 = 28
Less than 15	3 + 8 = 11	More than 10	17 + 8 = 25
Less than 20	11 + 9 = 20	More than 15	8 + 9 = 17
Less than 25	20 + 4 = 24	More than 20	4 + 4 = 8
Less than 30	24 + 4 = 28	More than 25	4

Conversion of Cumulative Frequency Series into Simple Frequency Series

Cumulative frequency series may be converted into simple frequency series. Following examples explain this process:

Example 5. Convert the following cumulative frequency distribution into a simple frequency distribution.

Marks	No. of students
Less than 10	4
Less than 20	20
Less than 30	40
Less than 40	48
Less than 50	50

Solution:

Conversion of a Cumulative Frequency Distribution into a Simple Frequency Distribution

Marks	No. of students
0—10	4
10—20	20 – 4 = 16
20—30	40 – 20 = 20
30—40	48 – 40 = 8
40—50	50 – 48 = 2

Example 6. Convert the following cumulative frequency distribution into a simple frequency distribution.

Marks	No. of students
More than 0	55
More than 5	51
More than 10	43
More than 15	28
More than 20	16
More than 25	6
More than 30	0

Solution:

Conversion of Cumulative Frequency Distribution into Simple Frequency Distribution

Marks	No. of students (f)
0—5	55 – 51 = 4
5—10	51 – 43 = 8
10—15	43 – 28 = 15
15—20	28 – 16 = 12
20—25	16 – 6 = 10
25—30	6 – 0 = 6
30—35	0

Construction of Frequency Distribution

The technique of constructing frequency distribution is illustrated as follows:

Example 7. Given below are the marks of 20 students of a class. Make a discrete frequency distribution.

10	12	18	14	13	10	12	15	17	19
18	16	14	15	17	11	20	13	12	14

Solution: **Construction of a Discrete Frequency Distribution**

Marks	Tally bars	No. of students (*f*)
10	||	2
11	|	1
12	|||	3
13	||	2
14	|||	3
15	||	2
16		1
17	||	2
18	||	2
19	|	1
20	|	1
	Total	20

Example 8. Following is the record of marks obtained by 75 students in an examination. Form a frequency distribution on exclusive basis:

84	19	58	44	87	58	43	40	73	43	56	55	40	91	35
18	59	27	92	13	45	61	39	78	23	11	71	62	22	41
63	47	39	19	22	35	30	80	37	80	52	73	65	50	43
40	27	84	53	19	35	72	44	19	51	67	58	76	38	16
37	74	45	50	53	70	36	33	63	67	85	45	55	41	49

Solution: Total number of students whose marks are given as 75. Lowest marks obtained by a student is 11 and highest marks obtained is 92. Range is 92 – 11, *i.e.*, 81. So if we take class interval of 10, 9 classes will be formed. So we take the lowest class interval as 10 to 20.

Frequency Distribution of Marks Obtained by 75 Students

Marks	Tally bars	Frequency
10—20	𝍸 ||	7
20—30	||||	5
30—40	𝍸 𝍸 |	11
40—50	𝍸 𝍸 𝍸	15
50—60	𝍸 𝍸 |||	13
60—70	𝍸 ||	7
70—80	𝍸 |||	8
80—90	𝍸 |	6
90—100	|||	3
	Total	75

Example 9. For the following raw data prepare a frequency distribution with a class interval of 5 on inclusive basis:

Marks in English									
12	36	40	16	10	10	19	20	28	30
19	27	15	21	33	45	7	19	20	26
26	37	6	5	20	30	37	17	21	20

Solution: The lowest value is 5 and highest is 45. We take class interval of 5 on inclusive basis. The various classes will be 5—9, 10—14 and so on upto 45—49.

Formation of a Frequency Distribution

Marks	Tally bars	No. of students (f)
5—9	|||	3
10—14	|||	3
15—19	~~||||~~ |	6
20—24	~~||||~~ |	6
25—29	||||	4
30—34	|||	3
35—39	|||	3
40—44	|	1
45—49	|	1
	Total	30

Example 10. The following are the marks obtained by 30 students of B.Com.II class in statistics:

Marks out of 50					
15	10	8	7	6	11
20	12	14	16	18	13
0	5	4	7	9	17
8	16	18	19	20	0
24	28	26	25	29	4

Prepare a frequency distribution by taking a class interval of 5 on exclusive basis.

Solution: The lowest value is 0 and highest is 29. We have to take a class interval of 5. The various classes will be 0—5, 5—10, and so on upto 25—30.

Frequency Distribution

Marks	Tally bars	Frequency
0—5	||||	4
5—10	~~||||~~ ||	7
10—15	~~||||~~	5
15—20	~~||||~~ ||	7
20—25	|||	3
25—30	||||	4
	Total	30

Example 11. The following is a record of weight of 70 students (in lbs). Tabulate the data in the form of frequency distribution taking the lowest class as (60—69).

61	73	93	107	112	76	78	69	96	72
80	88	96	109	103	84	84	106	91	75
91	92	102	91	101	90	77	105	90	86
113	101	114	72	77	118	95	63	99	82
100	106	87	89	92	107	111	76	83	86
106	107	62	94	73	108	115	85	98	93
109	97	74	98	67	82	104	88	88	92

Solution:

Preparation of Frequency Distribution

Weight in lbs	Tally bars	Frequency
60—69	𝍸	5
70—79	𝍸 𝍸 |	11
80—89	𝍸 𝍸 ||||	14
90—99	𝍸 𝍸 𝍸 |||	18
100—109	𝍸 𝍸 𝍸 |	16
110—119	𝍸 |	6
	Total	70

IMPORTANT TYPICAL EXAMPLES

Example 12. If the class mid-points in a frequency distribution of age of a group of persons are 25, 32, 39, 46, 53 and 60, find.

(*i*) the size of the class interval

(*ii*) the class boundaries

(*iii*) the class limits, assuming that the age quoted is the age completed last birthday.

Solution: (*i*) Size of class interval = Difference between the mid-value of any two consecutive classes.

$$= 32 - 25 = 39 - 32 = ... = 60 - 53 = 7$$

(*ii*) Since, the size of the class is 7 and the mid values of classes are 25, 32, 39, 46, 53 and 60, the corresponding class boundaries for different classes are obtained by using the formula:

$$l_1 = m - i/2 \qquad l_2 = m + i/2$$

where, m = mid-value, i = size of class.

The class boundaries for the first class will be:

$$l_1 = 25 - \frac{7}{2} = 25 - 3.5 = 21.5,\ l_2 = 25 + \frac{7}{2} = 25 + 3.5 = 28.5$$

The class boundaries for the 2nd class will be:

(32 – 3.5, 32 + 3.5), *i.e.*, (28.5, 35.5) and so on.

Thus, the various classes with class boundaries are given as

Classes	Mid-value	Technique
21.5—28.5	25	$l_1 = 25 - \frac{7}{2} = 21.5,\ l_2 = 25 + \frac{7}{2} = 28.5$
28.5—35.5	32	$l_1 = 32 - \frac{7}{2} = 28.5,\ l_2 = 32 + \frac{7}{2} = 35.5$
35.5—42.5	39	$l_1 = 39 - \frac{7}{2} = 35.5,\ l_2 = 39 + \frac{7}{2} = 42.5$
42.5—49.5	46	$l_1 = 46 - \frac{7}{2} = 42.5,\ l_2 = 46 + \frac{7}{2} = 49.5$
49.5—56.5	53	$l_1 = 53 - \frac{7}{2} = 49.5,\ l_2 = 53 + \frac{7}{2} = 56.5$
56.5—63.5	60	$l_1 = 60 - \frac{7}{2} = 56.5,\ l_2 = 60 + \frac{7}{2} = 63.5$

(*iii*) Assuming that the age quoted (X) is the age completed on last birthday, then X will be a discrete variable which takes only integral values. Hence, the given distribution can be expressed in an inclusive type of classes with class interval magnitude 7, as given in the adjoining table:

Age (on last birthday)	Mid-value
22—28	25
29—35	32
36—42	39
43—49	46
50—56	53
57—63	60

EXERCISE 3.1

1. Arrange the following data in an ascending order:
   18, 30, 15, 20, 10, 25, 19, 28
2. Following are the marks obtained by 25 students in statistics. Construct a discrete frequency distribution:

5	6	8	10	11	13	6	8	5	13	8	10	5
11	6	8	5	13	11	8	5	8	5	8	6	

3. From the following data relating to wages of 20 workers, prepare frequency distribution with a class interval of 5 on exclusive and inclusive basis:

10	15	25	27	29	20	24	23	22	12
14	16	17	18	19	18	16	15	5	9

[**Hint:** See Example 16]

4. From the following data related to the weight of college students in kg, prepare a frequency distribution with a class interval of 10 on exclusive and inclusive basis:

40	70	63	53	85
92	72	65	53	79
49	42	43	47	50
52	50	48	65	42
69	60	54	82	55

5. The weights in grams of 50 apples picked from a box are as follows:

110	103	89	75	98	121
110	108	93	128	185	123
113	92	86	70	126	78
139	120	129	119	105	120
100	116	85	99	114	189
205	111	141	136	123	90
115	128	160	78	90	107
81	137	25	84	104	100
87	115				

Construct a frequency distribution with class intervals of 15 gms on exclusive and inclusive basis.

6. Convert the following simple frequency distribution into cumulative frequency distribution by using 'less than method' and 'more than method'.

Marks	No. of students
0—10	7
10—20	10
20—30	23
30—40	30
40—50	3

7. Convert the following into exclusive form:

Class	Frequency
15—19	2
20—24	7
25—29	20
30—34	11
35—39	5

8. Prepare a frequency distribution from the following information:

7 students get less than 10 marks
18 students get less than 20 marks
38 students get less than 30 marks
63 students get less than 40 marks
70 students get less than 50 marks

9. Construct the simple frequency distribution from the following data:

Mid-values	Frequency
5	2
15	8
25	15
35	12
45	7
55	6

10. Prepare a simple frequency distribution from the following data:

Marks	No. of students
More than 0	21
More than 10	19
More than 20	14
More than 30	7
More than 40	2

MISCELLANEOUS SOLVED EXAMPLES

Example 13. From the following data related to the weight of 25 college students in kg, construct a grouped frequency distribution with class interval 40-49 and so on.

40	49	69	72	50	63	43	54	43	65	85	50	55
42	52	70	42	60	65	48	53	47	82	79	42	

Solution:

Weights (in kg)	Tally bars	No. of students
40-49	~~IIII~~ IIII	9
50-59	~~IIII~~ I	6
60-69	~~IIII~~	5
70-79	III	3
80-89	II	2
	Total	25

xample 14. Using **Sturges' rule** $k = 1 + 3.322 \log N$, prepare group frequency distribution from the marks obtained by 50 students:

4	47	84	65	15	44	13	42	60	20
12	50	92	54	17	55	15	25	70	24
25	48	17	72	80	67	20	30	72	35
45	70	22	14	70	70	25	12	75	42
65	72	34	20	72	12	30	15	80	45

ution: Here largest item is 92 and the smallest item is 4 and number of items are 50.

As per **Sturges' rule**

Number of classes $k = 1 + 3322 \log N = 1 + 3 \cdot 322 \log 50$

$$= 1 + 3.322\ (1.6990)$$

$$= 1 + 5.6 = 6.6$$

$$\text{Range} = \text{Largest} - \text{Smallest}$$

$$= 92 - 4 = 88$$

$$\text{Width of the classes} = \frac{Range}{No.\, of\ Class} = \frac{88}{6.6} = 13.3$$

In multiple of 5 width of the class intervals would be 15, so the starting class would be 0–15 and so on.

Marks	Tally bars	No. of students
0—15	~~IIII~~ I	6
15—30	~~IIII~~ ~~IIII~~ III	13
30—45	~~IIII~~ II	7
45—60	~~IIII~~ II	7
60—75	~~IIII~~ ~~IIII~~ II	12
75—90	IIII	4
90—105	I	1
	Total	50

Example 15. Classify the following data by taking class interval such that their mid-values are 17, 22, 27, 32 and so on:

30	42	30	54	40	48	14	17	51	42	25	41
30	27	42	36	28	28	37	54	44	31	36	40
36	22	30	31	19	48	16	42	32	21	22	40
33	41	21	16	17	36	37	41	46	47	52	53

Solution: If the mid-value of the first class is 17 and the subsequent mid-values are 22, 27, then the first class should be 15—19 as $\frac{15+19}{2}=17$, and the second class would be 20—24 as $\frac{20+24}{2}=22$ as so on. Therefore, the classified data is as shown below:

Class	Tally bars	Frequency
15—19	~~IIII~~ I	6
20—24	IIII	4
25—29	IIII	4
30—34	~~IIII~~ III	8
35—39	~~IIII~~ I	6
40—44	~~IIII~~ ~~IIII~~ I	11
45—49	~~IIII~~	4
50—54	~~IIII~~	5
	Total	48

Example 16. From the following data relating to wages of 20 workers, prepare frequency distribution with a class interval of 5 on exclusive and inclusive basis:

10	15	25	27	29	20	24	23	22	12
14	16	17	18	19	18	16	15	5	9

Solution: The lowest value is 5 and highest is 29. We have to take a class interval of 5. The classes will be (*i*) 5—10, 10—15, 15—20 ... 25—30 on exclusive basis and (*ii*) 5—9, 10—14, 15—19, ... 25—29 on inclusive basis:

(*i*) Formation of a Frequency Distribution on Exclusive Basis:

Wages	Tally bars	Frequency
5—10	𝍷𝍷	2
10—15	𝍷𝍷𝍷	3
15—20	𝍸 𝍷𝍷𝍷	8
20—25	𝍷𝍷𝍷𝍷	4
25—30	𝍷𝍷𝍷	3
	Total	20

(*ii*) Formation of a Frequency Distribution on Inclusive Basis:

Wages	Tally bars	Frequency
5—9	𝍷𝍷	2
10—14	𝍷𝍷𝍷	3
15—19	𝍸 𝍷𝍷𝍷	8
20—24	𝍷𝍷𝍷𝍷	4
25—29	𝍷𝍷𝍷	3
	Total	20

QUESTIONS

1. What is meant by classification of data? What are its various objectives? Also discuss various methods of classification.
2. What is frequency distribution? What are the problems in its construction?
3. What are the general rules of framing a frequency distribution with special reference to the choice of class interval and number of classes?
4. Explain the advantages of classification of data. Discuss the different methods of classification.
5. Explain giving examples the inclusive and exclusive form of class intervals.

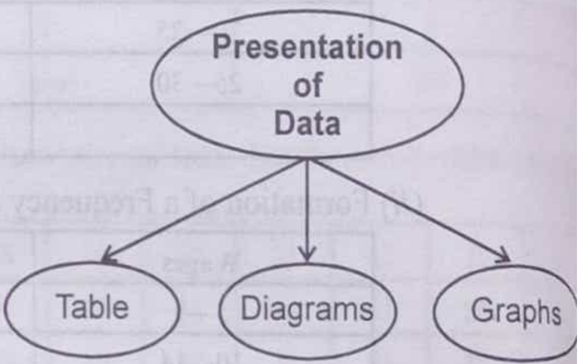
4 Presentation of Data

INTRODUCTION

The collection and classification of data lead to the problems of presentation of data. The presentation of data means exhibition of the data in such a clear and attractive manner that these are easily understood and conclusions are drawn thereof. There are many methods of presenting the data of which the following three are generally use:

(1) Tables, (2) Diagrams, and (3) Graphs.

Let us study them briefly.



TABULAR PRESENTATION OF DATA

After classification of the data, the same is presented in the form of tables. The main objective o the tables is to present the data in such a way as to attract the attention of an individual to importar information. A table is that method of presentation of data which symmetrically organises the dat rows and columns. In the words of **Neiswanger**, "*A statistical table is a systematic organisation data in columns and rows.*" Vertical disections in the formation of a table are known as column an horizontal disection (=) are known as rows. Tabulation is a process of presenting data in the form c table. According to **Prof. L.R. Connor**, "*Tabulation involves the orderely and systemat presentation of numerical data in a form designed to elucidate the problem under consideration.*" I the words of **Prof. Blair**, "*Tabulation in its broadest sense is any orderly arrangement of data columns and rows.*"

OBJECTIVES OF TABULATION

(1) Simple: The principal objective of tabulation is to organise the data in such a manner th these become simple to understand.

(2) Brief: The tabulation presents a large volume of statistical data in a very brief form.

(3) Facilities Comparison: The tabulation facilitates comparision of data by presenting tl data in different classes.

(4) Helpful in Presentation: Tabulation makes the data very brief. Tables are, therefore, ve useful in graphic or diagrammatic presentation of the data.

(5) Helpful in Analysis: It is very easy to analyse the data from Tables. It is by organising t data in the form of table that one finds out mean and dispersion.

(6) Clarifies the Chief Characteristis of Data: The tabulation highlights the characteristics of data. Accordingly, it becomes easy to remember the statistical facts.

(7) Economy: Tabular presentation is a very economical mode of data presentation. It saves time as well as space. Important figures can be easily located in a Table.

DIFFERENCE BETWEEN CLASSIFICATION AND TABULATION

The main difference between classification and tabulation are as follows:

(1) Classification and tabulation have to be done in the sequence. First data are classified and then they are presented in tables.

(2) Classification forms the basis of tabulation.

(3) In classification, data are classified into different classes according to their similarities and dissimilarities. On the other hand, in tabulation, the classified data are placed in rows and columns. Thus, tabulation is a mechanical function of classification.

(4) Classification is a process of statistical analysis whereas tabulation is a process of presentation.

(5) Classification divides the data into classes and sub-classes, while tabulation presents the data under headings and sub-headings.

MAIN PARTS OR COMPONENTS OF A STATISTICAL TABLE

The main parts of a table are as follows:

(1) Table Number: First of all, a table must be numbered. Different tables must have different numbers, *e.g.,* 1, 2, 3..., etc. These numbers must be in the same order as the tables. Numbers facilitate location of the tables.

(2) Title of the Table: A table must have a title. Title must be written in bold letters. It should attract the attention of the readers. The title must be simple, clear and short. A good title must reveal: (*i*) the problem in hand, (*ii*) the time period of the study, (*iii*) the place of study; and (*iv*) the nature of classification of data. A good title is short but complete in all respects.

(3) Head Note: If the title of the table does not give complete information, it is supplemented with a head note. Head note completes the information in the title of the table. Thus, units of the data are generally expressed in the form of a head note below the title of the table.

(4) Stubs: Stubs are titles of the rows of a table. These titles indicate information contained in the rows of the table.

(5) Caption: Caption means title given to the columns of a table. A caption indicates information contained in the columns of the table. A caption may have sub-heads when information contained in the columns is divided in more than one class. For example, a caption of 'Students' may have boys and girls as sub-heads.

(6) Body or Field: Body of a table means sum total of the items in the table. Thus, body is the most important part of a table. It shows the whole of information contained in the table. Each item in the body is called 'cell'.

(7) Footnotes: Footnotes are given for clarification of the reader. These are generally given when information in the table is not self-explanatory.

(8) Source: When tables are based on secondary data, source of the data is to be given. Source of the data is specified below the footnote. It should give: name of the publication and publisher, year of publication, reference, page numbers, etc.

Format of a Table
Table Number,
Title,
Head Note,
Columns

Stub Heads	Caption	Caption	Total
Stub-Entries	Cell	Cell	Cell
....	B O	D	Y
....			
Total	....	....	....

→ Rows

→ Stubs

Footnote:

Source:

Example 1. In 2001-2002, total production of foodgrains was 1,928 lakh tons of which production of rice, wheat and other crops was, 860, 708 and 360 lakh tons, respectively. Percentage share of rice, wheat and other crops in the total production of foodgrains was 44.60, 36.72 and 18.68 respectively. Present this information in the form of a Table indicating its various parts.

Solution: **Table 1: Title: Production of Foodgrains in India in 2001-2002**

S. No.	Foodgrains	Production	
		Total Quantity	Percentage of Total Output
1.	Rice	860	44.60
2.	Wheat	708	36.72
3.	Other crops	360	18.68
	Total	1,928	100.00

→ Stub Head ← Column

Stubs Head

← Body

Footnote: In 'others', all remaining foodgrains are included.

Source: *Economic Survey, 2002.*

GENERAL RULES FOR THE CONSTRUCTION OF A TABLE OR ESSENTIALS OF A GOOD TABLE

Construction of a good table depends upon the objective of study. It also depends upon the intelligence of the statistician. There are no hard and fast rules for the construction of a table. However, certain guidelines may be noted in this context:

(1) Title: Title of a table must be provided at the top centre of the table.

(2) Special Emphasis: Some items in the table may need special emphasis. Such items should be placed in the head rows (top above) or head columns (extreme left). Moreover, such items should be presented in bold letters.

(3) Size of Table: To determine an ideal size of a Table, a rough draft or sketch must be drawn. Rough draft will give an idea as to how many rows and columns should be drawn for presentation of the data.

(4) Headings: Headings should generally be written in the singular form. For example, in the columns indicating goods, the word 'good' should be used.

(5) Abbreviations should not be Used: Use of abbreviations should be avoided in the headings or sub-headings of the table. Short forms of the words such as Govt., M.P. (monetary policy), etc. should not be used. Also such signs as "(ditto)" should not be used in the body of the table.

(6) Footnote: Footnote should be given only if needed. However, if footnote is to be given, it must bear some asterisk mark (*) corresponding to the concerned item.

(7) Units: Units used must be specified above the columns. If figures are very large, units may be noted in the short form as '000' hectare or '000' tons.

(8) Total: In the table, sub-totals of the items must be given at the end of each row. Grand total of the items must also be noted.

(9) Percentage and Ratio: Percentage figures should be provided in the table, if possible. This makes the data more meaningful.

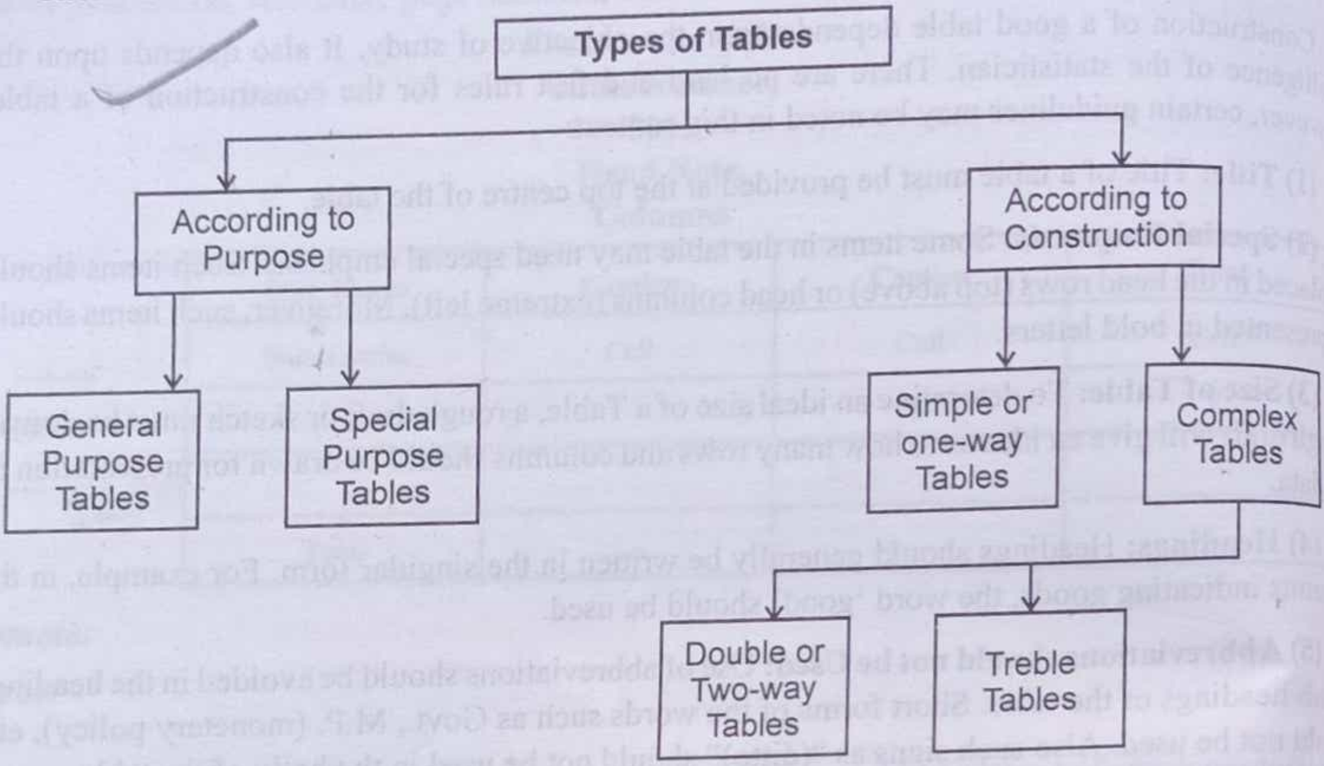
(10) Place of Approximation: If some approximate figures have been used in this table, the extent of approximation must be noted. This may be indicated at the top of the table as a part of head note or at the foot of the table as a footnote.

(11) Source of Data: Source of data must be noted at the foot of the table. It is generally noted next to the footnote.

(12) Simplicity, Economy and Attractiveness: A table must be simple, attractive and economical in space.

TYPES OF TABLES

There are three basis of classifying different types of tables:



(1) According to Purpose

According to purpose, there are two kinds of tables:

(*i*) **General Purpose Table:** General purpose table is that table which is of general use. It does not serve any specific purpose. Such tables are just 'data bank' for the use of researchers for their various studies. These tables are generally attached to some official reports, *e.g.*, Census Reports of India.

(*ii*) **Special Purpose Table:** Special purpose table is that table which is prepared with some specific purpose. Generally these are small tables and limited to the problem in hand. In these tables, data are presented in the form of result of the analysis.

(2) According to Construction

According to construction, also, tables are of two kinds:

(*i*) **Simple Table:** A simple table is that which shows only one characteristics of the data. Table 2, for example, is simple table. It shows number of students in a college:

Table 2: No. of Students in a College

Class	No. of students
BBA (I)	200
B.A. (I)	100
B.A. (II)	80
B.A. (III)	60
Total	440

(*ii*) **Complex Table:** A complex or multiple table is one which shows more than one characteristics of the data. On the basis of the characteristics shown, these tables may be further classified as:

(*a*) **Two-way Table:** A two-way table is that which shows two characteristics of the data. For example, table 3, showing the number of students in different classes according to their sex is a two-way table:

Table 3: No. of Students in a College (According to Sex and Class)

Class	No. of students		Total
	Boys	Girls	
BBA (I)	160	40	200
B.A. (I)	40	60	100
B.A. (II)	60	20	80
B.A. (III)	50	10	60
Total	310	130	440

(*b*) **Three-way Table:** A three-way table is that which shows three characteristics of the data. For example, table 4 shows number of students in a college according to class, sex and habitation.

Table 4: No. of Students in a College (According to Class, Sex and Habitation)

Class	Boys			Girls			Total		
	Rural	Urban	Total	Rural	Urban	Total	Rural	Urban	Total
BBA (I)	10	10	20	5	5	10	15	15	30
B.Com (I)	10	30	40	15	45	60	25	75	100
B.Com (II)	15	45	60	5	15	20	20	60	80
B.Com (II)	10	40	50	5	5	10	15	45	60
Total	45	125	170	30	70	100	75	195	270

Similarly, a manifold table showing more than three characteristics can be constructed. Such a type of table gives information about so many related questions. These tables are also called Higher Order Tables.

Example 2. Prepare a blank table to show the percentage of rural and urban population in India in 1971, 1981, 1991 and 2001.

Solution: **Percentage Distribution of Rural and Urban Population of India**

Census Year	Percentage of population		Total
	Rural	Urban	
1971			
1981			
1991			
2001			

Example 3. Point out the mistakes in the following table and rearrange it in the form of a good table.

Literate	Less than 20	20-30	30-40	40 and above
Male				
Female				

Solution: The following mistakes may be noted in the above table:

1. There is no table number and no title of the table.
2. The table is without any head note.
3. Caption and sub-entries have not been properly noted. Literacy and age be divided according to caption and sub-entries respectively.
4. Total of the rows and columns has not been provided.

Removing these mistakes, the Table may be presented in the following form:

Distribution of Population
(According to Age, Sex and Literacy)

Age (Years)	Boys			Girls		
	Male	Female	Total	Male	Female	Total
Less than 20 years						
20-30 years						
30-40 years						
40 and above						
Total						

Example 4. In 1995-96, total production of foodgrains was 1,720 lakh tons of which production of rice, wheat and other crops was, 795, 625 and 300 lakh tons, respectively. Percentage share of rice, wheat and other crops in the total production of foodgrains was 46.22, 36.34 and 17.44 respectively. Present this information in the form of a table.

Solution: **Production of Foodgrains in India in 1995-96** (lakh tonnes)

S.No.	Foodgrains	Production	
		Total Quantity	Percentage of Total Output
1.	Rice	795	46.22
2.	Wheat	625	36.34
3.	Others	300	17.44
	Total	1,720	100.00

Footnote: In 'others', all remaining foodgrains are included.
Source: *Economic Survey, 1997, P. S-16.*

Example 5. In a sample study about the coffee habits in two town, following data were observed:

Town X: 52% persons were males
25% were coffee drinkers, and
16% were male coffee drinkers

Town Y: 55% persons were males
28% were coffee drinkers, and
18% were male coffee drinkers.

Represent the above data in a tabular form.

Solution: **Percentage of Coffee Drinkers in Town X and Town Y** (in percentage)

Attribute	Town X			Town Y		
	Males	Females	Total	Males	Females	Total
Coffee Drinkers	16	9	25	18	10	28
Non-coffee Drinkers	36	39	75	37	35	72
Total	52	48	100	55	45	100

Example 6. In a trip organised by a college there were 80 persons each of whom paid Rs. 15.50 on an average. There were 60 students each of whom paid Rs. 16. Members of the teaching staff were charged at a higher rate. The number of servants was 6 (all males) and they were not charged anything. The number of ladies was 20% of the total of which one was a lady staff member.

Solution: No. of ladies $= \frac{20 \times 80}{100} = 16$

Trip Goers and their Contribution

Trip-Goers	Sex			Contribution Per Member	Total Contribution
	Male	Femlae	Total		
Students	45	15	60	Rs. 16	Rs. 960
Staff	13	1	14	Rs. 20	Rs. 280
Servants	6	Nil	6	Nil	Nil
Total	64	16	80	Rs. 15.50	Rs. 1,240

EXERCISE 4.1

1. In 1998, the contribution of agriculture, to India's National Income, Industry and services was 48%, 21% and 31% respectively. In 1999 these shares were 32%, 27% and 41% respectively. This information is based on Economic Survey of 2000-2001. Present this information in the form of a table.

2. Point out the mistakes in the following table. Re-arrange it in the form of a good table.

No. of Students	Subject			
	Economics	English	Hindi	History
Boys:				
Girls:				

3. Following information relates to the marks secured by 50 boys and girls in their paper in Economics. Present the information in the form of a two way table.

Marks	0—10	10—20	20—30	30—40
Boys:	10	7	6	1
Girls:	5	5	12	4

4. Draw a blank table to show the distribution of population according to sex, age and literacy.
5. Prepare a blank table in which can be shown the prices per kilograms of rice and wheat for the years 1994 and 1995 for three important markets in Punjab.
6. 'Point out the mistakes made in the following table and re-arrange the following table with a view to make it more intelligible.

Sex	Hindus		Muslims		Sikhs	
	Literate	Illiterate	Literate	Illiterate	Literate	Illiterate
Male:						
Female:						

7. Prepare a blank table showing the particulars of students studying in B.Com classes of Himachal Pradesh University College according to sex and class.
8. Present the following information in a suitable tabular form:
   (*i*) In 1985, out of total 2000 workers in a factory, 1550 were members of a trade union. The number of women workers employees was 250, out of which 200 did not belong to any trade union.
   (*ii*) In 1990, the number of union worker was 1725 of which 1600 were men. The number of non-union workers was 380, among which 155 were women.

DIAGRAMMATIC PRESENTATION OF DATA

Data may be presented in a simple and attractive manner in the form of diagrams. Diagrammatic presentation is the technique of presenting data in the form of Bar diagrams, Rectangles, Pie-diagrams, Pictographs and Cartographs.

Utility of Diagrammatic Representation

(1) Make Data Simple and Understandable: The most complex statistical data is made simple with the help of diagrams. One can understand the features of data merely by having a look at the diagrams.

(2) Remembrance for Long Period: In the form of diagrams, data are easily remembered for a long period. These are not easily forgotten.

(3) No Need of Training or Special Knowledge: One needs no training or special knowledge in reading the diagrams. Diagrams are easily understood even by a layman.

(4) Attractive and Effective Means of Presentation: Diagrams are very attractive and effective means of presenting data. It is rightly said that, a picture is worth of a thousand words.

(5) Saving of Time and Labour: Diagrammatic presentation of data saves lot of time and labour compared to other techniques of data presentation.

(6) Facilities Comparison: Diagrams facilitate comparison of data. Thus, data on investment in Private and Public sectors, when presented in the form of diagrams, can be easily compared. One can easily note the difference between the two.

(7) Informative and Entertaining: Besides being informative, diagrammatic presentation is an entertaining means of data presentation.

(8) Helpful in Predictions: Diagrammatic presentation helps in predicting the behaviour of variables.

(9) Helpful in Transmission of Information: Diagrams are very helpful in transmission of statistical information.

General rules for Constructing Diagrams

Some of the general rules for constructing diagrams are as follows:

(1) Attractive and Effective: Diagrams must be attractive and effective in communicating the information.

(2) Proper Size: Diagrams must suit the size of the paper. It should be neither too big nor too small.

(3) Proper Heading: Diagrams must bear proper headings. A heading must be simple, short and informative.

(4) Proper Scale: Before making a diagram, its scale should be properly determined and the same be mentioned on the diagram.

(5) Use of Signs and Colour: Diagrams must carry some signs on the nature and classification of information. Colours may be used to indicate different aspects of a diagram. These signs and colours must be clarified.

(6) Less Use of Words or Numerical: In diagrammatic presentation of data one should make use of minimum number of words and numericals.

(7) Drawing the Border: Diagrams must be bordered with bold lines to make them attractive.

(8) Simple: Simplicity is the principal feature of a diagram. It should not be ignored.

(9) From Left to Right or Bottom to Top: The construction of diagrams should flow from left to right or from bottom to the top.

(10) Statistics: One should indicate the statistics or data used in the construction of diagrams.

Limitations of Diagrammatic Presentation

The following are the main limitations of diagrammatic presentation of data:

(1) Estimate: Diagrammatic presentation of data shows only an estimate of the actual behaviour of the variables. In other words, diagrams show only an aggregate behaviour of the variables.

(2) Limited Use: Only a limited set of data can be presented in the form of a diagram. In fact, diagrams are generally used only when comparisons are involved.

(3) More Time: Diagrammatic presentation of data is a time consuming process. It involves too much verification of the data.

(4) Misuse: Diagrams may be misused for false projection of the statistical facts, especially in case of advertisement.

(5) Analysis: It is not very easy to arrive at final conclusions after seeing the diagrams. Generally, a diagram offers preliminary conclusions.

TYPES OF DIAGRAMS

Several types of diagrams are used to present statistic data. The following main types are discussed below:

(1) Bar Diagrams

(2) Rectangular Diagrams

(3) Pie Diagrams

(4) Pictograms and Cartograms.

(1) Bar Diagrams

Bar diagrams are the most common type of diagrams used in practice. In these diagrams, only the length of the bars are taken into account. The width of the bar is adjusted in accordance with the space available and the number of the bars to be used. The gap between one bar and another is also kept constant. In the construction of bar diagrams, either vertical or horizontal bars are used. But vertically bars are generally preferred.

Types of Bar Diagrams

Bar diagrams are of the following types:

(*i*) Simple bar diagrams

(*ii*) Multiple bar diagrams

(*iii*) Sub-divided bar diagrams

(*iv*) Percentage bar diagrams

(*v*) Deviation bar diagrams.

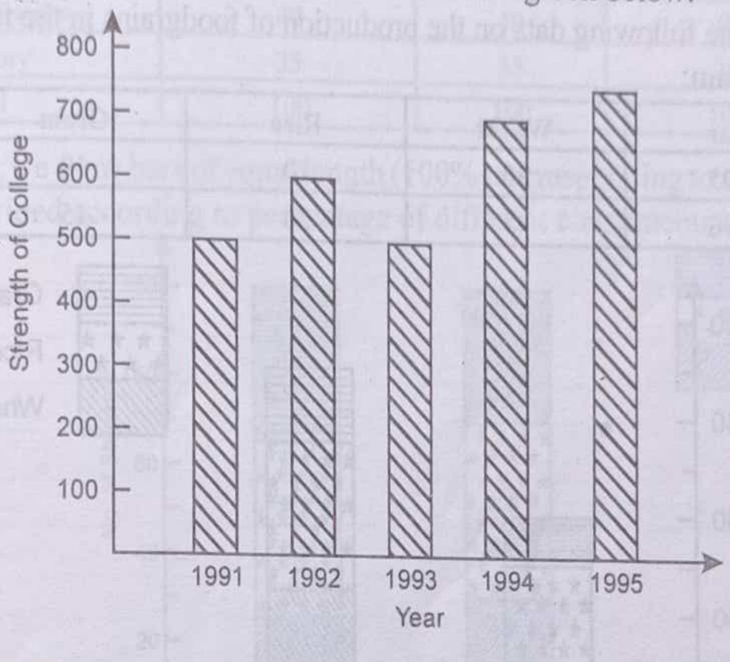
(*i*) Simple bar diagrams: A simple bar diagram is used to represent the values of a single variable with respect to time or geographical location etc. For example, the data of sales, production, population, etc., for various years may be shown by means of simple bar diagrams. Different values of the variable are shown by bars of proportionate lengths. The bars may be shaded or coloured to make the diagram more attractive.

Example 7. The strength of a college from 1991 to 1995 are given below:

Year:	1991	1992	1993	1994	1995
Strength of College:	500	600	500	700	750

Represent the data by a simple bar diagram.

Solution: The simple bar diagram of the above data is given below:

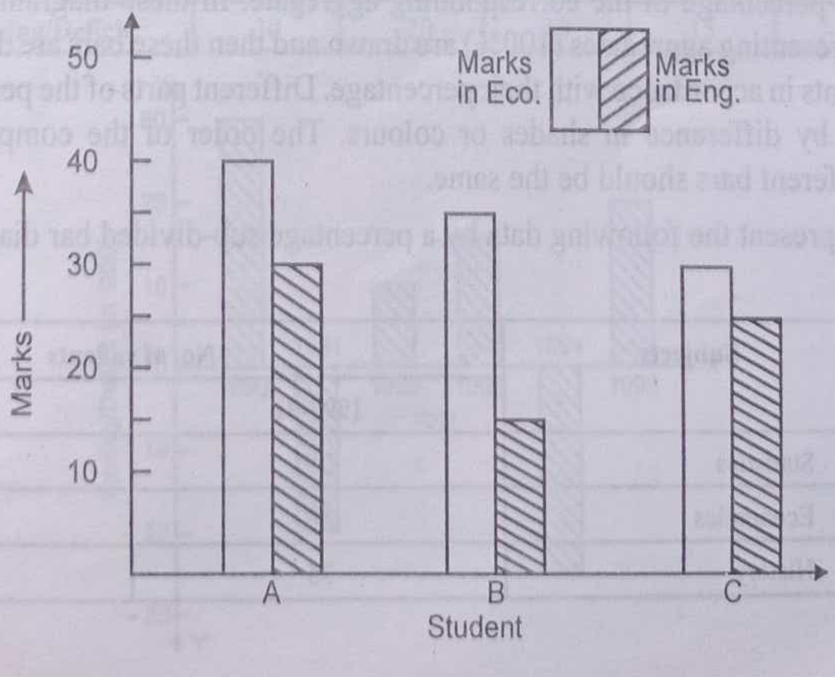


(ii) **Multiple bar diagrams:** The multiple bar diagrams are used to show two or more related variables with respect to time and location. In other words, they represent one or more than one type of data at a time. For every set of values, the separate bars are drawn but they are kept adjacent to each other. These bars are shaded or coloured differently to distinguish between variables.

Example 8. Draw a multiple bar diagram to show the following data:

Student:	A	B	C
Marks in Economics:	40	35	30
Marks in English:	30	15	25

Solution:

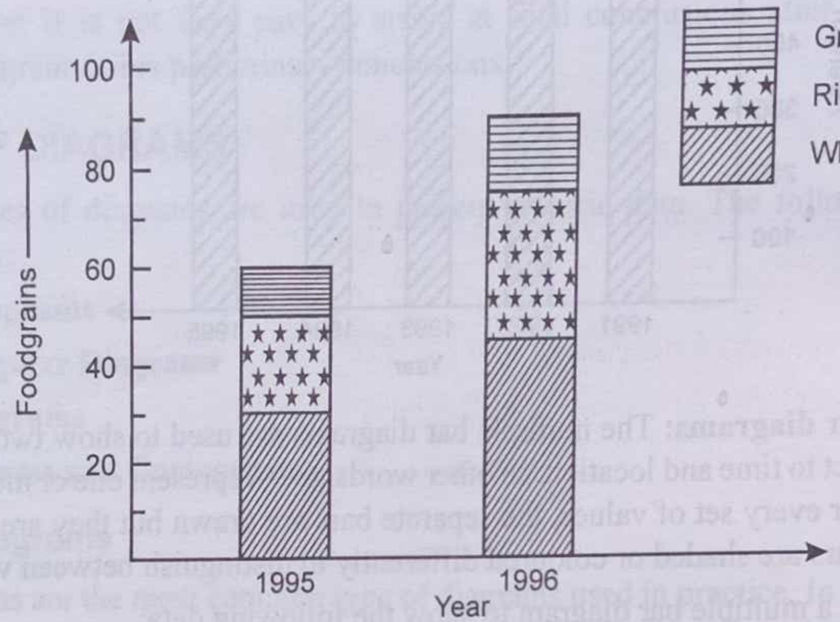


(iii) Sub-divided bar diagrams (or Component bar diagrams): Sub-divided bar diagram are used to present the constituent (or component) parts of the aggregates with respect to time location. In drawing such diagrams, first bars are drawn for the total sum of the values of differe variables and then these bars are divided to show the various components of the variables. Differe parts of the bar must be distinguished by difference in shades or colours. The order of th components of the various aggregates on different bars should be the same.

Example 9. Present the following data on the production of foodgrains in the form of a sub-divide bar diagram:

Year	Wheat	Rice	Gram	Total
1995	30	20	10	60
1996	45	30	15	90

Solution:



(iv) Sub-divided bar diagrams drawn on percentage basis: The sub-divided bar diagrams drawn on percentage basis are those diagrams which are used to present parts of the value of the aggregate with respect to some characteristic on percentage basis. Here, the values of the components are expressed on percentage of the corresponding aggregate. In these diagrams, first the bars of equal lengths representing aggregates (100%) are drawn and then these bars are divided to show the various components in accordance with their percentage. Different parts of the percentage bars must be distinguished by difference in shades or colours. The order of the components of various aggregates on different bars should be the same.

Example 10. Represent the following data by a percentage sub-divided bar diagram:

(in '000)

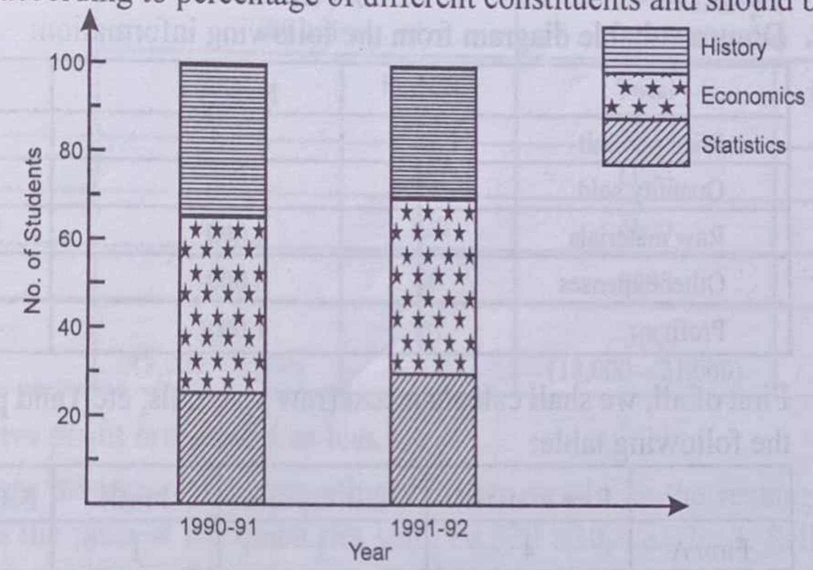
Subjects	No. of sudents	
	1990-91	1991-92
Statistics	25	30
Economics	40	42
History	35	28

Solution: First, we prepare a percentage table.

Percentage Table

Subject	1990-91		1991-92	
	No. of students ('000)	%	No. of students ('000)	%
Statistics	25	25	30	30
Economics	40	40	42	42
History	35	35	28	28
Total	100	100	100	100

Now, we draw bars of equal length (100%) corresponding to each year. Then each bar is divided according to percentage of different constituents and should be different.

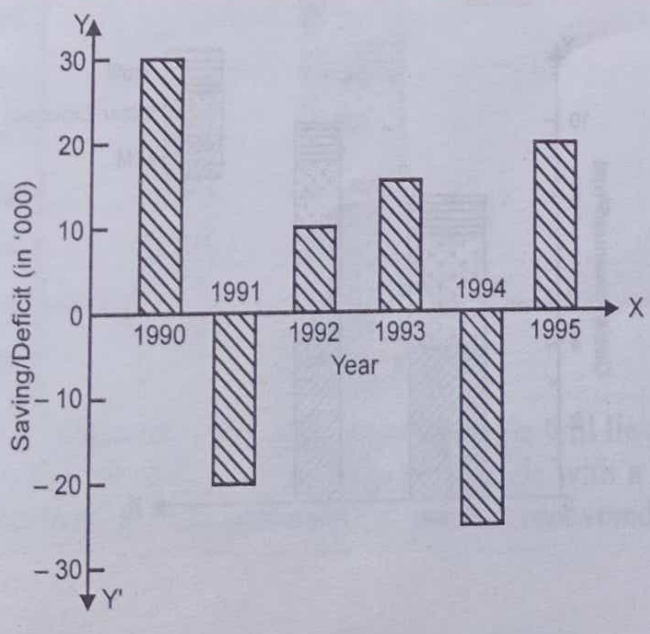


(*v*) **Deviation bar diagrams:** The deviation bar diagrams are used to compare the net deviation of related variables with respect to time and location. Bars representing positive and negative deviations are drawn above and below the base line. Such type of diagrams represent the deviations in magnitude as well as in direction.

Example 11. Represent the following data by a deviation bar diagrams:

Years:	1990	1991	1992	1993	1994	1995
Saving/Deficit:	30	–20	10	15	–25	20

Solution:



(2) Rectangular Diagrams

In case of bar diagrams, only length of the bar is taken into account as the comparison is on dimensional. On the other hand, in case of rectangular diagrams, both the lengths and breadth ar taken into account and data are represented by the area. Here, comparison is two dimensional. These diagrams may be constructed in either of the two ways: (*i*) By representing the figures as the are given, and (*ii*) By converting the figures into percentages and then sub-dividing the length int various components.

The following examples would illustrate both of the methods of constructing rectangula diagrams:

Sub-divided Rectangular Diagrams

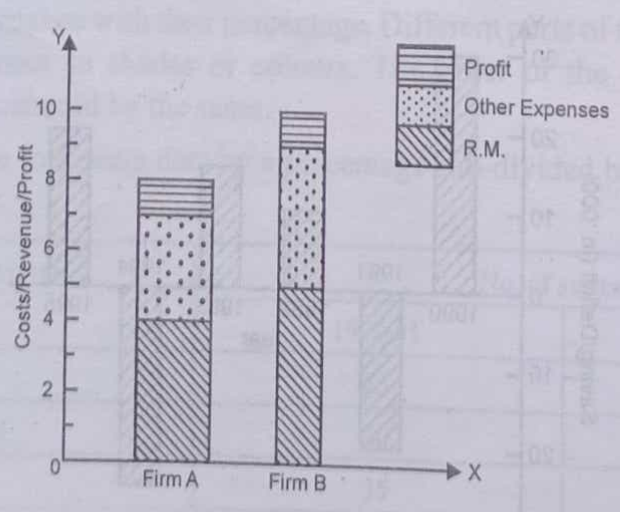
Example 12. Draw a suitable diagram from the following information:

	Firm A	Firm B
Price per unit	Rs. 8	Rs. 10
Quantity sold	1,000	600
Raw materials	4,000	3,000
Other expenses	3,000	2,400
Profit	1,000	600

Solution: First of all, we shall calculate cost (raw materials, etc.) and profit per unit as given in the following table:

	Raw materials	Other expenses	Profit	Selling price	Quantity sold
Firm A	4	3	1	8	1000
Firm B	5	4	1	10	600

An appropriate diagram for representing the data would be the rectangular diagram whose widths are in the ratio of quantities sold, *i.e.*, 1,000 : 600, *i.e.*, 10 : 6. Selling price would be represented by the corresponding heights (or lengths) of the rectangles with various cost (raw materials, others, etc.) and profit represented by various divisions of the rectangules as shown in the following diagram:



Example 13. Draw a suitable diagram to represent the following information:

	Selling price per unit (in Rs.)	Quantity sold	Total Cost (in Rs.)			
			Wages	Materials	Misc.	Total
Factory X	400	20	3,200	2,400	1,600	7,200
Factory Y	600	30	6,000	6,000	9,000	21,000

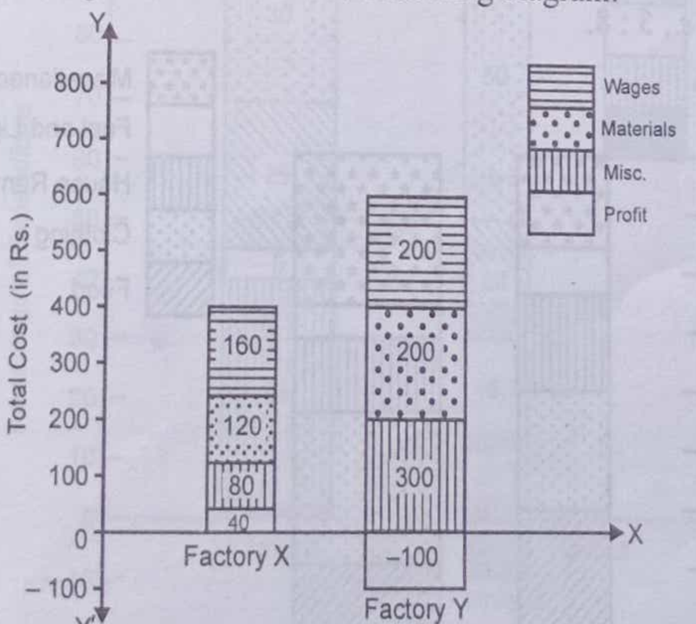
Show also the profit or loss as the case may be.

Solution: First of all we shall calculate the cost (wages, materials, misc.) and profit per unit as given in the following table:

Total Cost	Factory X (20 units)		Factory Y (30 units)	
	Total (Rs.)	Per Unit (Rs.)	Total (Rs.)	Per Unit (Rs.)
Wages	3200	160	6000	200
Materials	2400	120	6000	200
Misc.	1600	80	9000	300
Profit/Loss	800 (8,000 – 7,200)	40	–3000 (18,000 – 21,000)	–100

Note: Negative profit is regarded as loss.

An appropriate diagram for representing this data would be the rectangles whose widths are in the ratio of the quantities sold, *i.e.*, 20 : 30, *i.e.*, 2 : 3. Selling prices would be represented by the corresponding heights of the rectangles with various costs (wages, materials, misc.) and profit or loss represented by the various divisions of the rectangles as shown in the following diagram:



Note: In case of profit, *i.e.*, when SP > CP, the entire rectangle will lie above the X-axis. But in case of loss, *i.e.*, when SP < CP, we will have rectangle with a portion lying below the X-axis which will reflect the loss incurred, *i.e.*, can not recovered through sales.

▶ Percentage Sub-divided Rectangular Diagrams

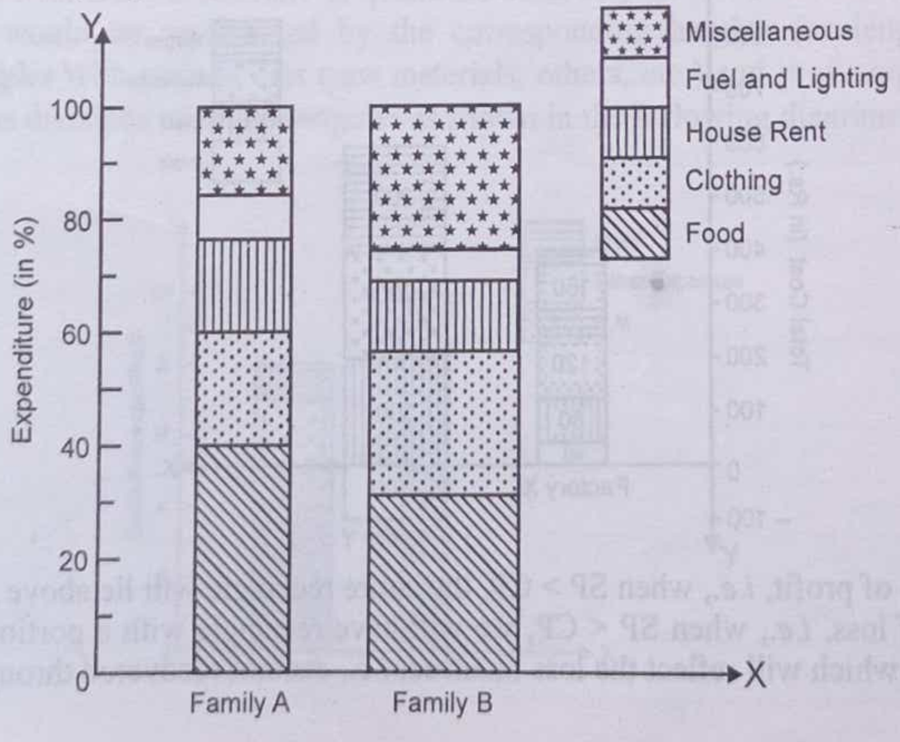
Example 14. Represent the following data through a suitable diagram:

Items of Expenditure	Family A (Income Rs. 500)	Family B (Income Rs. 800)
Food	200	250
Clothing	100	200
House Rent	80	100
Fuel and Lighting	40	50
Miscellaneous	80	200
Total	Rs. 500	Rs. 800

Solution: Since the total incomes of the two families are different, an appropriate diagram for the above data will be rectangular diagram on percentage basis. We first find the percentage of expenditure on each items for each family.

Item	Family A (Income Rs. 500)			Family B (Income Rs. 800)		
	Expenditure (Rs.)	%	Cumulative %	Expenditure (Rs.)	%	Cumulative %
Food	200	40	40	250	31.2	31.2
Clothing	100	20	60	200	25.0	56.2
House Rent	80	16	76	100	12.5	68.6
Fuel and Lighting	40	8	84	50	6.3	75.0
Miscellaneous	80	16	100	200	25.0	100
Total	500	100		800	100	

The width of the rectangles will be taken in the ratio of total income of families, *i.e* 500 : 800, *i.e.*, 5 : 8.



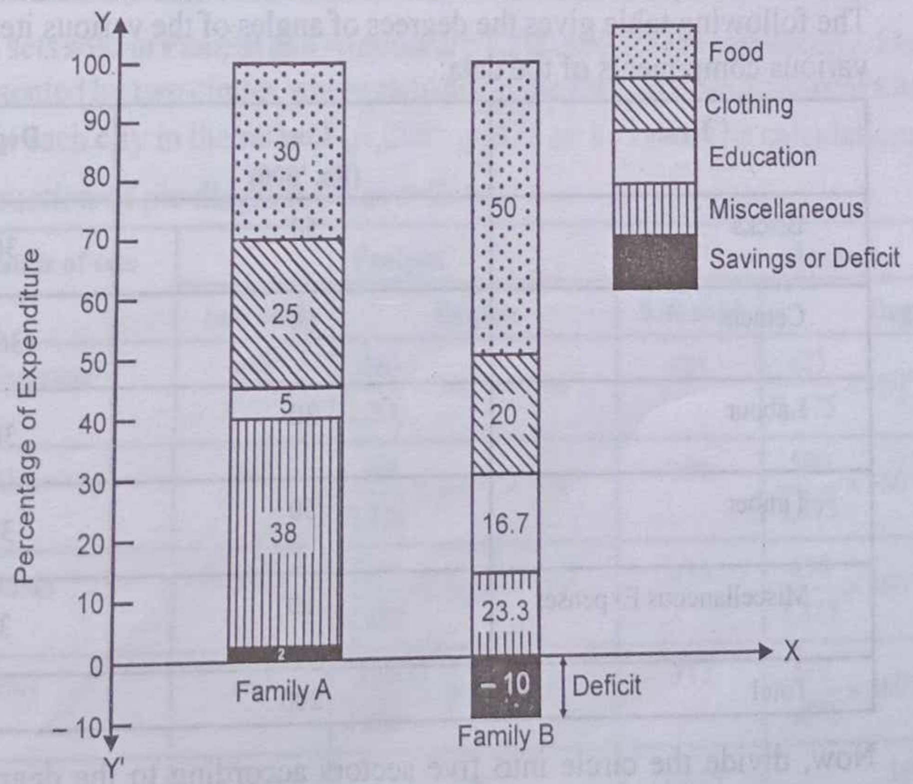
Example 15. Present the following information through a suitable diagram:

Items of Expenditure	Family A (Income Rs. 500)	Family B (Income Rs. 300)
Food	150	150
Clothing	125	60
Education	25	50
Miscellaneous	190	70
Savings or Deficit	+10	–30

Solution: Since the total incomes of the two families are different, an appropriate diagram for the above data will be rectangular diagram on percentage basis. We first find the percentage of expenditure on each item for each family.

Items of Expenditure	Family A			Family B		
	Expenditure (Rs.)	%	Cumulative (%)	Expenditure (Rs.)	%	Cumulative (%)
Food	150	30	30	150	50	50
Clothing	125	25	55	60	20	70
Education	25	5	60	50	16.7	86.7
Miscellaneous	190	38	98	70	23.3	110.0
Savings or Deficit	+10	2	100	–30	–10	100
Total	500			300		

The width of the rectangles will be taken in the ratio of the total income of families, *i.e.*, 500 : 300, *i.e.*, 5 : 3.



Note: In case total expenditure = total income, *i.e.*, when TE = TI, the entire rectangle will lie above the X-axis. But in case of deficit, *i.e.*, when TE > TI, we will have rectangles with a portion lying below the X-axis which will reflect the deficit.

(3) Pie Diagrams

A pie diagram is a circle sub-divided into component sectors. Just as a sub-divided bar or rectangle represents the whole data sub-divided into various components, similarly, a sub-divided circle represents the whole data sub-divided in to various component parts. For example, a pie diagram may show the distribution of money spent by the government on various heads of expenditure. The whole circle represents the total expenditure and various sectors of the circle show the percentage of total expenditure spent on various heads.

The following example would illustrate the construction of a pie diagram.

Example 16. Present the following data in the form of a pie diagram relating to cost of construction of a house in a city:

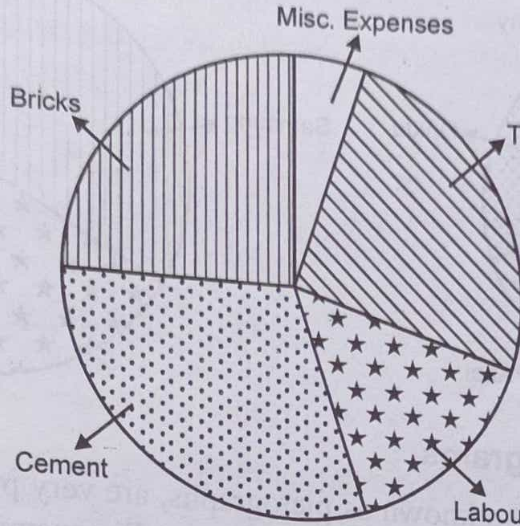
Items	Cost (Rs. '000)
Bricks	50
Cement	60
Labour	30
Timber	50
Miscellaneous Expenses	10
Total	200

Solution: First draw a circle of any radius. The circle, however, should neither be too big nor too small. The whole circle represents 100% of the total expenditure. The angle at the centre of the circle is of 360°. We calculate the degrees of angles of different items. The following table gives the degrees of angles of the various items representing the various components of the data:

Items	Cost (Rs. '000)	Degree of the angle
Bricks	50	$360°\times\frac{50}{200}=90°$
Cement	60	$360°\times\frac{60}{200}=108°$
Labour	30	$360°\times\frac{30}{200}=54°$
Timber	50	$360°\times\frac{50}{200}=90°$
Miscellaneous Expenses	10	$360°\times\frac{10}{200}=18°$
Total	200	360°

Now, divide the circle into five sectors according to the degrees of angles at the centre, as calculated in the above table. It is preferably to arrange the angles in the descending order and place the longest at the top.

The required pie diagram is as shown below:



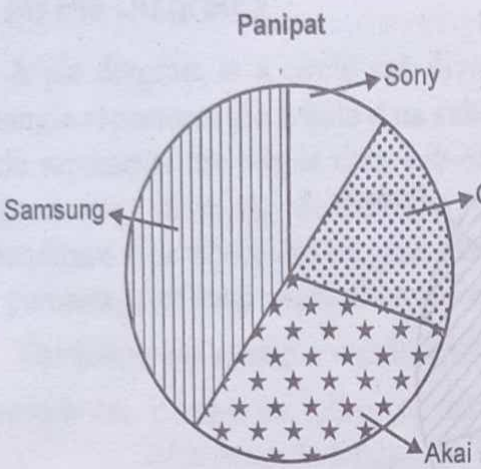
Example 17. Following are the data about the market share of 4 brands of T.V. sets sold in Panipat and Ambala. Present the data by a pie diagram.

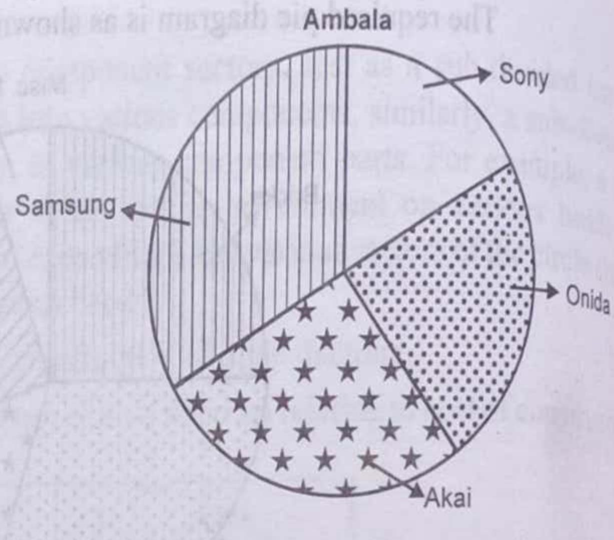
Brands of sets	Units sold in Panipat	Units sold in Ambala
Samsung	480	625
Akai	360	500
Onida	240	438
Sony	120	312

Solution: Total sets sold in Panipat and Ambala are 1,200 and 1,875 respectively. Data are to be represented by two circles whose radii are in the ratio of square roots of total T.V. sets sold in each city in the ratio of $\sqrt{1,200} : \sqrt{1,875}$ or 1 : 1.18. The calculations regarding construction of pie diagram are as follows:

Brands of sets	Panipat		Ambala	
	Sets sold	Degree	Sets sold	Degree
Samsung	480	$\frac{480}{1,200} \times 360^\circ = 144^\circ$	625	$\frac{625}{1,875} \times 360^\circ = 120^\circ$
Akai	360	$\frac{360}{1,200} \times 360^\circ = 108^\circ$	500	$\frac{500}{1,875} \times 360^\circ = 96^\circ$
Onida	240	$\frac{240}{1,200} \times 360^\circ = 72^\circ$	438	$\frac{438}{1,875} \times 360^\circ = 84^\circ$
Sony	120	$\frac{120}{1,200} \times 360^\circ = 36^\circ$	312	$\frac{312}{1,875} \times 360^\circ = 60^\circ$
Total	1,200	360°	1,875	360°
Square root of total	36.64		43.30	

Pie Diagram





(4) Pictograms and Cartograms

(*i*) **Pictograms:** Pictograms also known as pictographs, are very popularly used in presentin statistical data. Pictographs show data in the form of pictures. For examples, data on scooters woul be represented by pictures of scooters, data on milk diary by picture of cows, data on populatio would be represented by pictures of men, data on milk would be represented by the bottles of mil and the like.

(*ii*) **Cartograms:** Cartograms also known as cartographs are used to give quantitative informatio on a geographical basis. Cartographs show data in the form of maps. For example, rainfall i different parts of the country, size of population in different regions, sugar mills located in differen areas and the like. Can be represented by maps.

EXERCISE 4.2

1. Present the following data diagrammatically by using simple bar diagram:

Years:	1990	1991	1992	1993
No. of students:	400	500	700	400

2. Represent the following data by multiple bar diagram:

Years:	1991-92	1992-93	1993-94	1994-95
Imports (Rs. crore):	600	700	800	900
Exports (Rs. crore):	500	600	700	800

3. Represent the following data by sub-divided bar diagram:

Production of Electricity from Different Sources in India

('000 million Kwh)

Year	Hydro-Electricity	Thermal Electricity	Total Production
1992-93	46	64	110
1993-94	49	72	121
1994-95	48	82	130
1995-96	51	89	140

4. Represent the following data by a percentage sub-divided bar diagram:

Faculty	Number of students	
	1990–1991	2000–2001
Arts	300	720
Commerce	480	800
Science	420	480

5. Represent the following data diagrammatically:

Particulars	Firm 'A'	Firm 'B'	Firm 'C'
	Rs.	Rs.	Rs.
Wood	4	4	5
Labour	3	3.5	4
Polishing	2	2.5	2
Cost	9	10	11
Selling Price	10	10	10
Profit or Loss	+ 1	0	– 1

[**Hint:** See Example 34]

6. Represent the following data diagrammetically:

Items of Expenditure	Family A (Income Rs. 600)	Family B (Income Rs. 1000)
Food	200	400
Rent	100	200
Clothing	150	200
Fuel	100	100
Miscellaneous	50	50

[**Hint:** Saving of B is Rs. 50]

7. Represent the following data by a rectangular diagram:

	Commodities	
	A	B
Price per unit of commodity (Rs.)	10	12
Quantity sold	20	24
Cost of raw materials used (Rs.)	100	120
Other costs (Rs.)	60	96
Profit (Rs.)	40	72

8. Represent the following data by means of a pie diagram:

	Cost (Rs.)
Cost of Labour	10
Cost of Material	25
Cost of Electricity	5
Cost of Transportation	15
Overhead	35
Total	90

GRAPHIC PRESENTATION OF DATA

Graphic presentation is another method of data presentation. In this method, statistical d presented on the graph paper. Graphic presentation is the technique of presenting the data i form of curves or lines on the graph paper.

Utility of Graphic Presentation

The main advantage of graphic presentation of data are as follows:

(1) Presentation of Time Series and Frequency Distribution: Graphic Presentation is a effective technique of data presentation in case of time series data and frequency distributions

(2) Location of Averages: Using graphic technique, we can easily locate the value of ce averages, such as mode and median. It is not possible with the help of diagrams.

(3) Easy Estimation: Graphic presentation facilities interpolation and extrapolation of the in a more convenient and precise manner. For example, given the population data for the years l and 1971, one can easily make an estimate of the population in the years 1981 or 1991.

(4) Study of Correlation: Graphic technique helps in studying correlation between diffe variables, such as price and demand, cost and output, and the like.

(5) Comparison of Multiple-Dam: Data of different dimensions can be easily compared the help of graphic presentation.

In short, **Habbard** rightly observed, wherever keeping of record of the data, drawing conclusions, describing of facts are necessary, these graphs provide such an important means wh power we have started experiencing and using.

Limitations of Graphic Presentation

The following are some of the important limitations of graphic presentation of statistical d

(1) Less Significant: Graphs are not of equal significance to all the people. It is gener difficult for a layman to interpret graphs. Accordingly these are of little value to them.

(2) Only a Measure of Tendency: Graphs show only tendency of the data. Actual values not always clear from the graphs.

(3) Lack of Precise Value: Since graphs are based on brief information, these do not sh precise values.

(4) Wrong Conclusions: Graphs may sometimes suggest wrong conclusions. In fact eve small change in the scale of the graph causes a lot of difference in the structure of the graph. T may lead to wrong concusions.

General Rules for Constructing a Graph

The following points must be kept in mind while constructing a graph:

(1) Heading: Every graph must have a suitable and precise heading. Heading must be explanatory about the nature of information in the graph.

(2) Choice of Scale: One should fix an appropriate scale on which data should be presented. An appropriate scale is that scale by which the entire data is easily represented by the graph. The graph should be on the middle of the graph paper to make it attractive.

(3) Proportion of Axis: As far as possible, length of X-axis in the graph should be one and a half 1-1/2 times the length of Y-axis.

(4) Method of Plotting the Points: Economic and business statistics are generally positive. These are to be presented in the first quadrant of the graph. Accordingly the point of origin is fixed to the left and lower portion of the graph. On the X-axis, the points are plotted from left to right and on the Y-axis, the points are plotted upward from bottom to top.

(5) Lines of Different Types: If more than one line or curve are to be drawn in the same graph, these lines should be differentiated from each other in the form of broken lines (_ _ _), dotted lines (..........), bold lines (____), etc.

(6) Table of Data: It would be useful to give the table of data along with the graph of data. This helps in the verification of the curve.

(7) Use of false Base line: If the values in a series are very large and the difference between the smallest value and zero is high, then a false base line is drawn.

(8) To draw a line or curve: We mark different points on the graph paper corresponding to different values of a series. These points are joined to make line or curve. The joining line must be uniform throughout its length. It should not be of different thickness of its different points.

TYPES OF GRAPHS

Graphs can be broadly classified into the following two main types:

(A) Time Series Graphs

(B) Frequency Distribution Graphs.

(A) Time Series Graphs

When we observe the values of a variable at different points of time, the series is called time series. A time series is, thus, a chronological arrangement of statistical data. The time series can be represented geometrically with the help of a graph. In time series graph, the time, *i.e.*, years, months, weeks, etc., are taken on the X-axis and the corresponding values of the variables are shown on the Y-axis.

There are many types of time series graphs which are described as follows:

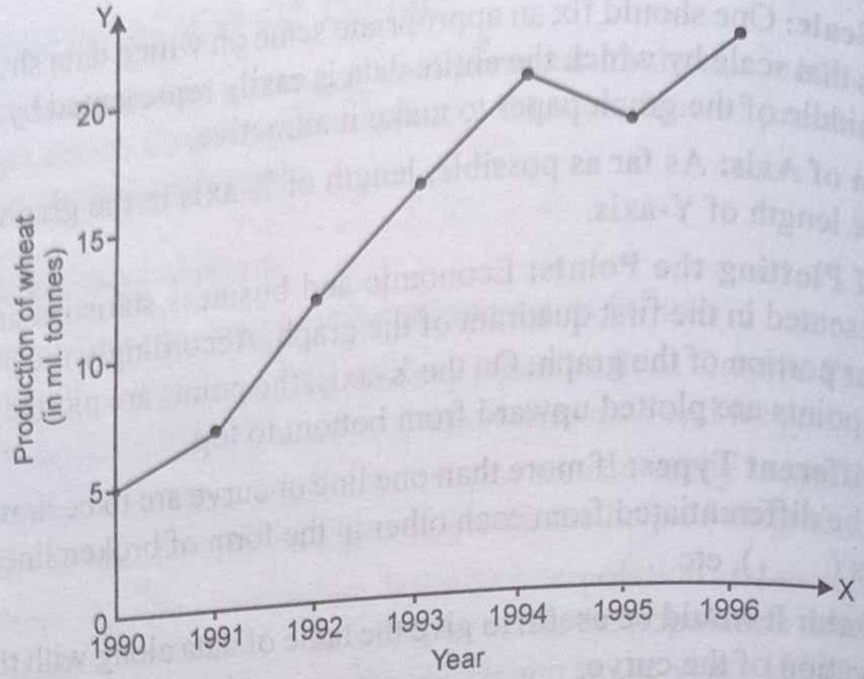
(1) Time Series Graph of One Variable

Time series graphs can be constructed for one variable in which the points are plotted and then joined to form a curve.

Example 18. Represent the following data graphically:

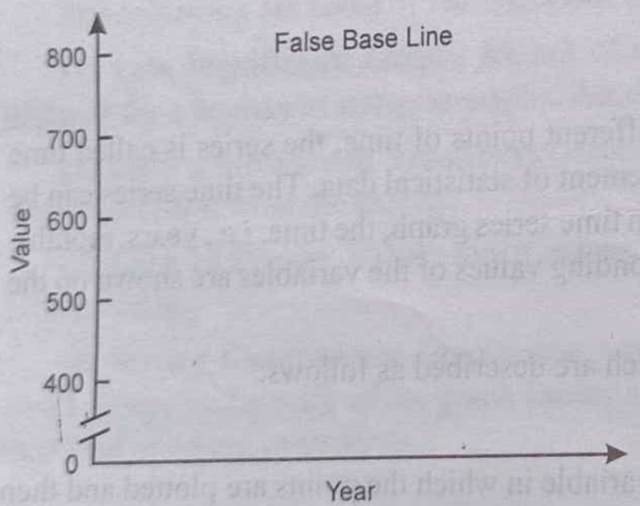
Years:	1990	1991	1992	1993	1994	1995	1996
Production of wheat: (in ml. tonnes)	5	8	13	16	20	17	22

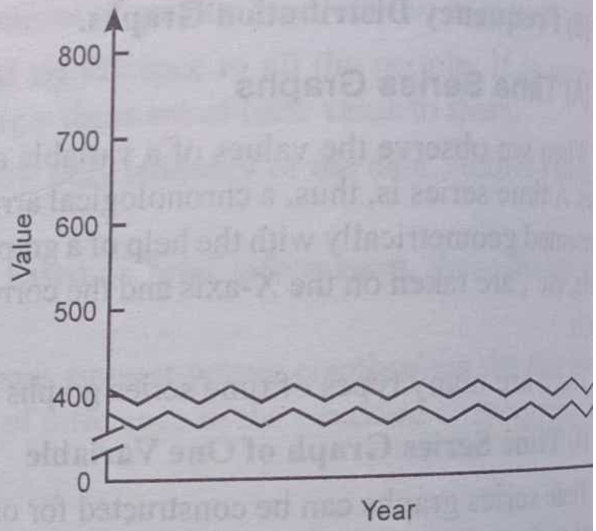
Solution:



False Base Line

A basic rule of the construction of graphs is that the scale on the Y-axis must begin from z
Such a rule can be followed in all such cases where (*i*) values of the dependent variable vary gre
and (*ii*) these values lie between zero and the highest value in the scale (or the graph). In many c
this is not so, *i.e.*, (a) the values of the Y or dependent variables lie within a narrow range and (b
least value is far away from zero. In such cases, false base line technique is used. The techniqu
false base line is very simple. The vertical scale is broken very close to the base line or near 'z
which is shown by leaving the space blank or by using a zig-zig line. The following graphs illust
the idea:

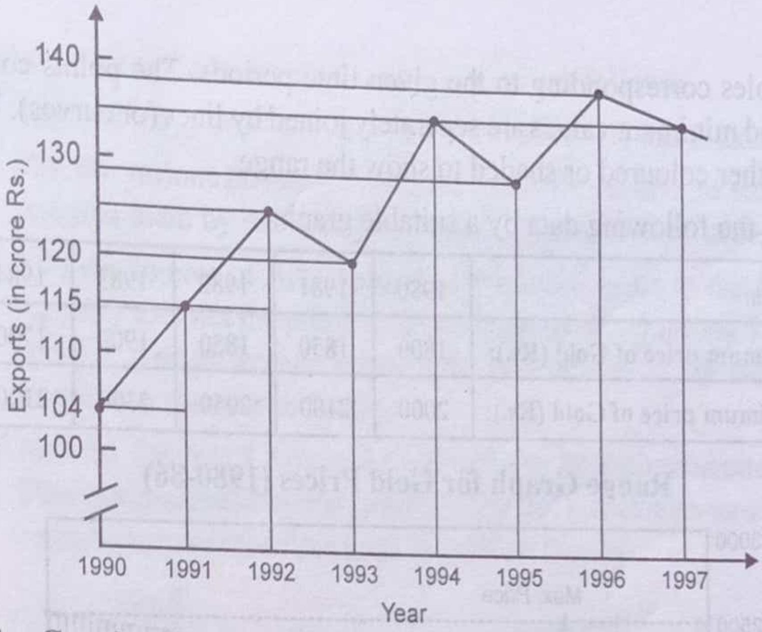




Example 19. Represent the following data graphically:

Years:	1990	1991	1992	1993	1994	1995	1996	19
Exports (in crore Rs.):	104	115	125	120	135	128	138	13

Solution:



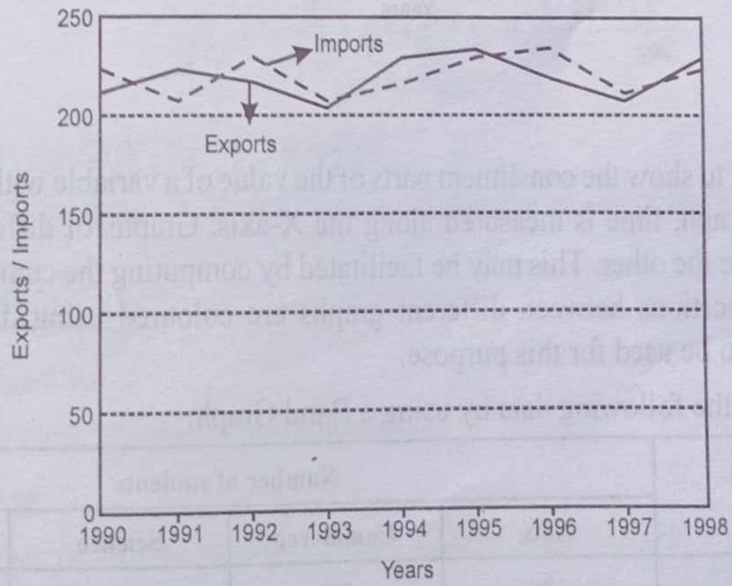
(2) Time Series Graphs of Two or More Variables

Two or more variables can also be shown in a graph provided the units of measurement is the same. If two or more variables are represented on the same graph, it would be better if lines (curves) of different types are drawn. For example, one line is dotted, another thin and a third one thick, etc. can be drawn to distinguish between the variables.

Example 20. Represent the following data graphically:

Years:	1990	1991	1992	1993	1994	1995	1996	1997	1998
Exports (in crore Rs.):	212	220	215	205	225	228	216	208	225
Imports (in crore Rs.):	220	210	225	210	215	225	228	212	220

Solution:



(3) Range Graph

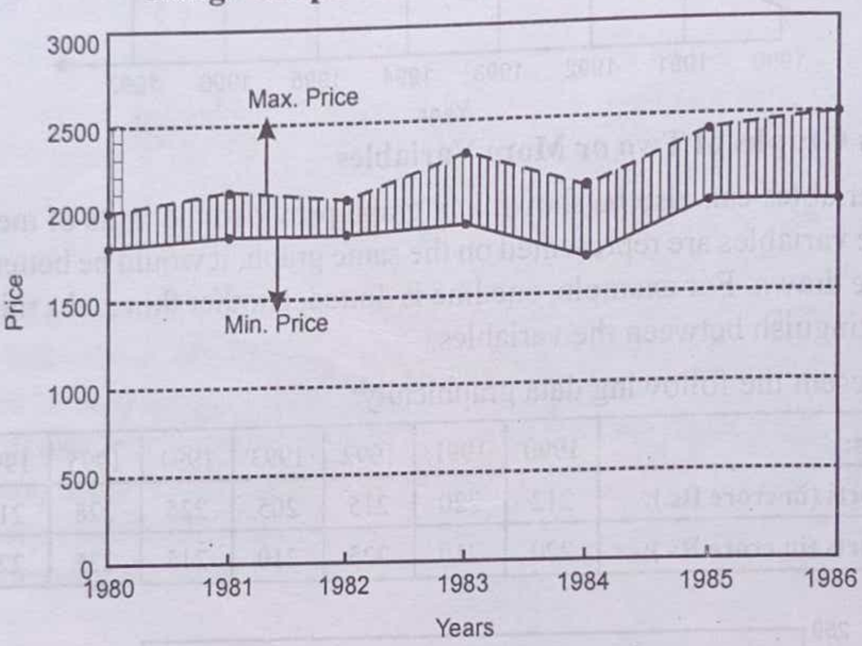
Range graphs or charts are used to depict the maximum and minimum values of variables with respect to time. For example, we may use range graph to show the maximum and minimum prices of commodities, shares, gold, etc. In range graph, time is measured along the X-axis and maximum and minimum value on the Y-axis. Points are plotted corresponding to the maximum and minimum values of the variables corresponding to the given time periods. The points corresponding to th maximum values and minimum values are separately joined by lines (or curves). The space betwe the two curves is either coloured or shaded to show the range.

Example 21. Show the following data by a suitable graph:

Years:	1980	1981	1982	1983	1984	1985	1986
Minimum price of Gold (Rs.):	1800	1850	1850	1900	1700	2000	2000
Maximum price of Gold (Rs.):	2000	2100	2050	2300	2100	2400	2500

Solution:

Range Graph for Gold Prices (1980-86)



▶ (4) Band Graph

Band graphs are used to show the constituent parts of the value of a variable with respect to tim For constructing band graph, time is measured along the X-axis. Graphs of different constitue parts are drawn one above the other. This may be facilitated by computing the cumulative totals f each time period. The portions between different graphs are coloured using different colou Different shades may also be used for this purpose.

Example 22. Represent the following data by using a Band Graph:

Years	**Number of students**			
	Arts	**Commerce**	**Science**	**Total**
1992	300	100	50	450
1993	400	150	100	650
1994	600	100	150	850
1995	1000	200	150	1350
1996	1000	200	200	1400

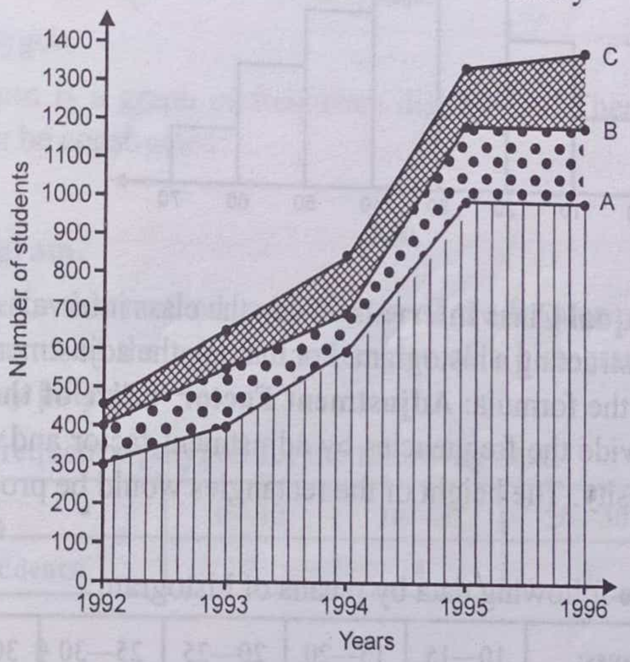
Solution: The procedure of constructing such a graph is as follows:

(*i*) Take the years on the X-axis and the variable on the Y-axis.

(*ii*) Plot the various points for different years for number of students in arts faculty and join them by straight lines. This is represented by curve A.

(*iii*) Add the figures of Arts' students for various years to the figures of commerce students and plot the points and join them by straight line. This is represented by curve B. The difference between the two curves, *i.e.*, B and A, gives number of students in commerce faculty.

(*iv*) Add the figures of Arts and Commerce to Science students and plot the points. This is represented by the curve C. The difference between curve C and curve B represents number of students in science faculty.



(B) Frequency Distribution Graphs

A frequency distribution can be presented graphically in any of the following ways:

(1) Histogram

(2) Frequency Polygon

(3) Frequency Curve

(4) Cumulative Frequency Curve or Ogive

(1) Histogram

The histogram is the most popular and widely used method of presenting a frequency distribution graphically. Histogram is a set of adjoining rectangles whose areas are proportional to class frequencies. For constructing historgram, class intervals are taken on the X-axis and frequencies on Y-axis. Histograms are mainly of two types:

(*a*) Histogram for equal class intervals.

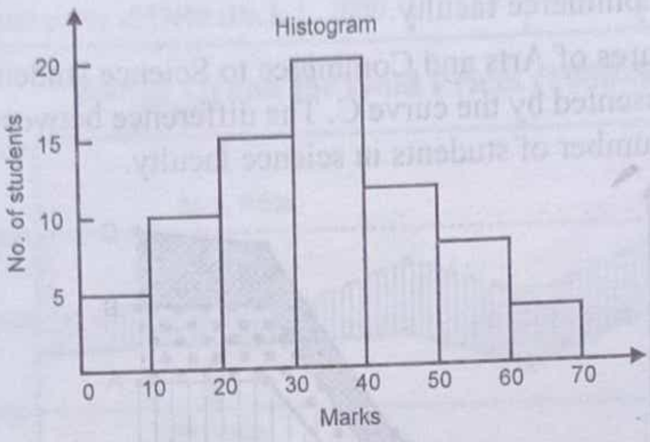
(*b*) Histogram for unequal class intervals.

(*a*) **Histogram for Equal Class Intervals:** When class intervals are equal, we take class frequency on the Y-axis, the class intervals on the X-axis and construct adjacent rectangles. The height of the rectangles will be proportional to the frequencies.

Example 23. Represent the following data by means of a histogram:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of students:	5	10	15	20	12	8	4

Solution:



(*b*) **Histogram for Unequal Class Intervals:** When the class intervals are unequal, frequencies are first adjusted before constructing a histogram. For making the adjustment, we first determine the adjustment factor by using the formula: **Adjustment Factor = Size of the class interval/Lowest class interval.** Then we divide the frequencies by adjustment factor and obtain the new adjusted frequency or frequency density. The height of the rectangles would be proportional to the adjusted frequencies.

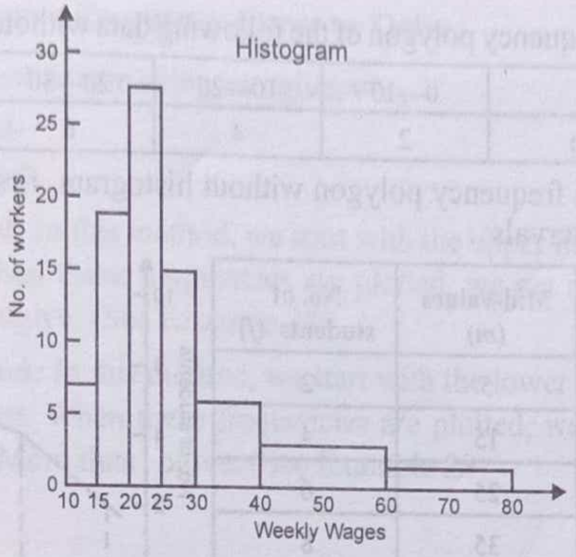
Example 24. Represent the following data by means of histogram:

Weekly wages:	10—15	15—20	20—25	25—30	30—40	40—60	60—80
No. of workers:	7	19	27	15	12	12	8

Solution: Since the class intervals are unequal, frequencies are to be adjusted.

Computations of New Frequency or Frequency Density

Weekly wages	**No. of workers (*f*)**	**Adjustment factor**	**New frequencies**
10—15	7	5/5 = 1	7
15—20	19	5/5 = 1	19
20—25	27	5/5 = 1	27
25—30	15	5/5 = 1	15
30—40	12	10/5 = 2	12 ÷ 2 = 6
40—60	12	20/5 = 4	12 ÷ 4 = 3
60—80	8	20/5 = 4	8 ÷ 4 = 2



▶ (2) Frequency Polygon

A frequency polygon is a graph of frequency distribution. There are two ways in which a frequency polygon may be constructed:

***(a)* By Histogram**

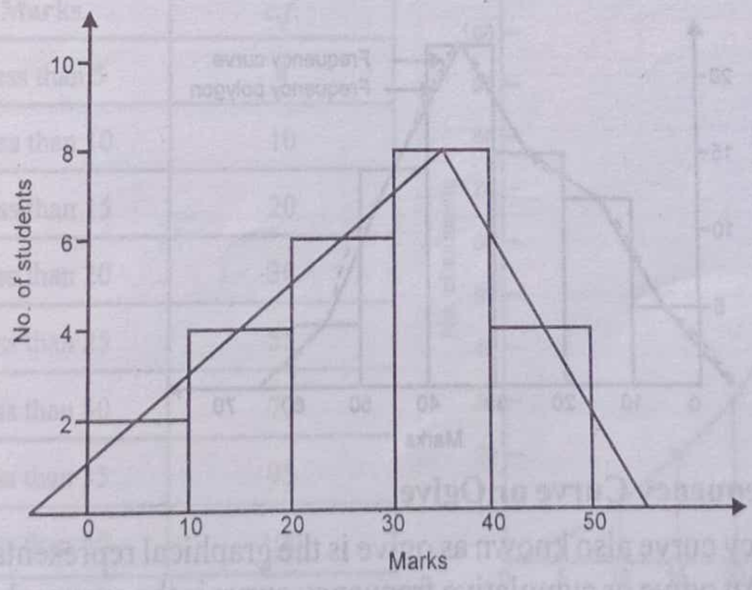
***(b)* Without Histogram.**

***(a)* By Histogram:** In this method, a histogram of the frequency distribution is first drawn. Then the upper mid-points of the adjacent rectangles of a histogram are joined. The figure so obtained is the frequency polygon.

Example 25. Draw a frequency polygon for the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	2	4	6	8	4

Solution:



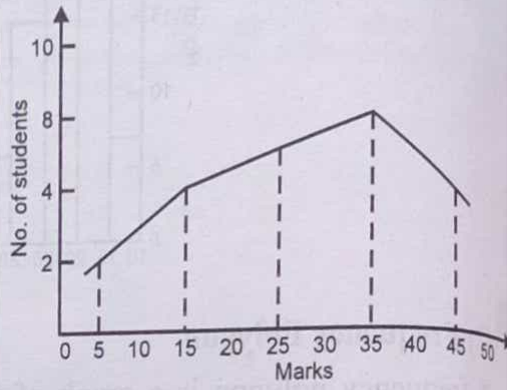
(b) Without Histogram: Frequency polygon can also be constructed without the help of histogram. Under this method, mid-points of various class intervals are taken and the corresponding frequencies to each mid-points are plotted. Then we join all these points by straight lines. The figure so obtained would be the same as obtained by the previous method.

Example 26. Construct a frequency polygon of the following data without constructing a histogram

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	2	4	6	8	4

Solution: To construct a frequency polygon without histogram, firstly we find the mid-point of the class intervals:

Marks	Mid-values (m)	No. of students (f)
0—10	5	2
10—20	15	4
20—30	25	6
30—40	35	8
40—50	45	4



▶ (3) Frequency Curve

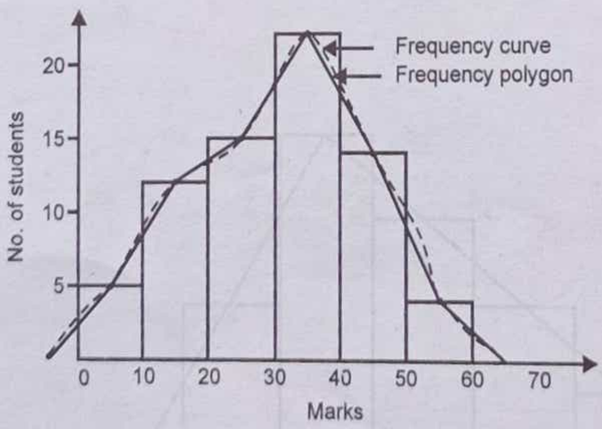
Frequency curve is the smoothed form of a frequency polygon. That is why it is known as smoothed **frequency curve**. A frequency curve can be drawn through different points of the polygon. The curve is drawn freehand in such a way that the area included under it is approximately the same as that of the polygon. With a freehand, the angularities of the frequency polygon are eliminated.

For drawing a frequency curve, it is necessary to first draw the frequency polygon and then smooth it out in such a way that all sudden turns are eliminated.

Example 27. Draw a histogram, a frequency polygon and frequency curve of the following data

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	5	12	15	22	14	4

Solution:



▶ (4) Cumulative Frequency Curve or Ogive

Cumulative frequency curve also known as ogive is the graphical representation of a cumulative frequency distribution. An ogive or cumulative frequency curve is the curve which is constructed by plotting the cumulative frequencies in the form of a smooth curve. From such curves, we come to know about the frequencies corresponding to certain lower limits or upper limits in the distribution of data. For example, such curves would indicate how many students in a class secured more than 40 marks or how many students secured less than 35 marks in an examination.

Construction of a Cumulative Frequency Curve or Ogive

There are two methods of constructing an ogive, viz.,

(*a*) **'Less than' method.**

(*b*) **'More than' method.**

(*a*) **'Less than' Method:** In this method, we start with the upper limits of the classes and go on adding the frequencies. When these frequencies are plotted, we get a rising curve. The resultant curve is called 'Less than' ogive. (See Example 28)

(*b*) **'More than' Method:** In this method, we start with the lower limits of the classes and add frequencies from the bottom. When these frequencies are plotted, we get a declining curve. The resultant curve is called a 'More than' ogive. (See Example 28)

Utility of Ogives

(1) To determine the number or proportion of cases below or above a given value.

(2) To compare two or more frequency distributions.

(3) Ogives are also used to determine certain values graphically such as median, quartiles, deciles, etc.

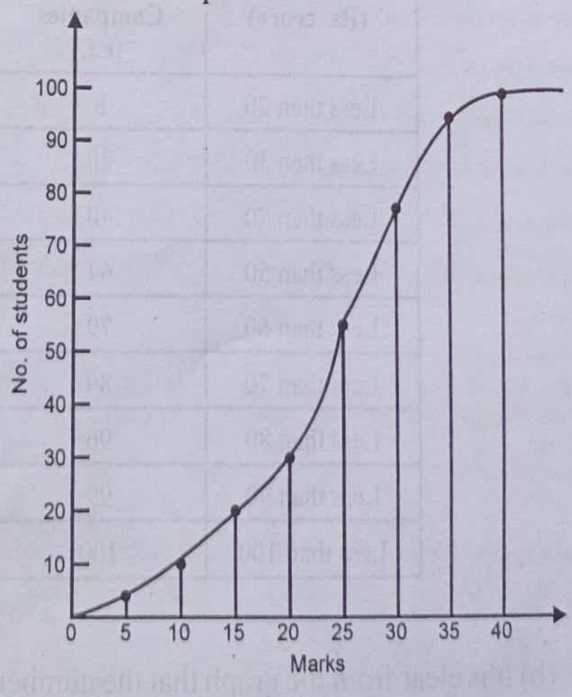
The following examples would illustrate the construction of ogives:

Example 28. Draw less than and more than ogives from the data given below:

Marks:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
No. of students:	4	6	10	10	25	22	18	5

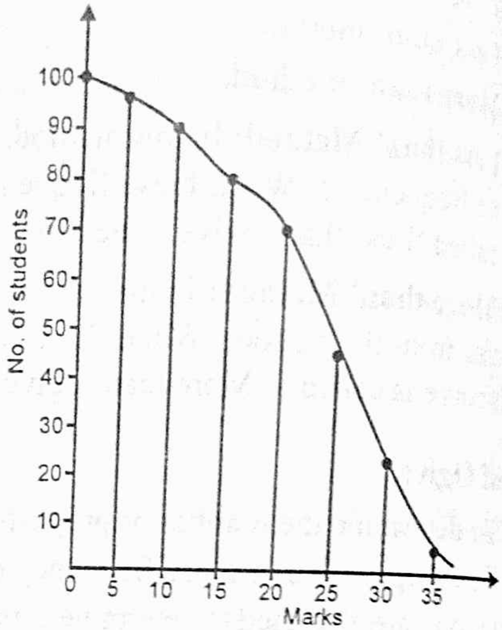
Solution: **Less than ogive:** In order to draw less than ogive, we start with the upper limits of the classes and go on adding the frequencies from the top.

Marks	*c.f.*
Less than 5	4
Less than 10	10
Less than 15	20
Less than 20	30
Less than 25	55
Less than 30	77
Less than 35	95
Less than 40	100



More than ogive: In order to draw more than ogive, we start with the lower limits of the classes and go on adding the frequencies from the bottom:

Marks	c.f.
More than 0	100
More than 5	96
More than 10	90
More than 15	80
More than 20	70
More than 25	45
More than 30	23
More than 35	5

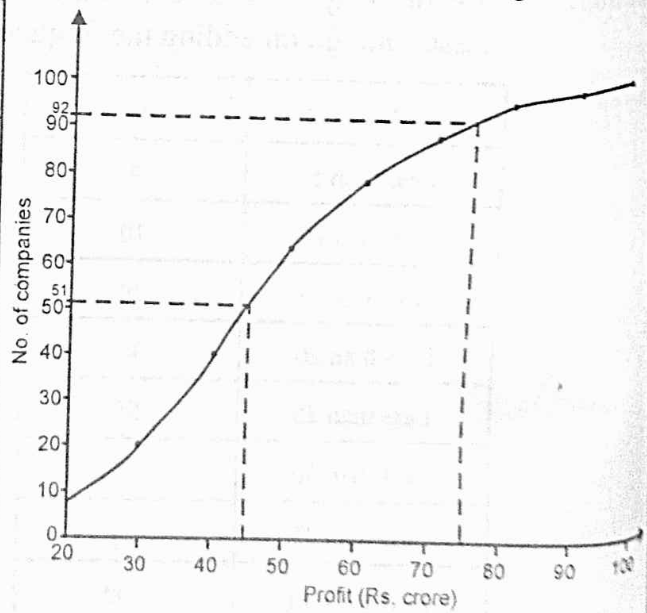


Example 29. Draw a less than ogive and determine the number of companies getting profits between Rs. 45 crores and Rs. 75 crores.

Profits (Rs. crore):	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90	90—100
No. of companies:	8	12	20	24	15	10	7	3	1

Solution: Firstly, we arrange the frequency distribution for less than method as given below:

Profits (Rs. crore)	No. of Companies (c.f.)
Less than 20	8
Less than 30	20
Less than 40	40
Less than 50	64
Less than 60	79
Less than 70	89
Less than 80	96
Less than 90	99
Less than 100	100



(b) It is clear from the graph that the number of companies getting profits less than Rs.75 crore is 92 and the number of companies getting profits less than Rs. 45 crores is 51. Hence the number of companies getting profits between Rs. 45 crores and Rs. 75 crores is $92 - 51 = 41$.

EXERCISE 4.3

1. Represent the following data graphically:

Years:	1980	1981	1982	1983	1984	1985
Wheat Production (in '000 tonnes):	5	12	8	15	20	17

2. Represent the following data graphically:

Years:	1980-81	1981-82	1982-83	1983-84	1984-85	1985-86	1986-87
Exports (Rs. in lakh):	51	56	58	65	60	66	61
Imports (Rs. in lakh):	53	54	60	70	65	73	70

3. Following are the maximum and minimum market values of the shares of a company. Represent the data by using a range chart.

Years:	July '86	Aug. '86	Sept. '86	Oct. '86	Nov. '86	Dec. '86
Minimum (in Rs.):	45	41	47	42	50	55
Maximum (in Rs.):	50	60	67	68	68	70

4. Represent the following data by using band chart:

Year	Production (in '000 tonnes)		
	Wheat	Rice	Pulses
1996	20	8	4
1997	22	10	3
1998	25	7	5
1999	23	11	7
2000	30	13	4
2001	25	12	5
2002	31	15	7

5. Construct a histogram from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of students:	5	12	20	35	24	12	4

6. Represent the following data by means of a histogram:

Daily wages (Rs.):	10—20	20—30	30—40	40—60	60—100
No. of workers:	5	20	30	40	20

7. Construct a histogram with the help of data given below:

Marks:	10—19	20—29	30—39	40—49	50—59	60—69
No. of students:	4	25	45	60	35	10

8. Draw a frequency polygon from the following data by using (a) histogram (b) without using histogram:

Daily wages (in Rs.):	10—15	15—20	20—25	25—30	30—35
No. of workers:	40	70	60	80	60

9. From the following data, draw: (*i*) Histogram, (*ii*) Frequency polygon and (*iii*) Frequency curve.

Wage groups (Rs.):	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90
No. of workers:	2	4	11	15	25	18	15	4	2

10. From the following data, draw: (*i*) Histogram, (*ii*) Frequency polygon and (*iii*) Frequency curve.

Wage (in Rs.):	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of persons:	2	7	15	25	18	15	8

11. Draw 'less than' as well as 'more than' ogives for the following data:

Weight (in kg)	30—34	35—39	40—44	45—49	50—54	55—59	60—64
Frequency:	3	5	12	18	14	6	2

(Hint: Convert into Exclusive Groups]

12. Prepare a less than cumulative frequency curve for the following data:

Weekly wages:	0—20	20—40	40—60	60—80	80—100
No. of workers:	40	51	64	38	7

Find the number of workers earning between Rs. 55 and Rs. 65 per week.

MISCELLANEOUS SOLVED EXAMPLES

Example 30. In a sample study about coffee habits in two towns, the following information is given:

Town A: Females were 40%, total coffee drinkers were 45% and male non-coffee drinkers were 20%.

Town B: Males were 55%, male non-coffee drinkers were 30% and female coffee drinkers were 15%.

Present the above data in a tabular form.

Solution:

Percentage of coffee drinkers in Towns A and B

	Town A			Town B		
	Male	Female	Total	Male	Female	Total
Coffee Drinkers	40	5	45	25	15	40
Non-coffee Drinkers	20	35	55	30	30	60
Total	60	40	100	55	45	100

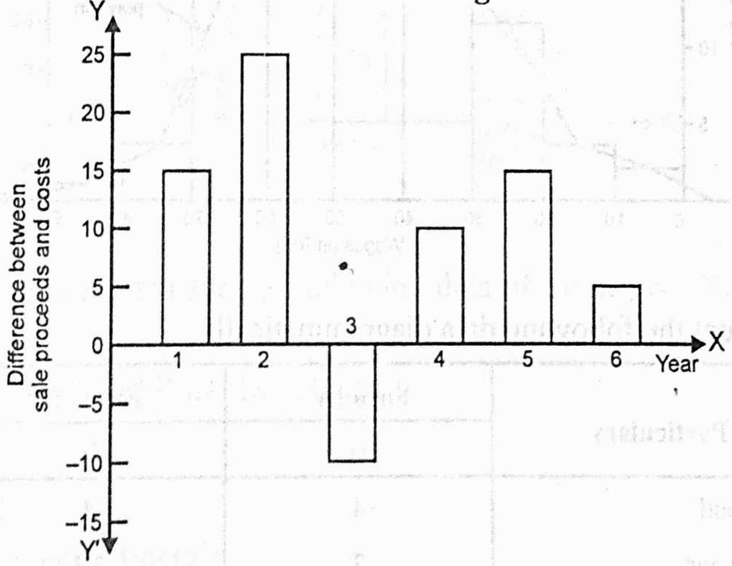
Example 31. Present the following data by a deviation bar diagram, showing the difference between proceeds and costs of a firm.

Year	Sale proceeds (Rs. in lakh)	Costs (Rs. in lakh)
1	115	100
2	140	115
3	145	155
4	150	140
5	160	145
6	170	165

Solution: In order to present the data by deviation bar diagram, we first calculate the deviations between sale proceeds and cost as follows:

Year	Sale proceeds	Costs	Difference between sale proceeds and costs
1	115	100	15
2	140	115	25
3	145	155	–10
4	150	140	10
5	160	145	15
6	170	165	5

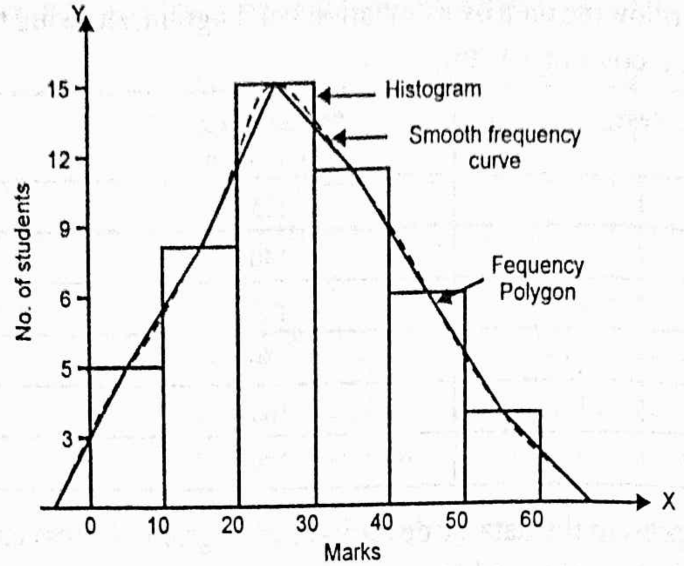
Deviation Bar Diagram



Example 32. Prepare histogram, frequency polygon and smooth frequency curve from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	5	8	15	11	6	4

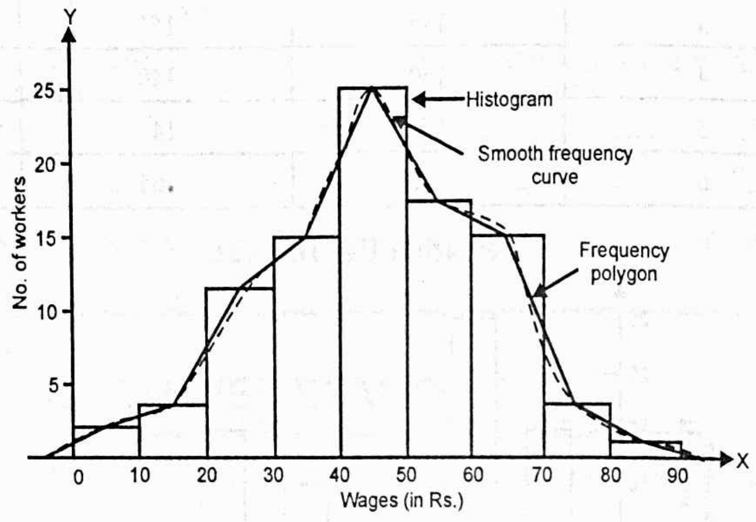
Solution:



Example 33. From the following data, construct (1) Frequency Histogram (2) Frequency Polygon and (3) Frequency Curve.

Wages (in Rs.):	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90
No. of workers:	2	4	11	15	25	18	15	4	1

Solution:



Example 34. Represent the following data diagrammatically:

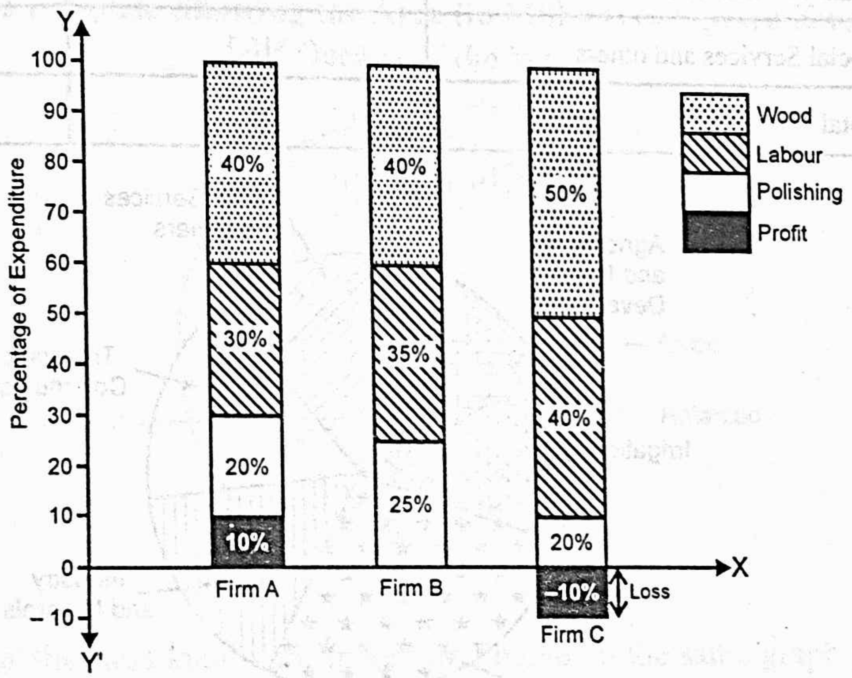
Particulars	Firm 'A'	Firm 'B'	Firm 'C'
	Rs.	Rs.	Rs.
Wood	4	4	5
Labour	3	3.5	4
Polishing	2	2.5	2
Cost	9	10	11
Selling Price	10	10	10
Profit or Loss	+ 1	0	– 1

Solution:

Selling Price of Chair on Percentage Basis

Particulars	Firm 'A'		Firm 'B'		Firm 'C'	
	Actual	%	Actual	%	Actual	%
Wood	4	40	4.0	40	5	50
Labour	3	30	3.5	35	4	40
Polishing	2	20	2.5	25	2	20
Cost	9	90	10.0	100	11	110
Selling Price	10	100	10.0	100	10	100
Profit or Loss	+ 1	+ 10%	—	—	− 1	− 10%

Diagram Showing % Expenditure of Chair of 3 Firms



Example 35. Draw a pie diagram for the following data of Sixth Five Year Plan Public Sector outlays:

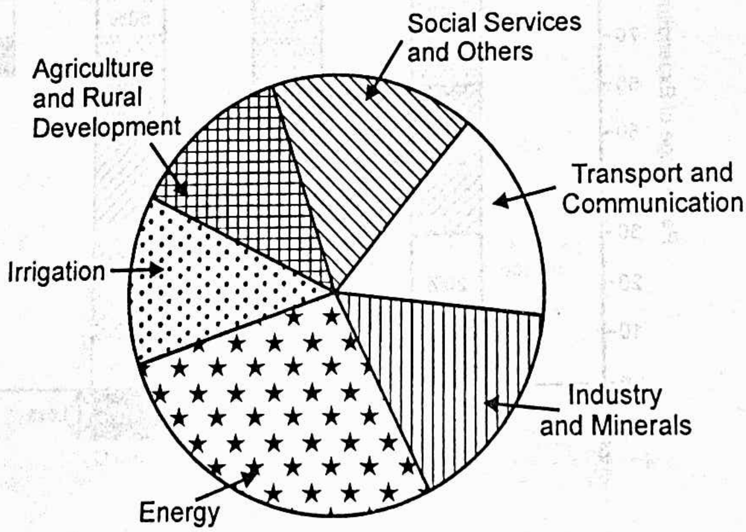
Agricultural and Rural Development	12.9%
Irrigation	12.5%
Energy	27.2%
Industry and Minerals	15.4%
Transport and Communication	15.9%
Social Services and Others	16.1%

Solution: The angle at the centre is given by

$$\frac{\text{Percentage outlay}}{100} \times 360 = \text{Percentage outlay} \times 3.6^\circ$$

Computations for Pie Diagram

Sector	Percentage outlays	Angle of the centre
(1)	(2)	(3) = (2) × 3.6°
Agriculture and Rural Development	12.9	12.9 × 3.6 = 46°
Irrigation	12.5	12.5 × 3.6 = 45°
Energy	27.2	27.2 × 3.6 = 98°
Industry and Minerals	15.4	15.4 × 3.6 = 56°
Transport and Communication	15.9	15.9 × 3.6 = 57°
Social Services and others	16.1	16.1 × 3.6 = 58°
Total	100.0	360°



Example 36. The following table shows the area in millions of sq. km of oceans of the world:

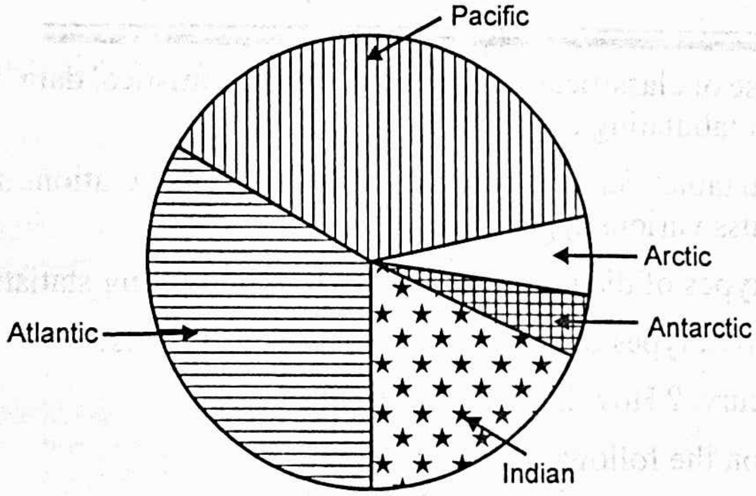
Ocean	Area (million sq. km)
Pacific	70.8
Atlantic	41.2
Indian	28.5
Antarctic	7.6
Arctic	4.8

Draw a pie diagram to represent the data.

Solution: **Calculations for Pie Diagram**

Ocean	Area	Degrees
Pacific	70.8	$\frac{70.8}{152.9} \times 360 = 166.7°$
Atlantic	41.2	$\frac{41.2}{152.9} \times 360 = 97.0°$
Indian	28.5	$\frac{28.5}{152.9} \times 360 = 67.1°$
Antarctic	7.6	$\frac{7.6}{152.9} \times 360 = 17.9°$
Arctic	4.8	$\frac{4.8}{152.9} \times 360 = 11.3°$
Total	152.9	360°

Pie Diagram Showing the Area (in Millions of Square Kms) of Oceans of the World



Example 37. Draw the 'less than' and 'more than' ogive on the same graph paper from the data given below:

Weekly Wages (Rs.):	0—20	20—40	40—60	60—80	80—100
No. of Workers:	10	20	40	20	10

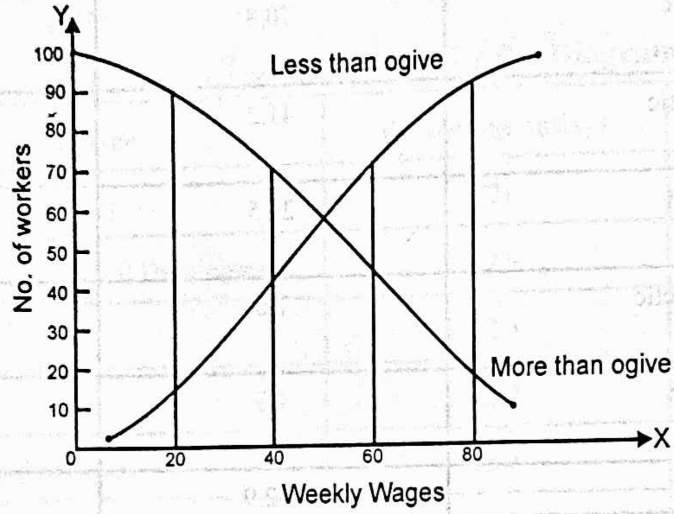
Solution: (*i*) Less than method

Weekly wages (Rs.)	***c.f.***
Less than 20	10
Less than 40	30
Less than 60	70
Less than 80	90
Less than 100	100

(*ii*) More than method

Weekly wages (Rs.)	**c.f.**
More than 0	100
More than 20	90
More than 40	70
More than 60	30
More than 80	10

Both less than and more than ogives based on the above data are presented in the following graph:



QUESTIONS

1. Explain the purpose of classification and tabulation of statistical data. Describe the rules that serve as a guide in tabulating the statistical data.
2. What is a statistical table? State its components. What considerations should govern framing a table? Also discuss various types of tables.
3. Explain different types of diagrams that are used in presenting statistical data.
4. What are the different types of frequency distributions graphs?
5. What is an ogive curve? How is it constructed?
6. Write short notes on the following:
   (*i*) Histogram; (*ii*) Ogive;
   (*iii*) Frequency Polygon; (*iv*) Frequency Curve;
   (*v*) Pie Diagram.
7. Discuss the utility and limitations of graphic method of presentation of statistical data.
8. Explain the advantages and limitations of diagrammetic and graphic presentations.
9. What is frequency distribution? Describe various ways by which a frequency distributio can be represented graphically.

Measures of Central Tendency

5

◘ INTRODUCTION

A large number of big figures is confusing to mind. It is also difficult to analyse them. In order to reduce the complexity of data and to make them comparable, it is essential that the various figures which are being compared are reduced to one single figure each. If, for example, a comparison is made between the marks obtained by 100 students of B.Com. II class of a college and the marks obtained by 100 students belonging to B.Com. II class of another college, it would be impossible to arrive at any conclusion, if the two series relating to these marks are directly compared. On the other hand, if each of these series is represented by one single figure, comparison and understanding would become very easy. The single figure which represents the whole set of data is called an average. Averages are also called **measures of central tendency** or **measures of location.**

◘ MEANING OF AVERAGE OR CENTRAL TENDENCY

An average is a single value which represents the whole set of figures and all other individual items concentrate around it. In other words, an average is single value within the range of the data that is used to represent all of the values in the series. Such an average is somewhere within the range of the data, it is therefore called measure of central tendency.

◘ DEFINITIONS

Some important definitions of average are given below:

1. "*An average is a single figure that represents the whole group.*" **—Clark**
2. "*An average is a single value selected from a group of values to represent them in some ways.*" **—A.E. Waugh**
3. "*An average is a typical value that represents all the individual values in a series.*" **—Croxten and Cowden**

The above definitions make it clear that an average is a single value that represents a group of values.

◘ PURPOSE AND FUNCTIONS OF AVERAGE

Some of the important functions and purposes of averages are as under:

***(i)* Brief Description:** The main purpose of average is to present a simple and systematic description of the raw data. The raw data may be complex and unorganised. An average reduces a mass of data into a single typical figure. It enables one to draw a general conclusion about the characteristics of the phenomena under study.

(ii) Helpful in Comparison: Averages help in making comparison of different sets of data. For example, a comparison of the per capita income of India and USA shows that per capita income of India is much less in comparison to the per capita income of USA. It leads us to the conclusion that India is a poor country in comparison to USA.

(iii) Helpful in the Formulation of Policies: Averages help in the formulation of policies. To illustrate, when the Government finds that the average per capita income in India is very low, it can formulate suitable policies to increase it.

(iv) Basis of Statistical Analysis: Averages constitute the basis of statistical analysis. For example, if one knows the average marks secured by the students of a class in their different subjects, one can easily analyse the general interest of the students in different subjects.

(v) Representation of the Universe: Averages represent the universe or the mass of statistical data. Accordingly, averages help in knowing the characteristics of the universe as a whole. For example, by calculating the per capita income of a nation, we get one single value that gives an idea of the economic condition of that nation.

CHARACTERISTICS/PROPERTIES OF A GOOD AVERAGE

A good average should possess the following properties:

(1) It should be easy to understand.

(2) It should be simple to compute.

(3) It should be uniquely defined.

(4) It should be based on all observations.

(5) It should not be unduly affected by extreme values.

(6) It should be capable of further algebraic treatment.

TYPES OF AVERAGES

There are different kinds of averages. The following are the important types of averages which are commonly used in business and industry:

1. Mathematical Averages

(i) Arithmetic Mean

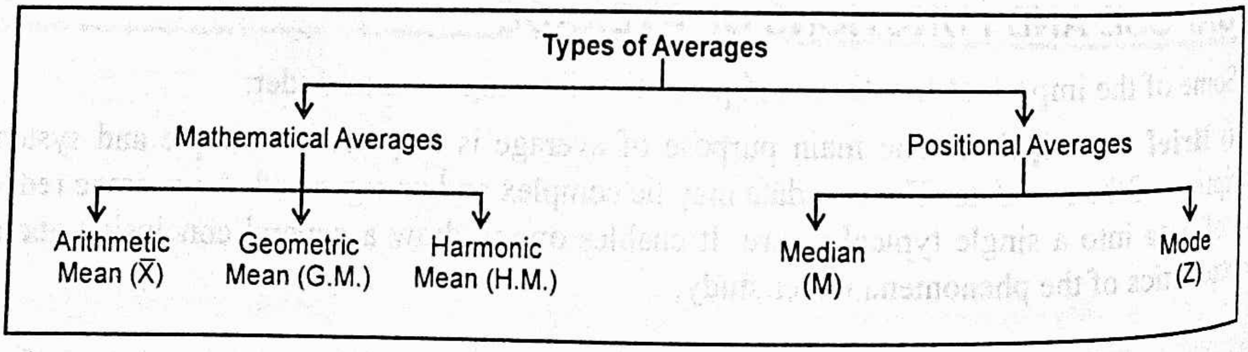
(ii) Geometric Mean

(iii) Harmonic Mean

2. Positional Averages

(i) Median

(ii) Mode



(1) ARITHMETIC MEAN

Arithmetic mean is the most popular and widely used measure of central tendency. Generally, when we talk of 'average', it signifies arithmetic mean. It is generally known as 'Mean'. Arithmetic mean is defined as the value which is obtained by adding all the items of a series and dividing this total by the number of items. Arithmetic mean may be of two types:

(a) Simple Arithmetic Mean.

(b) Weighted Arithmetic Mean.

Simple Arithmetic Mean

Individual Series

In case of individual series, arithmetic mean can be computed by applying any of the two methods:

(i) Direct Method

When direct method is used, the following formula is used:

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

Here, $\overline{X}$ = Arithmetic mean; ΣX = Sum of the values of the item of a series.
N = Number of observations.

Steps for Calculation

(i) Add together all the values of the variable X and obtain the total, *i.e.*, ΣX

(ii) Divide this total by the number of observations, *i.e.*, N. The result will give the value of arithmetic mean.

Example 1. The pocket allowances (in Rs.) of ten students are given below:

15, 20, 30, 22, 25, 18, 40, 50, 55 and 65

Calculate the arithmetic mean of pocket allowance.

Solution: Let pocket allowance be denoted by X

Students	Pocket Allowance (X)
1	15
2	20
3	30
4	22
5	25
6	18
7	40
8	50
9	55
10	65
$N = 10$	$\Sigma X = 340$

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

$$\therefore \quad \overline{X} = \frac{340}{10} = 34$$

Thus, the average pocket allowance is Rs. 34.

▶ (*ii*) Short-cut Method

When the number of observations are large, the arithmetic mean can be calculated by using short-cut method or assumed mean method. When deviations are taken from an assumed mean, the following formula is used:

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

Where, d = Deviations of the items from the assumed mean, *i.e.*, $X - A$

A = Assumed Mean.

Steps for Calculation

(*i*) Any one of the items in the series is taken as assumed mean A.

(*ii*) Take the deviations of the items from the assumed mean, *i.e.*, $X - A$ and denote these deviations by 'd'.

(*iii*) Obtain the sum of these deviations, *i.e.*, Σd.

(*iv*) Substitute the values of A, Σd and N in the above formula. The result will give the value of arithmetic mean.

Example 2. The pocket allowances (in Rs.) of ten students are given below:

15, 20, 30, 22, 25, 18, 40, 50, 55 and 65

Calculate the arithmetic mean by taking 40 as assumed mean.

Solution: **Calculation of Arithmetic Mean**

Students	Pocket Allowances (X)	$A = 40$ $d = X - 40$
1	15	$15 - 40 = -25$
2	20	$20 - 40 = -20$
3	30	$30 - 40 = -10$
4	22	$22 - 40 = -18$
5	25	$25 - 40 = -15$
6	18	$18 - 40 = -22$
7	40 = A	$40 - 40 = 0$
8	50	$50 - 40 = +10$
9	55	$55 - 40 = +15$
10	65	$65 - 40 = +25$
$N = 10$		$\Sigma d = -60$

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

Substituting to values, we get

$$\overline{X} = 40 + \frac{(-60)}{10} = 40 - 6 = 34$$

Hence, the average pocket allowance is Rs. 34.

o Discrete Series

For calculating arithmetic mean in discrete series, the following two methods may be used:

(*i*) **Direct method**

(*ii*) **Short-cut method.**

▶ (*i*) Direct Method

When direct method is used, the following formula is used:

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

Where, f = frequency, X = values of the variable, N = total number of observations, *i.e.*, Σf.

Steps for Calculation

(*i*) Multiply the frequency of each item with the values of variable and obtain total ΣfX.

(*ii*) Find the sum of frequencies, *i.e.*, Σf or N.

(*iii*) Divide the total obtained (ΣfX) by the number of observations N or Σf. The result would be the required arithmetic mean.

Example 3. Calculate the arithmetic mean from the following data:

Wages (Rs.):	10	20	30	40	50
No. of workers:	4	5	3	2	5

Solution: Denoting wages by X and number of workers by f.

Calculation of Arithmetic Mean

Wages (X)	**No. of workers (f)**	fX
10	4	40
20	5	100
30	3	90
40	2	80
50	5	250
	N or $\Sigma f = 19$	$\Sigma fX = 560$

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{560}{19} = 29.47$$

Thus, the mean wage is Rs. 29.47.

(ii) Short-cut Method

When this method is used, the formula for calculating arithmetic mean is:

$$\bar{X} = A + \frac{\Sigma fd}{N}$$

Where, A = Assumed mean; $d = X - A$; N = Total number of observations, *i.e.*, Σf

Steps for Calculation

(*i*) Any one of the item in the series is taken as assumed mean A.

(*ii*) Take the deviations of the items from the assumed mean, *i.e.*, $X - A$ and denote these deviations by d.

(*iii*) Multiply these deviations with the respective frequency and obtain the total Σfd.

(*iv*) Divide the total obtained (Σfd) by the total frequency Σf or total number of observations N.

Example 4. From the following data of the wages obtained by 19 workers of a factory, calculate the arithmetic mean by using short-cut method:

Wages (Rs.):	10	20	30	40	50
No. of workers:	4	5	3	2	5

Solution: **Calculation of Arithmetic Mean**

Wages (Rs.) (X)	No. of workers (f)	$A = 30$ $d = X - 30$	fd
10	4	$10 - 30 = -20$	-80
20	5	$20 - 30 = -10$	-50
30 A	3	$30 - 30 = 0$	0
40	2	$40 - 30 = +10$	$+20$
50	5	$50 - 30 = +20$	$+100$
	Σf or $N = 19$		$\Sigma fd = -10$

$$\bar{X} = A + \frac{\Sigma fd}{N} = 30 - \frac{10}{19} = 30 - 0.53 = 29.47$$

Thus, the mean wage is 29.47.

Continuous Series

In continuous series, arithmetic mean can be calculated by using any one of the method:

(*i*) **Direct Method**

(*ii*) **Short-cut Method**

(*iii*) **Step deviation Method.**

(i) Direct Method:

When direct method is used, we apply the following formula to calculate arithmetic mean:

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

Where, m = mid-point of various classes, f = frequency of each class, N = total frequency.

Steps for Calculation

(*i*) Obtain the mid-value of each class and denote it by m.

(*ii*) Multiply each mid-value by the corresponding frequency and obtain the total Σfm.

(*iii*) Divide the total obtained (Σfm) by the sum of frequencies, *i.e.*, N.

Example 5. Calculate the arithmetic mean from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	20	24	40	36	20

Solution: **Calculation of Arithmetic Mean**

Marks	f	**Mid-value** (m)	fm
0—10	20	$\frac{0+10}{2}=5$	100
10—20	24	$\frac{10+20}{2}=15$	360
20—30	40	$\frac{20+30}{2}=25$	1,000
30—40	36	$\frac{30+40}{2}=35$	1,260
40—50	20	$\frac{40+50}{2}=45$	900
	Σf or $N=140$		$\Sigma fm=3620$

$$\overline{X}=\frac{\Sigma fm}{N}=\frac{3620}{140}=25.85$$

▶ **(*ii*) Short-cut Method**

When this method is used, arithmetic mean is computed by applying the following formula:

$$\overline{X}=A+\frac{\Sigma fd}{N}$$

Where, A = assumed mean; d = deviations of mid-value from assumed mean, *i.e.*, $m-A$;

N = total number of observations, *i.e.*, Σf.

Steps for Calculation

(*i*) Find the mid-value of each class and denote it by m.

(*ii*) Take any mid-value as assumed mean A.

(*iii*) Take deviations of the mid-value (m) from the assumed mean $m-A$ and denote it by d.

(*iv*) Multiply the respective frequencies of each class by these deviations and obtain the total Σfd.

(*v*) Divide the total obtained (Σfd) by the total frequency Σf or total number of observations.

Example 6. Calculate the arithmetic mean by the short-cut method from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	20	24	40	36	20

Solution:

Calculation of Arithmetic Mean

Marks	No. of students (f)	Mid-value (m)	$A = 25$ $d = m - A$	fd
0—10	20	5	–20	– 400
10—20	24	15	–10	– 240
20—30	40	25 = A	0	0
30—40	36	35	+10	+360
40—50	20	45	+20	+400
	N = 140			$\Sigma fd = 120$

$$\overline{X} = A + \frac{\Sigma fd}{N} = 25 + \frac{120}{140} = 25 + 0.85 = 25.85$$

Thus, the mean marks = 25.85

(*iii*) Step Deviation Method

In case of continuous series with class intervals of equal magnitude, the arithmetic mea computed by applying the following formula:

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

Where, $d' = \frac{m - A}{i}$; m = mid-value of the class; i = common magnitude of the class interv

A = assumed mean.

Note: Step deviation method is most commonly used in case of continuous series.

Steps for Calculation

(*i*) Find the mid-value of each class and denote it by m.

(*ii*) Take any mid-value as assumed mean A.

(*iii*) Take deviations of the mid-value (m) from the assumed mean $m - A$ and denote it by

(*iv*) Compute step deviations d'. These are obtained by dividing the deviations by magnitude of class intervals, *i.e.*, $d' = d / i$.

(*v*) Multiply the respective frequencies of each class by these deviations (d') and obtain total $\Sigma fd'$.

(*vi*) Divide the total obtained ($\Sigma fd'$) by the total frequency Σf or N and then multiply by the formula for getting arithmetic mean.

Example 7. From the following data, compute arithmetic mean by step deviation method:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	20	24	40	36	20

Solution:

Calculation of Arithmetic Mean

Marks	**Mid-value (m)**	f	**A = 25** $d = m - 25$	$d' = d/10$	fd'
0—10	5	20	–20	–2	–40
10—20	15	24	–10	–1	–24
20—30	25 = A	40	0	0	0
30—40	35	36	+10	+1	+36
40—50	45	20	+20	+2	+40
		N = 140			$\Sigma fd' = 12$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 25 + \frac{12}{140} \times 10$$

$$= 25 + 0.85 = 25.85$$

Inclusive Class Intervals

While calculating mean in a continuous series with inclusive class intervals, it is not necessary to convert the series into exclusive class intervals by adjusting class limits because the mid-values remain the same whether or not the adjustment is made:

Example 8. Calculate the arithmetic mean from the following data:

Size:	20—29	30—39	40—49	50—59	60—69
Frequency:	10	8	6	4	2

Solution:

Calculation of Arithmetic Mean

Size	**Mid-value (m)**	f	**A = 44.5** $d = m - A$	$d' = d/10$	fd'
20—29	24.5	10	– 20	–2	–20
29—39	34.5	8	– 10	–1	–8
40—49	44.5 = A	6	0	0	0
50—59	54.5	4	+10	+1	+4
60—69	64.5	2	+20	+2	+4
		N = 30			$\Sigma fd' = -20$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 44.5 + \frac{-20}{30} \times 10$$

$$= 44.5 - \frac{20}{3} = 44.5 - 6.67 = 37.83$$

Cumulative Frequency Series

While calculating mean in a cumulative frequency series, it is necessary to convert the series into a simple frequency series, and only after that the arithmetic mean is calculated.

Example 9. Calculate the arithmetic mean from the following data:

Marks	No. of students
Less than 10	5
Less than 20	17
Less than 30	31
Less than 40	41
Less than 50	49

Solution: Since, cumulative frequencies are given, first we find the simple frequencies:

Calculation of Arithmetic Mean

Marks	No. of students (f)	Mid-value (m)	$A = 25$ $d = m - 25$	$d' = d/10$	fd'
0—10	5	5	−20	−2	−10
10—20	$17-5=12$	15	−10	−1	−12
20—30	$31-17=14$	$25=A$	0	0	0
30—40	$41-31=10$	35	+10	+1	+10
40—50	$49-41=8$	45	+20	+2	+16
	$N=49$				$\Sigma fd'=+4$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 25 + \frac{4}{49} \times 10 = 25 + 0.81 = 25.81$$

Example 10. Calculate the arithmetic mean from the following data:

Marks	No. of students
More than 0	30
More than 2	28
More than 4	24
More than 6	18
More than 8	10

Solution: Since, cumulative frequencies are given, first we find the simple frequencies.

Calculation of Arithmetic Mean

Marks	No. of students (f)	Mid-value (m)	$A = 5$ $d = m - 5$	$d' = d/2$	fd'
0—2	$30-28=2$	1	−4	−2	−4
2—4	$28-24=4$	3	−2	−1	−4
4—6	$24-18=6$	$5=A$	0	0	0
6—8	$18-10=8$	7	+2	+1	+8
8—10	10	9	+4	+2	+20
	$N=30$				$\Sigma fd'=20$

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times i = 5 + \frac{20}{30} \times 2$$

$$= 5 + 1.33 = 6.33$$

Missing Frequency

Example 11. From the following data, calculate the missing value when its mean is 115.86:

Wages (Rs.):	110	112	113	117	—	125	128	130
No. of workers:	25	17	13	15	14	8	6	2

Solution: Let the missing item be denoted by a

Calculation of Missing Value

Wages (Rs.) (X)	No. of workers (f)	fX
110	25	2750
112	17	1904
113	13	1469
117	15	1755
a	14	$14a$
125	8	1000
128	6	768
130	2	260
	$N = 100$	$\Sigma fX = 9906 + 14a$

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

or $$115.86 = \frac{9906 + 14a}{100}$$

or $$115.86 \times 100 = 9906 + 14a$$

or $$14a = 1680$$

or $$a = \frac{1680}{14}$$

$$a = 120$$

Hence, the missing value is 120.

Example 12. Find the missing frequency from the following data for which the mean is 52:

Marks:	10—20	20—30	30—40	40—50	50—60	60—70	70—80
No. of students:	5	3	4	—	2	6	13

Solution: Let the missing frequency be denoted by f

Calculation of Missing Frequency

Marks	No. of students (f)	Mid-value (m)	fm
10—20	5	15	75
20—30	3	25	75
30—40	4	35	140
40—50	f	45	$45f$
50—60	2	55	110
60—70	6	65	390
70—80	13	75	975
	$N = 33 + f$		$\Sigma fm = 1765 + 45f$

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

$\Rightarrow$
$$52 = \frac{1765 + 45f}{33 + f}$$

or $52(33 + f) = 1765 + 45f$

or $1716 + 52f = 1765 + 45f$

or $7f = 49$

$$f = \frac{49}{7} = 7$$

Thus, the missing frequency is 7.

IMPORTANT TYPICAL EXAMPLES

Example 13. The following are the monthly salaries in rupees of 20 employees of a firm:

130	62	145	118	125	76	151	142	110	98
65	116	100	103	71	85	80	122	132	95

The firm gives bonus of Rs. 10, 15, 20, 25 and 30 for individuals in the respectiv salary groups exceeding Rs. 60 but not exceeding Rs. 80, exceeding Rs. 80 but nc exceeding Rs.100 and so on upto exceeding Rs. 140 but not exceeding Rs. 160. Fin the average bonus paid per employee.

Solution:

Monthly salary	Tally Bars	No. of workers (f)	Bonus (X)	fX
Exceeding 60 but not exceeding 80	~~IIII~~	5	10	50
Exceeding 80 but not exceeding 100	IIII	4	15	60
Exceeding 100 but not exceeding 120	IIII	4	20	80
Exceeding 120 but not exceeding 140	IIII	4	25	100
Exceeding 140 but not exceeding 160	III	3	30	90
		Σf or $N = 20$		$\Sigma fX = 380$

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{380}{20} = \text{Rs. } 19$$

$\therefore$ Average bonus = Rs. 19.

Example 14. The sum of the deviations of a certain number of items measured from 2.5 is 50 and from 3.5 is – 50. Find N and $\overline{X}$.

Solution: Now, when assumed mean (A) is 2.5, $\Sigma d = 50$

$$\therefore \quad \overline{X} = A + \frac{\Sigma d}{N} = 2.5 + \frac{50}{N} \quad \text{...(i)}$$

When assumed mean (A) is 3.5, $\Sigma d = -50$

$$\therefore \quad \overline{X} = A + \frac{\Sigma d}{N} = 3.5 - \frac{50}{N} \quad \text{...(ii)}$$

From (i) and (ii),

$$2.5 + \frac{50}{N} = 3.5 - \frac{50}{N}$$

$$\frac{100}{N} = 1 \Rightarrow N = 100$$

From (i), $\overline{X} = 2.5 + \frac{50}{100} = \frac{250+50}{100}$

$$= \frac{300}{100} = 3$$

$\therefore \quad \overline{X} = 3, N = 100$

EXERCISE 5.1

1. Following are the marks obtained by 8 students. Calculate the arithmetic mean:

Marks:	15	18	16	45	32	40	30	28

[**Ans.** $\overline{X} = 28$ **marks**]

2. Following are the marks obtained by 25 students in Economics. Find out the mean marks by using direct and short-cut method:

Marks:	10	20	30	40	50	60
No. of students:	5	2	3	8	4	3

[**Ans.** $\overline{X} = 35.2$]

3. Calculate the arithmetic mean from the following data:

Class interval:	20—25	25—30	30—35	35—40	40—45	45—50	50—55
Frequency:	10	12	8	20	11	4	5

[**Ans.** $\overline{X} = 35.5$]

4. Calculate the average marks from the following data by using Short-cut method:

Marks:	6—10	11—15	16—20	21—25	26—30
No. of students:	20	30	50	40	10

[**Ans.** $\overline{X} = 17.67$]

5. Find out the arithmetic mean from the following data:

Marks (less than):	5	10	15	20	25	30
No. of students:	10	23	30	54	69	80

[**Ans.** $\overline{X} = 15.87$]

6. Find out the arithmetic mean from the following data:

Bonus (more than):	0	4	8	12
No. of workers:	15	11	3	1

[**Ans.** $\overline{X} = 6$]

7. Calculate the arithmetic mean from the following:

X:	–40 to –30	–30 to –20	–20 to –10	–10 to 0	0 to 10	10 to 20	20 to 30
f:	10	28	30	42	65	180	10

[**Ans.** $\overline{X} = 4.287$]

8. Find out the missing frequencies of the following series if the A.M. is 35 and total number of items is 100:

Class interval:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
Frequency:	5	10	?	4	20	3	?

[**Ans.** $f_1 = 41$ app.; $f_2 = 17$ app.]

9. The following are the monthly salaries in rupees of 30 employees in a firm:

140	139	120	114	100	88	62	77	99	103
108	129	144	148	134	63	69	148	132	118
142	116	123	104	95	80	85	106	123	133

The firm gave bonus of Rs. 10, 15, 20, 25, 30 and 35 for individuals in the respective salary groups: exceeding 60 but not exceeding 75, exceeding 75 but not exceeding 90 and so on upto exceeding 135 but not exceeding 150. Find out the average bonus paid.

[**Ans.** Average bonus = Rs. 24.5]

10. The sales of a balloon seller on seven days of a week are given as below:

Days:	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Sales (in Rs.):	100	150	125	140	160	200	250

If the profit is 20% of sales, find his average profit per day.

[**Hint:** Calculate profit per day $= \frac{20}{100} \times$ Sales] [**Ans.** Rs. 32.14]

11. The sum of the deviations of a certain number of items measured from 4 is 72 and the sum of the deviations of the items from 7 is – 3. Find the number of items and their mean.

[**Ans.** $N = 25$, $\overline{X} = 6.88$]

COMBINED ARITHMETIC MEAN

If we have the arithmetic mean and the number of items of two or more than two related sub-groups, we can calculate the combined arithmetic mean of the whole group by applying the following formula:

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

Where, $\bar{X}_{12}$ = Combined arithmetic mean of the two groups.

$\bar{X}_1$ = A.M. of the first group; $\bar{X}_2$= A.M. of the second group.

The above formula can be extended to calculate the arithmetic mean of three or more groups. For example, combined arithmetic mean of three groups is given by:

$$\bar{X}_{123} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3}{N_1 + N_2 + N_3}$$

The following examples illustrate the application of the above formula:

Example 15. The mean height of 25 male workers in a factory is 61 cms. and the mean height of 35 female workers in the same factory is 58 cm. Find the combined mean height of 60 workers in the factory.

Solution: Given: $N_1 = 25,\ \bar{X}_1 = 61,\ N_2 = 35,\ \bar{X}_2 = 58$

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

$$= \frac{(25 \times 61) + (35 \times 58)}{25 + 35} = \frac{1525 + 2030}{60} = \frac{3555}{60} = 59.25$$

$\therefore \quad \bar{X}_{12} = 59.25$

Thus, the combined mean height of 60 workers is 59.25 cm.

Example 16. The mean wage of 100 workers in a factory, running two shifts of 60 and 40 workers respectively is Rs. 38. The mean wage of 60 labourers working in the morning shift is Rs. 40. Find the mean wage of 40 workers working in the evening shift.

Solution: Given: $N = 100, \bar{X}_{12} = 38, N_1 = 60,\ \bar{X}_1 = 40;\ N_2 = 40,\ \bar{X}_2 = ?$

Using the formula of combined arithmetic mean, we have:

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

$$38 = \frac{60 \times 40 + 40\,\bar{X}_2}{100}$$

or $\quad 3800 = 2400 + 40\,\bar{X}_2$

or $\quad 1400 = 40\,\bar{X}_2$

$\therefore \quad \bar{X}_2 = \frac{1400}{40} = 35$

$\therefore \quad \bar{X}_2 = 35$

IMPORTANT TYPICAL EXAMPLES

Example 17. The mean monthly salary paid to all employees in a certain company was Rs.600. The mean monthly salaries paid to male and female employees were Rs. 620 and Rs. 520 respectively. Find the percentage of male to female employees in the company.

Solution: Given: $\overline{X}_{12} = 600,\ \overline{X}_1 = 620,\ \overline{X}_2 = 520$

Let N_1= Male Employees and N_2= Female Employees.

Using the formula of combined arithmetic mean, we have

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

$$600 = \frac{N_1(620) + N_2(520)}{N_1 + N_2}$$

or $600(N_1 + N_2) = 620\,N_1 + 520\,N_2$

or $600N_1 + 600N_2 = 620\,N_1 + 520\,N_2$

$\therefore$ $80N_2 = 20\,N_1$

or $\dfrac{N_1}{N_2} = \dfrac{80}{20} = \dfrac{4}{1}$

or $N_1 : N_2 = 4 : 1$

Hence, the percentage of male employees $= \dfrac{4}{4+1} \times 100 = 80\%$ and the percentage of female employees $= \dfrac{1}{4+1} \times 100 = 20\%$

Aliter: The problem can also be solved by using the following method:

Let x be the percentage of male in the combined group. Therefore, the percentage of women = $100 - x$.

We are given that $\overline{X}_1$(men) = Rs. 620 and $\overline{X}_2$(women) = Rs. 520

Also $\overline{X}_{12}$(combined mean) = 600

$$\therefore \quad 600 = \frac{620x + 520(100 - x)}{100} = \frac{620x + 52000 - 520x}{100}$$

$\Rightarrow$ $60{,}000 - 52{,}000 = 100x$

$\Rightarrow$ $x = \dfrac{8000}{100} = 80\%$

Thus, there are 80% men and 20% women in this group.

Example 18. A bookseller has 150 books of Economics and Accountancy. The average price of these books is Rs. 40 per book. Average price of books on Economics is Rs. 43 and that of Accountancy is Rs. 35. Find the number of books on Economics with the seller.

Solution: Let, N_1 = Economics, N_2 = Accountancy

$N_1 + N_2 = 150$

$\Rightarrow \quad N_2 = 150 - N_1$

$\overline{X}_{12} = 40, \overline{X}_1 = 43, \ \overline{X}_2 = 35$

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

$$40 = \frac{43\,N_1 + 35(150 - N_1)}{150}$$

$$40 = \frac{43\,N_1 + 5250 - 35\,N_1}{150}$$

$$40 = \frac{8\,N_1 + 5250}{150}$$

$$6000 = 8N_1 + 5250$$

$$\therefore \quad 8N_1 = 750$$

$$\Rightarrow \quad N_1 = \frac{750}{8} = 93.75 \cong 94$$

Thus, books on Economics with the seller are 94.

EXERCISE 5.2

1. If the average marks obtained by the students of sections A and B of B.Com. class in a college are 40 and 30 respectively whereas the number of students in sections A and B are 60 and 40 respectively. Find out the combined mean marks. **[Ans.** $\overline{X}_{12}$ = 36 marks]
2. The mean monthly salary paid to 77 employees in a company was Rs. 78. The mean salary of 32 of them was Rs. 45 and of the other 25 was Rs. 82. What was the mean salary of the remaining? **[Ans.** $\overline{X}_3 = 125.8$]
3. The average rainfall for a week excluding Sunday was 10 cm. Due to heavy rainfall on Sunday, the average for the week rose to 15 cm. How much rainfall was on Sunday?
[Hint: $\overline{X}_{12} = 15, \ \overline{X}_1 = 10, N_1 = 6, N_2 = 1, \overline{X}_2 = ?$] **[Ans.** 45 cm]
4. The mean annual salary paid to all employees of a company was Rs. 5000. The mean annual salaries paid to male and female employees were Rs. 5200 and Rs. 4200 respectively. Determine the percentage of male and female employed by the company. **[Ans.** Male = 80%; Female = 20%]
5. The average marks of 39 students of a class is 50. The marks obtained by 40th student are 39 more than the average marks of all the 40 students. Find the mean marks of all the 40 students.
[Hint: $\overline{X} + 39 + 39 \times 50 = 40\overline{X}$] **[Ans.** $\overline{X} = 51$]
6. The mean of 99 items is 55. The value of 100th item is 99 more than the mean of 100 items. What is the value of 100th item? **[Ans.** $X_{100} = 155$]

7. 50 students took a test. The result of those who passed the test is given below:

Marks:	4	5	6	7	8	9
Students:	8	10	9	6	4	3

If the average marks for all the 50 students were 5.16, find out the average marks of the students who failed. **[Ans.** $\overline{X}_F = 2.12$]

◘ CORRECTING INCORRECT VALUES OF MEAN

Sometimes, due to mistake in copying, certain items may be wrongly taken in calculating the arithmetic mean. The problem now arises is to find out correct mean. In this case, without calculating the arithmetic mean from the beginning, we can directly calculate the arithmetic mean. The process is very simple. Wrong values are deducted and correct values are added to ΣX. Then correct ΣX is divided by the number of observations. The result, thus, obtained will give us the correct mean.

The following examples would illustrate the procedure:

Example 19. The mean of 100 items is 80. By mistake one item is misread as 92 instead of 29. Find the correct mean.

Solution: We are given, $N = 100,\ \overline{X} = 80$

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

$$\therefore\ \Sigma X \text{ (Incorrected)} = 100 \times 80 = 8000$$

$$\text{Corrected } \Sigma X = 8000 + \text{Correct item} - \text{Incorrect item}$$

$$= 8000 + 29 - 92$$

$$= 7937$$

$$\therefore\ \text{Corrected Mean} = \frac{\text{Corrected } \Sigma X}{N} = \frac{7937}{100} = 79.37$$

Example 20. The mean of 5 observations is 7. Later on, it was found that two observations 4 and 8 were wrongly taken instead of 5 and 9. Find the correct mean.

Solution: We are given, $N = 5,\ \overline{X} = 7$

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

$$\therefore\ \Sigma X \text{ (Incorrected)} = 5 \times 7 = 35$$

$$\text{Corrected } \Sigma X = 35 + \text{Correct item} - \text{Incorrect item}$$

$$= 35 + 5 + 9 - 4 - 8$$

$$= 37$$

$$\therefore\ \text{Corrected } \overline{X} = \frac{\text{Corrected } \Sigma X}{N} = \frac{37}{5} = 7.4$$

EXERCISE 5.3

1. The mean marks of 100 students were found to be 40. Later on it was discovered that a score of 53 was misread as 83. Find the correct mean. [Ans. $\bar{X} = 39.7$]
2. Mean of 100 observations is found to be 40. If at the time of computation two items are wrongly taken as 30 and 27 instead of 3 and 72. Find the correct mean. [Ans. $\bar{X} = 40.18$]
3. The mean of 20 observations is 6.21. Later on it was discovered that two items +5 and +3 were taken as –5 and –3. Find the correct mean. [Ans. $\bar{X} = 7.01$]
4. The average daily price of share of a company from Monday to Friday was Rs. 130. If the highest and lowest price during the week were Rs. 200 and Rs. 100 respectively, find the average daily price when the highest and lowest price are not included. [Ans. $\bar{X} = 116.67$]

MATHEMATICAL PROPERTIES OF ARITHMETIC MEAN

The following are some of the important mathematical properties of the arithmetic mean:

(1) The sum of the deviations of the items from arithmetic mean is always zero. Symbolically,

$$\Sigma(X - \bar{X}) = 0$$

This property is verified from the following example:

X	$\bar{X} = 30$ $X - \bar{X}$
10	$10 - 30 = -20$
20	$20 - 30 = -10$
30	$30 - 30 = 0$
40	$40 - 30 = +10$
50	$50 - 30 = +20$
$\Sigma X = 150$ $N = 5$ $\bar{X} = \frac{150}{5} = 30$	$\Sigma(X - \bar{X}) = 0$

It is clear that when the deviations from the actual mean, *i.e.*, 30 is taken, its sum comes to be zero.

(2) The sum of the squared deviations of the items from arithmetic mean is minimum. That is, it is less than the sum of the squared deviations of the items from any other value. This property is clear from the following example:

X	$\overline{X}=5$ $X-\overline{X}$	$(X-\overline{X})^2$	$A=4$ $X-4$	$(X-4)^2$
3	–2	4	–1	1
4	–1	1	0	0
5	0	0	+1	1
6	+1	1	+2	4
7	+2	4	+3	9
$\Sigma X=25$ $N=5$ $\overline{X}=5$		$\Sigma(X-\overline{X})^2=10$		$\Sigma(X-4)^2=15$

It is clear that the sum of squared deviations taken from the arithmetic mean is 10 whereas the sum of the squared deviations taken from the assumed mean 4 is 15. Therefore,

$$\Sigma(X-\overline{X})^2<\Sigma(X-A)^2$$

(3) If we have the arithmetic mean and number of items of two groups, then the combined arithmetic mean of these groups can be calculated by using the following formula:

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

The above formula can be extended to calculate the combined arithmetic mean of three or more groups.

(4) If each item of a series is increased, decreased, multiplied or divided by some constant, then A.M. is also increases, decreases, multiplied or divided by the same constant. The following example would clarify this property:

X	$X+2$	$X-2$	$X\times 2$	$X\div 2$
10	12	8	20	5
20	22	18	40	10
30	32	28	60	15
40	42	38	80	20
50	52	48	100	25
$\Sigma X=150$ $N=5$ $\overline{X}=30$	$\Sigma X=160$ $N=5$ $\overline{X}=32$	$\Sigma X=140$ $N=5$ $\overline{X}=28$	$\Sigma X=300$ $N=5$ $\overline{X}=60$	$\Sigma X=75$ $N=5$ $\overline{X}=15$

The above example shows that if 2 is added or subtracted to different items of a series or the items of a series are multiplied or divided by 2, the A.M. will also be affected accordingly.

(5) The product of the arithmetic mean and number of items on which mean is based is equal to the sum of all given items. That is,

As $\overline{X}=\dfrac{\Sigma X}{N}\Rightarrow\Sigma X=N\cdot\overline{X}$

(6) If each item of the original series is replaced by the actual mean, then the sum of the substitutions will be equal to the sum of the individual items.

Some Examples Based on Properties of Arithmetic Mean

Example 21. The arithmetic mean ($\overline{X}$) of a series is 15. What is the new mean $\overline{X}$ if each item is increased by 5 and then divided by 3?

Solution: When each item is increased by 5, new $\overline{X} = 15+5 = 20$

When each item is divided by 3, new $\overline{X} = \frac{15}{3} = 5$

Example 22. If the average salary of a firm is 400 and the number of workers is 60, find the total salary bill of the firm.

Solution: Given: $\overline{X} = 400, N = 60$

As $\overline{X} = \frac{\Sigma X}{N} \Rightarrow \Sigma X = N \cdot \overline{X}$

Total Salary Bill $= (\Sigma X) = 60 \times 400 =$ Rs. 24,000

Example 23. If the arithmetic mean of series is 28, what will be resultant mean if each item of a series is increased by 3, decreased by 5 or divided by 4 or is multiplied by 10?

Solution: When each item is increased by 3, the mean $\overline{X} = 28+3 = 31$

When each item is decreased by 5, the mean $\overline{X} = 28-5 = 23$

When each item is multiplied by 10, the mean $\overline{X} = 28 \times 10 = 280$

When each item is divided by 4, the mean $\overline{X} = \frac{28}{4} = 7$

Example 24. Prove that $\Sigma(X - \overline{X}) = 0$

Solution: $\Sigma(X - \overline{X}) = \Sigma X - \Sigma \overline{X} = \Sigma X - N\overline{X}$

$= N\overline{X} - N\overline{X} = 0$ [As $\overline{X} = \frac{\Sigma X}{N} \Rightarrow \Sigma X = N\overline{X}$]

Hence, proved.

Merits and Demerits of Arithmetic Mean

Merits:

(1) It is easy to calculate and simple to understand.

(2) It is based on all observations.

(3) It is capable of further algebraic treatment.

(4) It is rigidly defined.

(5) It is a calculated value and not based on the position of the series.

Demerits:

(1) It is highly affected by extreme values. It is not advised if the series has a few extreme items.

(2) It cannot be calculated in open ended series.

(3) Sometimes, arithmetic mean gives misleading and surprising results such as average number of children born per married couple is 2.3, etc.

(4) It cannot be ascertained graphically.

(5) It cannot be determined in a situation when any value is missing.

Weighted Arithmetic Mean

Simple Arithmetic Mean, as discussed above gives equal importance (or weights) to each item of the series. But there can be some cases where all the items of a series are not of equal importance. In case of Science, Economics or Technology, each subject has its own importance. We have to provide different weights according to their importance and weighted arithmetic mean is used as an average in such cases. The following formula is used to calculate the weighted arithmetic mean:

$$\bar{X}_W = \frac{\Sigma WX}{\Sigma W}, \quad \text{Where, } \bar{X}_W = \text{Weighted A.M.}$$

$$\Sigma W = \text{Sum of weights}$$

$$X = \text{Variables}$$

Note: In weighted A.M., W is taken instead of f.

Steps for Calculation

(*i*) Multiply the values of the items (X) by the weights (W) and obtain the total, *i.e.*, ΣWX.

(*ii*) Now, divide this total, *i.e.*, ΣWX by the sum total of weights, *i.e.*, ΣW. The resultant value would give the weighted A.M.

The following examples would illustrate the computation of weighted A.M.:

Example 25. Calculate weighted mean from the following data:

Items:	81	76	74	58	70	73
Weight:	2	3	6	7	3	7

Solution: **Calculation of Weighted Mean**

X	W	WX
81	2	162
76	3	228
74	6	444
58	7	406
70	3	210
73	7	511
	$\Sigma W = 28$	$\Sigma WX = 1961$

$$\therefore \quad \bar{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{1961}{28} = 70.04$$

Example 26. A student obtained 60 marks in English, 75 in Hindi, 63 in Mathematics, 59 in Economics and 55 in Statistics. Calculate the weighted mean of the marks if weights are respectively 2, 1, 5, 5 and 3.

Solution: **Calculation of Weighted Mean**

Marks (X)	Weights (W)	WX
60	2	120
75	1	75
63	5	315
59	5	295
55	3	165
	$\Sigma W = 16$	$\Sigma WX = 970$

$$\overline{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{970}{16} = 60.63$$

Example 27. A train runs 25 miles at a speed of 30 *mph*, another 50 miles at a speed of 40 *mph*, then due to repairs of the track travels for 6 minutes at a speed of 10 *mph* and finally covers the remaining distance of 24 miles at a speed of 24 *mph*. What is the average speed in miles per hour?

Solution: Time taken in covering 25 miles at a speed of 30 *mph* = 50 minutes

Time taken in covering 50 miles at a speed of 40 *mph* = 75 minutes

Distance covered in 6 minutes at a speed of 10 *mph* = 1 mile

Time taken in covering 24 miles at speed of 24 *mph* = 60 minutes.

Therefore, taking the time taken as weights we have the weighted mean as

Speed in *mph* (X)	Time Taken (W)	WX
30	50	1,500
40	75	3,000
10	6	60
24	60	1,440
	$\Sigma W = 191$	$\Sigma WX = 6{,}000$

$$\therefore \quad \text{Average speed} = \frac{6{,}000}{191} = 31.41 \; mph$$

Example 28. Comment on the performance of the students of the three universities given below using simple and weighted averages:

University	Bombay		Calcutta		Madras	
Course of Study	Pass %	No. of students (in hundreds)	Pass %	No. of students (in hundreds)	Pass %	No. of students (in hundreds)
M.A.	71	3	82	2	81	2
M.Com.	83	4	76	3	76	3.5
B.A	73	5	73	6	74	4.5
B.Com.	74	2	76	7	58	2
B.Sc.	65	3	65	3	70	7
M.Sc.	66	3	60	7	73	2

Solution: **Calculation of Simple and Weighted Arithmetic Means**

University	Bombay			Calcutta			Madras		
Course of Study	Pass %	No.of students (in hundreds)		Pass %	No.of students (in hundreds)		Pass %	No.of students (in hundreds)	
	X	W	WX	X	W	WX	X	W	WX
M.A.	71	3	213	82	2	164	81	2	162
M.Com.	83	4	332	76	3	228	76	3.5	266
B.A	73	5	365	73	6	438	74	4.5	333
B.Com.	74	2	148	76	7	532	58	2	116
B.Sc.	65	3	195	65	3	195	70	7	490
M.Sc.	66	3	198	60	7	420	73	2	146
	ΣX =432	ΣW =20	ΣWX =1,451	ΣX =432	ΣW =28	ΣWX =1,977	ΣX =432	ΣW =21	ΣWX =1,513

Simple and Weighted Arithmetic Means

Bombay : $\overline{X} = \frac{\Sigma X}{N} = \frac{432}{6} = 72;\ \overline{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{1,451}{20} = 72.55$

Calcutta : $\overline{X} = \frac{\Sigma X}{N} = \frac{432}{6} = 72;\ \overline{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{1,977}{28} = 70.61$

Madras : $\overline{X} = \frac{\Sigma X}{N} = \frac{432}{6} = 72;\ \overline{X}_W = \frac{\Sigma WX}{\Sigma W} = \frac{1,513}{21} = 72.05$

The simple arithmetic mean is the same for all the three universities, *i.e.*, 72 and hence, it may be concluded that the performance of students is alike. But this will be a wrong conclusion because what we should compare here is the weighted arithmetic mean. On comparing the weighted arithmetic means, we find that for Bombay the mean value is the highest and hence, we can say that in Bombay University the performance of students is best.

EXERCISE 5.4

1. A housewife uses 10 kg of Wheat, 20 kg of Fuel, 5 kg. of Sugar and 2 kg. of Oil. Price (per kg.) of these items are respectively Rs. 1.50, Rs. 0.50, Rs. 2.80 and Rs. 10. Taking quantities used as weights, find out weighted arithmetic average of the prices. [**Ans.** $\overline{X}_W$ = Rs. 1.59 per kg.]

2. Calculate the simple and weighted arithmetic mean price per tonne of coal purchased by an industry for the half year:

Months:	Jan.	Feb.	Mar.	Apr.	May	June
Price per tonne (Rs.)	42.50	51.25	50.00	52.00	44.25	54.00
Tonnes Purchased:	25	30	40	50	10	45

[**Hint:** Taking quantity purchased as weight] [**Ans.** $\overline{X} = 49$, $\overline{X}_W = 50.363$]

3. From the following results of Universities A and B, which is better?

Class	University A		University B	
	Appeared	Passed	Appeared	Passed
M.A.	100	90	240	200
M.Com.	60	45	200	160
B.A.	120	75	160	60
B.Com.	200	150	200	140
Total	480	360	800	560

[**Hint:** Taking number of students appeared as weights and pass percentage of each as X]

[**Ans.** $\overline{X}_w$(Univ. A) = 75%, $\overline{X}_w$(Univ. B) = 70%; University A is better]

(2) MEDIAN

Median is another important measure of central tendency. It is a positional average. Median is defined as the middle value of the series when arranged either in ascending or descending order. It is a value which divides the arranged series into two equal parts in such a way that the number of observations smaller than the median is equal to the number greater than it. Median is thus, a positional average. In the words of **Connor, "The median is that value of the variable which devides the group into two equal parts, one part comprising all values greater and the other values less than the median."** A point to be noted is that median is always determined by first arranging the series in an ascending or descending manner. Median is denoted by the symbol 'M'.

Calculation of Median

Individual Series

The formula used for calculating median in individual series is:

$$M = \textbf{Size of} \left(\frac{N+1}{2}\right) \textbf{th item.}$$

Where, M = median, N = total number of items in the series.

Steps for Calculation

(*i*) Arrange the data in ascending or descending order of the size.

(*ii*) Locate the median item by using the formula $\frac{N+1}{2}$.

(*iii*) The value or size of this item is the median.

Odd Number Series

If the number of items is odd, then the median is the middle value after the items are arranged in ascending or descending order of their magnitude.

Example 29. Calculate median from the following data:

22, 16, 18, 13, 15, 19, 17, 20, 23

Solution: The data is first arranged in ascending order

Sr. No.	Items (X)
1	13
2	15
3	16
4	17
5	18
6	19
7	20
8	22
9	23
$N = 9$	

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

Here, $N = 9$

$$\therefore \quad M = \text{Size of} \left(\frac{9+1}{2}\right) \text{th item.}$$

$$= \text{Size of 5th item} = 18$$

Hence, $M = 18$

Even Number Series

In case of even number of items, median is arithmetic median of two middle values after t items are arranged in ascending or descending order of their magnitude.

Example 30. Calculate median from the following data:

200, 217, 316, 264, 296, 282, 317, 299

Solution: The data is first arranged in ascending order

Sr. No.	Items (X)
1	200
2	217
3	264
4	282
5	296
6	299
7	316
8	317
$N = 8$	

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

Here, $N = 8$

$$\therefore \quad M = \text{Size of}\left(\frac{8+1}{2}\right)\text{th item.}$$

$$= \text{Size of 4.5th item}$$

$$= \frac{\text{Size of 4th items+ Size of 5th item}}{2}$$

$$\therefore \quad M = \frac{282+296}{2} = \frac{578}{2} = 289$$

Hence, $\quad M = 289.$

Example 31. Following are the marks obtained by a batch of 10 students in a certain class test in Statistics (X) and Accountancy (Y):

Roll No.	1	2	3	4	5	6	7	8	9	10
X:	63	64	62	32	30	60	47	46	35	28
Y:	68	66	35	42	26	85	44	80	33	72

In which subject is the level of knowledge of students higher?

Solution: **Calculation of Median**

X:	28	30	32	35	46	47	60	62	63	64
Y:	26	33	35	42	44	66	68	72	80	85

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

$$= \frac{10+1}{2}\text{th item} = \frac{11}{2}\text{th} = 5.5\text{th item.}$$

$$\text{Median (X)} = \text{Size of}\left(\frac{5\text{th}+6\text{th}}{2}\right)\text{item} = \frac{46+47}{2} = 46.5 \text{ marks.}$$

$$\text{Median (Y)} = \text{Size of}\left(\frac{5\text{th}+6\text{th}}{2}\right)\text{item} = \frac{44+66}{2} = 55 \text{ marks.}$$

Median marks of Accountancy (Y) are more than that of statistics (X). Therefore, level of knowledge of students is higher in Accountancy.

Discrete Series

The formula used for calculating median in discrete series is

$$M = \textbf{Size of}\left(\frac{N+1}{2}\right)\textbf{th item.}$$

▶ **Steps for Calculation**

(*i*) Arrange the data in ascending or descending order of size.

(*ii*) Then find the cumulative frequency column.

(*iii*) Apply the formula,

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

(*iv*) Now locate $\frac{N+1}{2}$th item in the cumulative frequency column. It is done by comparing $\frac{N+1}{2}$ with the cumulative frequency at each stage. The value of the variable is the value of the median.

Example 32. Calculate the median from the following data:

X:	10	12	14	16	18	20	22
f:	2	5	12	20	10	7	3

Solution:

Calculation of Median

X	f	c.f.
10	2	2
12	5	7
14	12	19
16	20	39 M
18	10	49
20	7	56
22	3	59
	$N = 59$	

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

$$= \text{Size of} \left(\frac{59+1}{2} \right) \text{th item.}$$

$$= \text{Size of 30th item.}$$

The value of 30th item lies against 39 whose value is 16.

Hence, $M = 16$.

⭘ Continuous Series

The formula used for calculating median in continuous series is

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

Where, l_1 = lower limit of the median class; $c.f.$ = cumulative frequency of the class preceeding the median class; f = frequency of the median class; i = size of class interval of the median class.

▶ Steps for Calculation

(*i*) Firstly, calculate cumulative frequency.

(*ii*) Then, find out the median size by using the formula: $\frac{N}{2}$

(*iii*) Determine the median class in which median lies.

(*iv*) Substitute the values in the above formula.

Example 33. Calculate median from the following data:

Marks:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40	40—45
No. of students:	6	12	17	30	10	10	8	5	2

Solution:

Marks	*f*	*c.f.*
0—5	6	6
5—10	12	18
10—15	17	35 *c.f.*
15—20	30 *f*	65 M
20—25	10	75
25—30	10	85
30—35	8	93
35—40	5	98
40—45	2	100
	N = 100	

Median item = Size of $\left(\frac{N}{2}\right)$th item = $\frac{100}{2}$th = 50th item.

50th item lies in class 15—20. Hence median class is 15–20.

Applying the formula,

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 15 + \frac{50 - 35}{30} \times 5$$

$$= 15 + \frac{15}{30} \times 5 = 15 + 2.5 = 17.5$$

Hence, $M = 17.5$.

Cumulative Frequency Series

Example 34. Calculate median from the following data:

Value	Frequency
Less than 10	4
Less than 20	16
Less than 30	40
Less than 40	76
Less than 50	96
Less than 60	112
Less than 70	120
Less than 80	125

Solution: Since we are given cumulative frequencies, firstly we find simple frequencies.

Calculation of Median

Value	f	$c.f.$
0—10	4	4
10—20	12	16
20—30	24	40
30—40	36	76
40—50	20	96
50—60	16	112
60—70	8	120
70—80	5	125
	N = 125	

$$\text{Median item} = \text{Size of}\left(\frac{N}{2}\right)\text{th item} = \frac{125}{2} = 62.5\text{th item.}$$

$\therefore$ Median lies in the class 30—40

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 30 + \frac{62.5 - 40}{36} \times 10$$

$$= 30 + \frac{22.5}{36} \times 10$$

$$= 30 + 6.25$$

$$= 36.25$$

Inclusive Series

Example 35. Calculate median from the following data:

Value:	1—10	11—20	21—30	31—40	41—50
Frequency:	4	12	20	9	5

Solution: Since we are given inclusive series, firstly we convert it into exclusive one by deducting 0.5 from the lower limits and adding 0.5 to the upper limits.

Calculation of Median

Value	f	$c.f.$
0.5—10.5	4	4
10.5—20.5	12	16
20.5—30.5	20	36
30.5—40.5	9	45
40.5—50.5	5	50
	$N = 50$	

$$\text{Median item} = \text{Size of} \left(\frac{N}{2}\right)\text{th item} = \frac{50}{2}\text{th} = 25\text{th item.}$$

$\therefore$ Median lies in the class 20.5—30.5.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20.5 + \frac{25-16}{20} \times 10 = 20.5 + \frac{9}{20} \times 10 = 20.5 + 4.5 = 25$$

$\therefore$ $M = 25.$

Unequal Class Interval

Example 36. Amend the following table and calculate the median from the amended table:

Size :	10—15	15—17.5	17.5—20	20—30	30—35	35—40	40 and onwards
f:	10	15	17	25	28	30	40

Solution: Since the class intervals are unequal, let us first convert it into a series with equal class intervals by adjusting the frequencies correspondingly.

Calculation of Median

Size	f	$c.f.$
10—20	10+15+17 = 42	42
20—30	25	67
30—40	28+30 = 58	125
40 and onwards	40	165
	$N = 165$	

$$\text{Median item} = \text{Size of}\left(\frac{N}{2}\right)\text{th item} = \frac{165}{2} = 82.50 \text{ item.}$$

$\therefore$ Median lies in class 30—40

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 30 + \frac{82.50 - 67}{58} \times 10$$

$$= 30 + 2.67 = 32.67$$

$$\therefore \quad M = 32.67$$

Descending Class Intervals

Example 37. Find the median from the distribution of marks obtained in Economics:

Marks:	30—35	25—30	20—25	15—20	10—15	5—10	0—5
Number of students:	4	8	12	16	10	6	4

Solution: **Ist Method.** Converting the decending order series into ascending order series.

Marks	f	$c.f.$
0—5	4	4
5—10	6	10
10—15	10	20 $c.f.$
15—20	16 f	36 M
20—25	12	48
25—30	8	56
30—35	4	60
	N=60	

$$\text{Median item} = \frac{N}{2} = \frac{60}{2} = 30\text{th item.}$$

Median lies in the class 15–20.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 15 + \frac{30 - 20}{16} \times 5$$

$$= 15 + \frac{50}{16} = 15 + 3.125 = 18.125$$

$$\therefore \quad \text{Median} = 18.125$$

2nd Method. Median can also be calculated by keeping the series in descending order. The formula used is:

$$M = l_2 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

Where, l_2 = upper limit of the median class.

Marks	f	$c.f.$
30—35	4	4
25—30	8	12
20—25	12	24 $c.f.$
15—20	16 f	40 M
10—15	10	50
5—10	6	56
0—5	4	60
	N=60	

$$\text{Median item} = \frac{N}{2} = \frac{60}{2} = 30\text{th item.}$$

Median lies in the class 15–20.

Applying the modified formula:

$$M = l_2 - \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 - \frac{30-24}{16} \times 5$$

$$= 20 - \frac{6}{16} \times 5 = 20 - \frac{30}{16} = 20 - 1.875 = 18.125$$

$\therefore$ Median = 18.125

Note: Students find that the answer of median will however, remain the same as is computed in ascending or descending order.

Mid-value Series

Example 38. Compute median from the following data:

Mid-value:	5	15	25	35	45	55	65	75
Frequency:	15	7	11	10	13	8	20	16

Solution: Since, we are given the mid-values, we should first find out the upper and lower limits of the various classes. As the mid-values are 5, 15, 25, 35, 45, 55, 65, 75, so, class size is 10. For determining limits of different classes, applying the formula.

$l_1 = m - \frac{i}{2}$ and $l_2 = m + \frac{i}{2}$, where, i = difference between two mid-values.

In this case, the upper and the lower limits of the first mid-value are:

$$l_1 = 5 - \frac{10}{2} = 0 \text{ and } l_2 = 5 + \frac{10}{2} = 10, \textit{ i.e., } 0\text{–}10 \text{ class interval.}$$

Similarly class intervals for other mid-values are obtained as:

Classes	f	$c.f.$
0—10	15	15
10—20	7	22
20—30	11	33
30—40	10	43 $c.f.$
40—50	13 f	56 M
50—60	8	64
60—70	20	84
70—80	16	100
	$N = 100$	

$$\text{Median item} = \text{Size of} \left(\frac{N}{2}\right)\text{th item} = \frac{100}{2} = 50\text{th item.}$$

∴ Median lies in the class 40—50.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 40 + \frac{50 - 43}{13} \times 10$$

$$= 40 + 5.385 = 45.385$$

IMPORTANT TYPICAL EXAMPLES

▶ To Locate Missing Frequency

Example 39. Find the missing frequency in the following distribution if $N = 100$ and $M = 30$.

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	10	?	25	30	?	10

Solution: Two frequencies are missing and let the missing frequencies be denoted by f_1 and f_2. We need two equations, we will get one from summation of frequencies and one from median formula.

Marks	f	$c.f.$
0—10	10	10
10—20	f_1	$10+f_1$
20—30*	25	$35+f_1$
30—40	30	$65+f_1$
40—50	f_2	$65+f_1+f_2$
50—60	10	$75+f_1+f_2$
	$N=100$	

1st Equation: From summation of frequencies,

$$75+f_1+f_2=100 \Rightarrow f_1+f_2=25 \qquad ...(i)$$

2nd Equation: From information regarding median,

$M=30$, Median class is 30–40.

Now,
$$M=l_1+\frac{\frac{N}{2}-c.f.}{f}\times i$$

$$30=30+\frac{50-(35+f_1)}{30}\times 10$$

$$0=\frac{15-f_1}{3}$$

or
$$f_1=15 \qquad ...(ii)$$

Substituting $f_1=15$ in equation (*i*),

$\therefore \quad f_1+f_2=25$

Put $f_1=15$

$\therefore \quad 15+f_2=25$

$\therefore \quad f_2=25-10=15$

Thus, $f_1=15,\ f_2=10$

Example 40. The following table gives distribution of marks secured by some students:

Marks:	0—20	20—30	30—40	40—50	50—60	60—70	70—80
No. of students:	42	38	120	84	48	36	31

Find (*i*) Median marks (*ii*) The percentage of failure if minimum for a pass is 35 marks.

Solution: (*i*) **Calculation of Median**

Marks	f	$c.f.$
0—20	42	42
20—30	38	80
30—40	120	200

The class 20—30 would cover upto 29.999 and 30 would be covered in class 30—40.

40—50	84	284
50—60	48	332
60—70	36	368
70—80	31	399
	$N = 399$	

$$\text{Median item} = \text{Size of } \frac{N}{2}\text{th item} = \frac{399}{2} = 199.5\text{th item.}$$

Hence, median lies in the class 30—40.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 30 + \frac{199.5 - 80}{120} \times 10$$

$$= 30 + 9.96 = 39.96 \approx 40$$

(*ii*) **Percentage of failure if the minimum for a pass is 35 marks.**

Under the assumption that observations in a class are uniformly distributed, th number of students getting less than 35 marks are:

$$= 42 + 38 + \left(\frac{35-30}{10}\right) \times 120$$

$$= 42 + 38 + 60 = 140$$

$\therefore$ Percentage of failure $= \frac{140}{399} \times 100 = 35.1$

Thus, the percentage of failure for a minimum pass marks is 35.1.

Example 41. The median of a few number of observations is given to be 48.67. The six iten whose values are 33.5, 38.9, 45.57, 49.03, 53.43 and 59.95 were added to give series, what will be the new median?

Solution: $M = 48.67$ (given)

Number of items less than median = 3

Number of items more than median = 3

Since, three items are less than given median and the same number of items are mor than median, hence the position of the given median, *i.e.*, 48.67 remains unchanged.

Graphical Location of Median

Median can also be located graphically with the help of ogive curve (or cumulative frequenc curve). Following are the steps involved in it:

(*i*) Firstly draw an ogive curve either 'less than ogive' or 'more than ogive'.

(*ii*) Compute $\frac{N}{2}$ and locate it (median item) on Y-axis.

(iii) Now draw a horizontal line from this point on Y-axis so as to meet at ogive curve.

(iv) Then draw a perpendicular line from the point of intersection on X-axis.

(v) The point of intersection on X-axis gives the value of median.

Aliter

Median can also be located by making use of both the ogive curves together. The point where 'less than ogive curve' and 'more than ogive curve' meet will give us the value of median.

Example 42. Locate the median graphically from the following data:

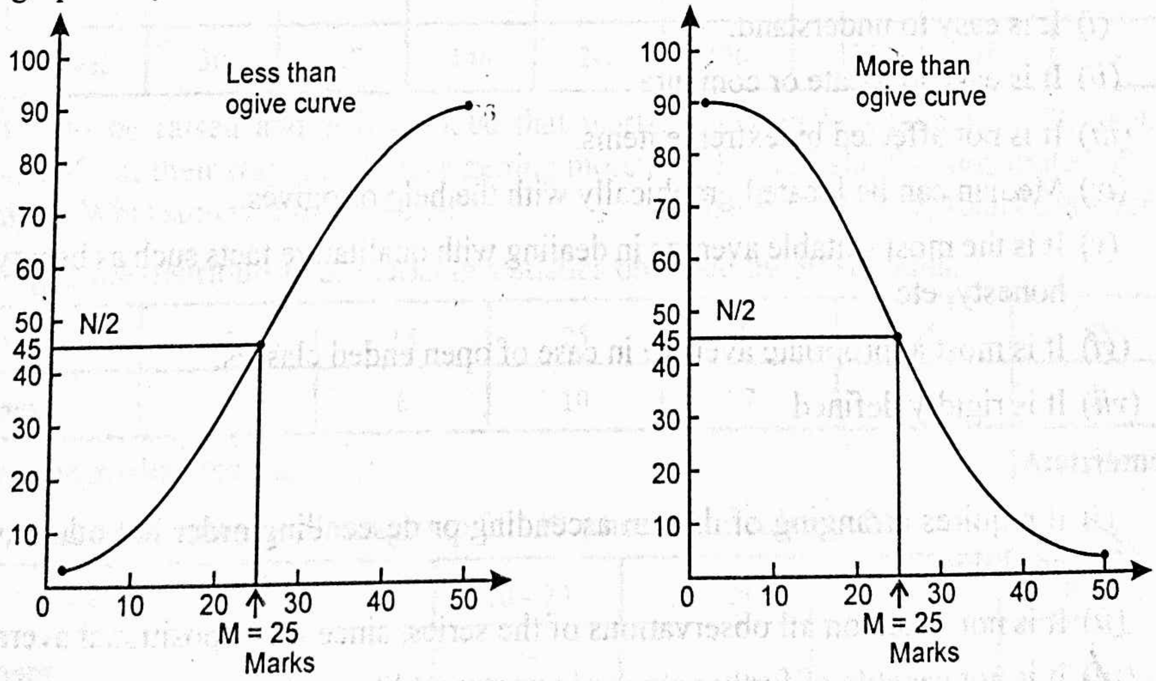
Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	10	20	30	20	10

Solution: Firstly we convert the given distribution into "less than" and "more than" cumulative frequencies distribution.

Less than method	
Marks	***c.f.***
Less than 10	10
Less than 20	30
Less than 30	60
Less than 40	80
Less than 50	90

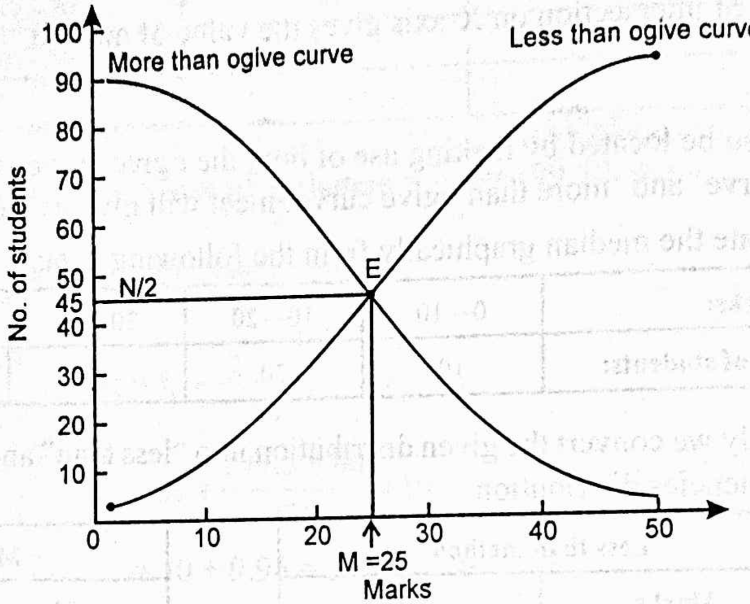
More than method	
Marks	***c.f.***
More than 0	90
More than 10	80
More than 20	60
More than 30	30
More than 40	10

Method I: Using the "less than ogive" and "more than ogive" method, the median is computed graphically as follows:



M = Size of $\frac{N}{2}$th item $= \frac{90}{2} = 45$th item. It lies in the class interval 20—30.

Method II: By making use of both the ogive curves together, the median is computed graphically as follows:



It is clear from the above figure that less than ogive curve and more than ogive curve meet each other at point E. The value of median is 25.

An Important Property

An important property of median is that the sum of the absolute deviations of the items from the median is less than the sum from any other value or average.

Merits and Demerits of Median

Merits:

(*i*) It is easy to understand.

(*ii*) It is easy to locate or compute.

(*iii*) It is not affected by extreme items.

(*iv*) Median can be located graphically with the help of ogives.

(*v*) It is the most suitable average in dealing with qualitative facts such as beauty, intelligence, honesty, etc.

(*vi*) It is most appropriate average in case of open ended classes.

(*vii*) It is rigidly defined.

Demerits:

(*i*) It requires arranging of data in ascending or descending order but other averages do not need this.

(*ii*) It is not based on all observations of the series, since it is a positional average.

(*iii*) It is not capable of further algebraic treatment like arithmetic mean.

(*iv*) It cannot be computed exactly where the number of items in a series is even.

(*v*) It is very difficult to calculate if the number of items is very small or large.

EXERCISE 5.5

1. Calculate the median of the following items:
25, 20, 15, 45, 18, 7, 10, 38, 12 [Ans. $M = 18$]

2. Calculate the median of the following items:
15, 20, 20, 23, 23, 25, 26, 27, 35, 40 [Ans. $M = 24$]

3. Calculate the median from the following data:

X:	15	20	25	30	35	40
f:	10	15	25	5	5	20

[Ans. $M = 25$]

4. Find out median marks from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	8	30	40	12	10

[Ans. $M = 23$]

5. Calculate the median from the following data:

Weight in gms.:	410—419	420—429	430—439	440—449	450—459	460—469	470—479
No. of apples:	14	20	42	54	45	18	7

[Ans. $M = 443.94$]

6. Find out median from the following table:

Monthly Wages (in Rs.):	50—80	80—100	100—110	110—120	120—130	130—150	150—180	180—200
No. of workers:	30	127	140	240	176	135	20	3

A fund is to be raised and it is decided that workers getting less than Rs. 120 should contribute 5% of their wages and those getting more than Rs. 120 should contribute 10% of their wages. What sum will be collected? [Ans. $M = 115.77$, Total Fund = 7261]

7. Following is the distribution of marks in statistics obtained by 50 students:

Mid-value:	5	15	25	35	45	55
Frequency :	4	6	10	7	3	2

Calculate the median marks. [Ans. $M = 26$]

8. Find the missing frequency of the group 20–30 when the median is 24.

Sale (in '000 Rs.)	0—10	10—20	20—30	30—40	40—50
No. of shops	5	25	—	18	7

[Ans. Missing frequency is 25]

9. The median of the following incomplete table is 46. Find the missing frequencies and calculate the arithmetic mean of the completed table:

Class interval:	10—20	20—30	30—40	40—50	50—60	60—70	70—80	Total
Frequency:	12	30	?	65	?	25	18	229

[Ans. $f_1 = 33 \cdot 5 \simeq 34$, $f_2 = 45$ and $\bar{X} = 45.82$]

10. Calculate median from the following data:

Marks less than:	80	70	60	50	40	30	20	10
No. of students:	100	90	80	60	32	20	13	5

[Ans. $M = 46.43$]

11. Below are given the marks obtained by 65 students in an examination. Find the median marks:

Marks more than:	70%	60%	50%	40%	30%	20%
No. of students:	7	18	40	40	63	65

[Ans. $M = 53.4$]

12. The following table gives the frequency distribution of the marks of 400 candidates in an examination:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90	90—100
No.of Candidates:	5	20	40	70	85	65	50	35	20	10

Find (*i*) Median marks, (*ii*) The percentage of pass if the minimum pass marks are 35.

[Ans. M = 47.64, % of Pass = 75%]

13. Find the median income graphically from the following data. Also verify the result algebraically.

Marks:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
No. of students:	4	6	10	10	25	22	18	5

[Ans. M = 2

(3) PARTITION VALUES—QUARTILES, DECILES AND PERCENTILES

Just as median divides the series into two equal parts, there are other useful measures whic divides the series into 4, 10 or 100 equal parts. They are called quartiles, deciles and percentiles.

(1) Quartiles: Quartiles divide a series into 4 equal parts. For any series there are three quartil denoted by Q_1, Q_2 and Q_3. Q_1 is known as first or lower quartile, covering 25% items. The seco quartile or Q_2 is the same as Median of the series. Q_3 is called third or upper quartile, covering 7 items.

(2) Deciles: Deciles divide a series into 10 equal parts. For any series, there are 9 deci denoted by $D_1, D_2 \ldots D_9$. These are called as first decile, second decile and so on.

(3) Percentiles: Percentiles divide a series into 100 equal parts. For any series, there are percentiles denoted by $P_1, P_2, P_3 \ldots P_{99}$.

Calculation of Quartiles, Deciles and Percentiles

The calculation of quartiles, deciles and percentiles is done in the same manner as the calculation of median.

For Individual and Discrete Series	For Continuous Series	Formula to be used in continuous series:
$Q_1 = \text{Size of } \frac{N+1}{4}\text{th item}$	$Q_1 = \text{Size of } \frac{N}{4}\text{th item}$	$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$
$Q_3 = \text{Size of } \frac{3(N+1)}{4}\text{th item}$	$Q_3 = \text{Size of } \frac{3N}{4}\text{th item}$	$Q_3 = l_1 + \frac{\frac{3}{4}N - c.f.}{f} \times i$
$D_1 = \text{Size of } \frac{N+1}{10}\text{th item}$	$D_1 = \text{Size of } \frac{N}{10}\text{th item}$	$D_1 = l_1 + \frac{\frac{N}{10} - c.f.}{f} \times i$
$D_9 = \text{Size of } \frac{9(N+1)}{10}\text{th item}$	$D_9 = \text{Size of } \frac{9N}{10}\text{th item}$	$D_9 = l_1 + \frac{\frac{9}{10}N - c.f.}{f} \times i$
$P_1 = \text{Size of } \frac{N+1}{100}\text{th item}$	$P_1 = \text{Size of } \frac{N}{100}\text{th item}$	$P_1 = l_1 + \frac{\frac{N}{100} - c.f.}{f} \times i$
$P_{99} = \text{Size of } \frac{99(N+1)}{100}\text{th item}$	$P_{99} = \text{Size of } \frac{99}{100}N\text{th item}$	$P_{99} = l_1 + \frac{\frac{99}{100}N - c.f.}{f}$

Individual Series

Example 43. From the following data, calculate Q_1, Q_3, D_5 and P_{25}

21, 15, 40, 30, 26, 45, 50, 54, 60, 65, 70

Solution: The data is first arranged in ascending order:

Sr. No.	X
1	15
2	21
3	26
4	30
5	40
6	45
7	50
8	54
9	60
10	65
11	70
$N = 11$	

$$Q_1 = \text{Size of}\left(\frac{N+1}{4}\right)\text{th item} = \text{Size of}\left(\frac{11+1}{4}\right)\text{th item} = \text{Size of 3rd item} = 26$$

Thus, $Q_1 = 26$

$$Q_3 = \text{Size of}\frac{3(N+1)}{4}\text{th item} = \text{Size of}\frac{3(11+1)}{4}\text{th item} = \text{Size of 9th item} = 60$$

Thus, $Q_3 = 60$

$$D_5 = \text{Size of}\frac{5(N+1)}{10}\text{th item} = \text{Size of}\frac{5(11+1)}{10}\text{th item}$$

$$= \text{Size of 6th item} = 45$$

Thus, $D_5 = 45$

$$P_{25} = \text{Size of}\frac{25(N+1)}{100}\text{th item} = \text{Size of}\frac{25(11+1)}{100}\text{th item}$$

$$= \text{Size of 3rd item} = 26$$

Thus, $P_{25} = 26$

Example 44. Calculate Q_1, Q_3, D_9 and P_{70} from the following data:

120, 150, 170, 180, 181, 187, 190, 192, 200, 210

Solution: The data is first arranged in ascending order:

Sr. No.	X
1	120
2	150
3	170
4	180
5	181
6	187
7	190
8	192
9	200
10	210
$N = 10$	

$$Q_1 = \text{Size of}\frac{N+1}{4}\text{th item} = \left(\frac{10+1}{4}\right)\text{th item}$$

$$= \text{Size of 2.75th item.}$$

$$= \text{Size of 2nd item} + \underline{0.75}\,(\text{Size of 3rd item} - \text{Size of 2nd item})$$

$$= 150 + \frac{3}{4}(170 - 150) = 150 + 15 = 165$$

$$Q_3 = \text{Size of } \frac{3(N+1)}{4}\text{th item} = \frac{3(10+1)}{4}$$

$= \text{Size of 8.25th item.}$

$= \text{Size of 8th item} + 0.25 \text{ (Size of 9th item} - \text{8th item)}$

$= 192 + \frac{1}{4}(200-192) = 192+2 = 194.$

$$D_9 = \text{Size of } \frac{9(N+1)}{10}\text{th item} = \frac{9(10+1)}{10} = \text{Size of 9.9 item.}$$

$= \text{Size of 9th item} + 0.9 \text{ (Size of 10th item} - \text{9th item)}$

$= 200 + \frac{9}{10}(210-200)$

$= 200+9 = 209$

$$P_{70} = \text{Size of } \frac{70(N+1)}{100}\text{th item} = \frac{70(10+1)}{100} = \text{Size of 7.7th item}$$

$= \text{Size of 7th item} + 0.7 \text{ (Size of 8th item} - \text{Size of 7th item)}$

$= 190 + \frac{7}{10}(192-190) = 190+1.40 = 191.40$

Thus, $Q_1 = 165, Q_3 = 194, D_9 = 209, P_{70} = 191.40$

Discrete Series

Example 45. Calculate Q_1, Q_3, D_6 and P_{85} from the following data:

X:	10	11	12	13	14	15	16	17	18
f:	3	4	5	12	10	7	5	2	1

Solution:

X	f	$c.f.$
10	3	3
11	4	7
12	5	12
13	12	24
14	10	34
15	7	41
16	5	46
17	2	48
18	1	49
	N = 49	

$$Q_1 = \text{Size of } \frac{N+1}{4}\text{th item} = \frac{49+1}{4} = \text{Size of 12.5th item} = 13$$

$$Q_3 = \text{Size of } \frac{3(N+1)}{4}\text{th item} = \frac{3(49+1)}{4} = \text{Size of 37.5th item} = 15$$

$$D_6 = \text{Size of } \frac{6(N+1)}{10}\text{th item} = \frac{6(49+1)}{10} = \text{Size of 30th item} = 14$$

$$P_{85} = \text{Size of } \frac{85(N+1)}{100}\text{th item} = \frac{85(49+1)}{100} = \text{Size of 42.5th item} = 16$$

Thus, $Q_1 = 13,\ Q_3 = 15,\ D_6 = 14,\ P_{85} = 16$

Continuous Series

Example 46. Calculate the values of Q_1, Q_3, D_8 and P_{56} from the following data:

Wages:	0—10	10—20	20—30	30—40	40—50
No. of workers:	22	38	46	35	19

Solution: Calculation of Q_1, Q_3, D_8 and P_{56}

Wages	f	$c.f.$
0—10	22	22
10—20	38	60
20—30	46	106
30—40	35	141
40—50	19	160
	N = 160	

$$Q_1 = \text{Size of } \frac{N}{4}\text{th item} = \text{Size of } \frac{160}{4}\text{th item} = \text{Size of 40th item.}$$

Q_1 lies in the class 10—20

Thus, $$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

$$= 10 + \frac{40-22}{38} \times 10 = 10 + 4.74 = 14.74$$

$\therefore$ $$Q_1 = 14.74$$

$$Q_3 = \text{Size of } \frac{3N}{4}\text{th item} = \frac{3(160)}{4}\text{th item} = \text{Size of 120th item}$$

Q_3 lies in the class 30—40.

Thus, $$Q_3 = l_1 + \frac{\frac{3}{4}N - c.f.}{f} \times i$$

$$= 30 + \frac{120-106}{35} \times 10 = 34$$

$\therefore \quad Q_3 = 34$

$$D_8 = \text{Size of } \frac{8N}{10}\text{th item} = \text{Size of } \frac{8\,(160)}{10}\text{ th item}$$

$$= \text{Size of 128th item}$$

D_8 lies in the class 30—40

Thus,
$$D_8 = l_1 + \frac{\frac{8N}{10} - c.f.}{f} \times i$$

$$= 30 + \frac{128 - 106}{35} \times 10 = 36.29$$

$\therefore \quad D_8 = 36.29$

$$P_{56} = \text{Size of } \frac{56N}{100}\text{th item} = \frac{56(160)}{100}\text{th item}$$

$$= \text{Size of 89.6th item.}$$

P_{56} lies in the class 20—30.

Thus,
$$P_{56} = l_1 + \frac{\frac{56}{100}N - c.f.}{f} \times i$$

$$= 20 + \frac{89.6 - 60}{46} \times 10 = 26.43$$

$\therefore \quad P_{56} = 26.43$

Example 47. Calculate Median, Quartiles, 6th decile and 70th percentile from the following data:

Marks less than:	80	70	60	50	40	30	20	10
No. of students:	100	90	80	60	32	20	13	5

Solution: The data is given in the form of a cumulative frequency distribution. First we convert it into simple frequency distribution and write it in ascending order:

Marks	f	$c.f.$
0—10	5	5
10—20	8	13
20—30	7	20
30—40	12	32
40—50	28	60
50—60	20	80
60—70	10	90
70—80	10	100
	$N = 100$	

$$\text{Median} = \text{Size of } \frac{N}{2}\text{th item} = \frac{100}{2} = \text{Size of 50th item}$$

Median lies in the class 40—50

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 40 + \frac{50-32}{28} \times 10$$

$$= 40 + \frac{18}{28} \times 10 = 40 + 6.42 = 46.42$$

$\therefore$ Median = 46.42

$$Q_1 = \text{Size of } \frac{N}{4}\text{th item} = \frac{100}{4} = 25\text{th item}$$

Q_1 lies in the class 30—40

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 30 + \frac{25-20}{12} \times 10$$

$$= 30 + \frac{5}{12} \times 10 = 30 + 4.16 = 34.16$$

$\therefore$ $Q_1 = 34.16$

$$Q_3 = \text{Size of } \frac{3N}{4}\text{th item} = \frac{3 \times 100}{4} = 75\text{th item.}$$

Q_3 lies in the class 50—60

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 50 + \frac{75-60}{20} \times 10$$

$$= 50 + \frac{15}{20} \times 10 = 50 + 7.5 = 57.5$$

$\therefore$ $Q_3 = 57.5$

$$D_6 = \text{Size of } \frac{6N}{10}\text{th item} = \frac{6 \times 100}{10} = 60\text{th item.}$$

D_6 lies in the class 40—50.

$$D_6 = l_1 + \frac{\frac{6}{10}N - c.f.}{f} \times i = 40 + \frac{60-32}{28} \times 10$$

$$= 40 + \frac{28}{28} \times 10 = 50$$

$\therefore$ $D_6 = 50$

$$P_{70} = \text{Size of } \frac{70N}{100}\text{th item} = \frac{70 \times 100}{100} = 70\text{th item}$$

P_{70} lies in the class 50—60

$$P_{70} = l_1 + \frac{\frac{70N}{100} - c.f.}{f} \times i = 50 + \frac{70-60}{20} \times 10$$

$$= 50 + \frac{10}{20} \times 10 = 55$$

$\therefore \quad P_{70} = 55$

IMPORTANT TYPICAL EXAMPLES

Example 48. Following is the distribution of marks in Economics obtained by 50 students:

Marks (more than):	0	10	20	30	40	50
No. of students:	50	46	40	20	10	3

Calculate the median marks. If 60% of the students pass this test, find the minimum marks obtained by a pass candidate.

Solution: Let us first of all convert cumulative frequencies into simple frequencies.

Marks	f	$c.f.$
0—10	$50 - 46 = 4$	4
10—20	$46 - 40 = 6$	10
20—30	$40 - 20 = 20$	30
30—40	$20 - 10 = 10$	40
40—50	$10 - 3 = 7$	47
50 and above	3	50
	$N = 50$	

(*i*) Median item = Size of $\left(\frac{N}{2}\right)$th item = Size of $\frac{50}{2}$th item = Size of 25th item.

Median lies in the class 20—30.

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{25-10}{20} \times 10$$

$$= 20 + \frac{15}{2}$$

$$= 20 + 7.5 = 27.5$$

(*ii*) Since, 60% candidates are passed and 40% have failed, the minimum pass marks are given by P_{40}. In other words, we have to find the value of P_{40}.

$$P_{40} = \text{Size of}\left(\frac{40N}{100}\right)\text{th item} = \text{Size of}\left(\frac{40\times50}{100}\right)\text{th item}$$

$$= \text{Size of 20th item.}$$

P_{40} lies in the class 20—30

$$\therefore \quad P_{40} = l_1 + \frac{\frac{40}{100}N - c.f.}{f} \times i$$

$$= 20 + \frac{20-10}{20} \times 10 = 20 + 5 = 25$$

Thus, the minimum marks obtained in a test by a pass candidate are 25.

Example 49. The first and third quartiles of the following data are given to be 25 marks and 50 marks respectively out of the data given below:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80
Frequency:	4	8	—	19	—	10	5	—

Find the missing frequencies when $N = 72$.

Solution: Let the missing frequencies be x, y and z. The given table is written as:

Marks	**Frequency (f)**	***c.f.***
0—10	4	4
10—20	8	12
20—30	x	$12+x$
30—40	19	$31+x$
40—50	y	$31+x+y$
50—60	10	$41+x+y$
60—70	5	$46+x+y$
70—80	z	$46+x+y+z$
	$N = 72$	

Since, $N = 72$

$$\therefore \quad x + y + z + 46 = 72$$

$$\Rightarrow \quad x + y + z = 72 - 46 = 26 \quad ...($$

$$Q_1 = \left(\frac{N}{4}\right)\text{th item} = \frac{72}{4} = 18\text{th item.}$$

Since x is still unknown, therefore, we cannot locate the position of Q_1 from c.f. column. As $Q_1 = 25$

$\therefore \quad Q_1$ lies in class 20–30.

Now, $$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

$$25 = 20 + \frac{18-12}{x} \times 10$$

or $$5 = \frac{6}{x} \times 10$$

$\Rightarrow$ $$x = 12 \qquad ...(ii)$$

$$Q_3 = \left(\frac{3N}{4}\right)\text{th item} = \frac{3 \times 72}{4} = 54\text{th item.}$$

Since, the value of y is still unknown, therefore, we cann not locate the position of Q_3 from c.f. column. As $Q_3 = 50, Q_3$ lies in 50–60 class.

Now, $$Q_3 = l_1 + \frac{\left(\frac{3N}{4} - c.f.\right)}{f} \times i$$

$$50 = 50 + \frac{(54 - 31 - x - y)}{10} \times 10$$

$$0 = 54 - 31 - x - y$$

$\Rightarrow$ $$x + y = 23 \qquad ...(iii)$$

Putting the value of x from (*ii*) in equation (*iii*), we have

$$12 + y = 23$$

$\Rightarrow$ $$y = 11$$

Now, $$x + y + z = 26$$

$$12 + 11 + z = 26$$

$\Rightarrow$ $$z = 3$$

$\therefore$ $x = 12$, $y = 11$ and $z = 3$

Example 50. Given the following distribution of income:

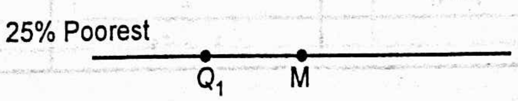
Income (thousands):	0—1	1—2	2—3	3—4	4—5	5—6	6—7	7—8
Number of families	4	6	10	15	8	5	4	2

Find out (*i*) Highest income among the poorest 25%.

(*ii*) Lowest income among the richest 30%.

Solution: (*i*) We are interested in finding that level of income less than or equal to which 25% of families have their income. So we are to find Q_1.

See Diagram **Ascending Order Distribution of Income**



Income (thousands) (X)	Number of families (f)	$c.f.$
0—1	4	4
1—2	6	10
2—3	10	20
3—4	15	35
4—5	8	43
5—6	5	48
6—7	4	52
7—8	2	54
	$N = 54$	

$$Q_1 = \text{Size of } \frac{N}{4}\text{th item}$$

$$Q_1 = \text{Size of} \left(\frac{54}{4}\right) \text{th item} = 13.5\text{th item}$$

Q_1 lies in interval 2—3.

$$Q_1 = l_1 + \frac{\left(\frac{N}{4} - c.f.\right)}{f} \times i$$

$$Q_1 = 2 + \frac{13.5 - 10}{10} \times 1 = 2 + 0.35 = \text{Rs. } 2.35 \text{ thousands.}$$

So, highest income among the poorest 25% is Rs. 2.35 thousand.

(*ii*) We are interested in finding that level of income equal to or above which have 30% of the families have their income. So we are to find P_{70} or D_7.

See Diagram **Ascending Order Distribution of Income**

Richest 30% families

D_7 or P_{70}

Income (thousands) (X)	Number of families (f)	$c.f.$
0—1	4	4
1—2	6	10
2—3	10	20
3—4	15	35
4—5	8	43
5—6	5	48
6—7	4	52
7—8	2	54
	$N = 54$	

D_7 is equal to the size of $\frac{(7 \times 54)}{10}$ th item = 37.8th item

It lies in interval 4—5.

$$D_7 = l_1 + \frac{\frac{7N}{100} - c.f.}{f} \times i$$

$$D_7 = 4 + \frac{37.8 - 35}{8} \times 1$$

$$D_7 = 4 + \frac{2.8}{8} = \text{Rs. } 4.35 \text{ thousands.}$$

Hence, the lowest income among the richest 30% is 4.35 thousand rupees.

Example 51. An investigator collected the following information regarding the wage distribution:

Wages (Rs. week):	0—500	500—1000	1000—1500	1500—2000	2000—2500
Number of workers:	4	8	12	15	20
Wages (Rs. week):	2500—3000	3000—3500	3500—4000	4000—4500	4500—5000
Number of workers:	27	15	8	5	3

(*i*) Find out median wages.

(*ii*) Find out the highest wage among the lowest paid 30% workers.

(*iii*) Within what limits do the middle 20% workers have their wages?

(*iv*) The management decides to give bonus to workers with wages less than or equal to Rs. 3300 per week. How many workers will get the bonus?

Solution:

Wages (Rs. week) (*X*)	No. of workers (*f*)	*c.f.*
0—500	4	4
500—1000	8	12
1000—1500	12	24
1500—2000	15	39
2000—2500	20	59
2500—3000	27	86
3000—3500	15	101
3500—4000	8	109
4000—4500	5	114
4500—5000	3	117
	$N = 117$	

(i) $$\text{Median} = \text{Size of} \left(\frac{N}{2}\right)\text{th item} = \frac{117}{2} = 58.5\text{th item}$$

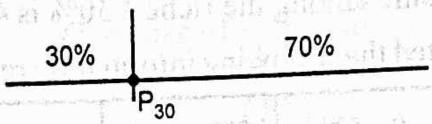
Median lies in the interval 2000—2500.

$$M = l_1 + \frac{\left(\frac{N}{2} - c.f.\right)}{f} \times i$$

$$M = 2000 + \frac{58.5-39}{20} \times 500 = 2000 + 487.5$$

$$M = \text{Rs. } 2487.5$$

(ii) Ascending Order Distribution of Wages



As shown here, the highest wage among the lowest paid 30% workers is given by P_{30}.

$$P_{30} = \text{Size of}\left(\frac{30N}{100}\right)\text{th item} = \frac{(30)(117)}{100} = 35.1\text{th item.}$$

P_{30} will lie in the class interval 1500—2000.

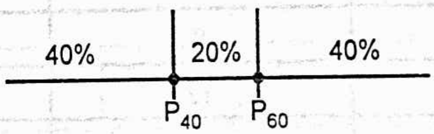
$$P_{30} = l_1 + \frac{\frac{30N}{100} - c.f.}{f} \times i$$

$$P_{30} = 1500 + \frac{35.1-24}{15} \times 500 = 1500 + \frac{11.1}{15} \times 1500$$

$$P_{30} = \text{Rs. } 1870.$$

(iii) The middle 20% of the workers will have wages between P_{40} and P_{60} as shown below:

Ascending Order Distribution of Wages



$$P_{40} = \text{Size of}\left(\frac{40N}{100}\right)\text{th item} = \frac{40 \times 117}{100} = 46.8\text{th item.}$$

It lies in the interval 2000—2005.

$$P_{40} = 2000 + \frac{46.8-39}{20} \times 500 = \text{Rs. } 2195$$

$$P_{60} = \text{Size of}\left(\frac{60N}{100}\right)\text{th item} = \frac{60 \times 117}{100} = 70.2\text{th item.}$$

It lies in the interval 2500—3000.

$$P_{60} = 2500 + \frac{70.2-59}{27} \times 500 = \text{Rs. } 2707.41$$

The limits within which the middle 20% of the workers will have their wages are Rs. 2195 and Rs 2707.41.

(*iv*) Suppose K workers have wages less than Rs. 3300. The class interval under question is 3000—3500. The value of K will be determined as below:

$$3300 = 3000 + \frac{K - c.f.}{f} \times i$$

$$3300 = 3000 + \frac{K-86}{15} \times 500 \quad \Rightarrow \quad 300 = \frac{K-86}{15} \times 500$$

$$\Rightarrow \quad K - 86 = \frac{300 \times 15}{500} = 9 \quad \Rightarrow \quad K = 95$$

The number of workers with wages equal to or less than Rs. 3300 is 95. So 95 workers will have to paid bonus.

Note: It is being assumed that distribution is uniform within class interval.

Alternative Approach to (*d*) Part:

Number of workers with wages from 0—3000 = 4 + 8 + 12 + 15 + 20 + 27
= 86 (given)

Number of workers within 3000—3500 = 15 (given)

Number of workers within 3000—3300 = $\frac{15}{500} \times 300 = 9$

Therefore, number of workers with wages less than or equal to Rs. 3300
= 86 + 9 = 95

xample 52. From the following data, calculate the percentage of workers getting wages:
(*i*) more than Rs. 44.
(*ii*) between Rs. 22 and Rs. 58.

Wage (in Rs.):	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80
Number of workers:	20	45	85	160	70	55	35	30

olution:

Wage (in Rs.)	**Number of workers (f)**	***c.f.***
0—10	20	20
10—20	45	65
20—30	85	150
30—40	160	310
40—50	70	380
50—60	55	435
60—70	35	470
70—80	30	500
	$N = 500$	

(i) Let K% workers get less than Rs. 44. $\therefore$ $P_K = 44$.

$\therefore$ P_K class is 40—50.

Now, $P_K = l_1 + \left(\dfrac{K\left(\dfrac{N}{100}\right) - c.f.}{f} \right) \times i \Rightarrow 44 = 40 + \left(\dfrac{K\left(\dfrac{500}{100}\right) - 310}{70} \right) \times 10$

$\Rightarrow \quad 4 = \dfrac{5K - 310}{7} \quad \Rightarrow 5K = 338$

$\Rightarrow \quad K = 67.6$

$\therefore$ 67.6% workers get less than Rs. 44.

$\therefore$ Percentage of workers getting more than Rs. 44

$= 100\% - 67.6\%$

$= 32.4\%$

Alternative Method:

The number of workers getting more than 44 can be obtained directly as :

$= \dfrac{50 - 44}{10} \times 70 + 55 + 35 + 30$

$= 42 + 55 + 35 + 30$

$= 162$

Percentage of workers $= \dfrac{162}{500} \times 100 = 32.4\%$

(ii) Let K% workers get less than Rs. 22. $\therefore$ $P_K = 22$.

$\therefore$ P_K class is 20—30.

Now, $P_K = l_1 + \left(\dfrac{K\left(\dfrac{N}{100}\right) - c.f.}{f} \right) \times i$

$\Rightarrow \quad 22 = 20 + \left(\dfrac{K\left(\dfrac{500}{100}\right) - 65}{85} \right) \times 10$

$\Rightarrow \quad 2 = \dfrac{(5K - 65) \times 2}{17} \quad \Rightarrow \quad 34 = 10K - 130$

$\Rightarrow \quad K = 16.4.$

$\therefore$ 16.4% workers get less than Rs. 22.

Let K% workers get less than Rs. 58. $\therefore$ $P_K = 58$

$\therefore$ P_K class is 50—60.

Now, $$P_K = l_1 + \left(\frac{K\left(\frac{N}{100}\right) - c.f.}{f}\right) \times i$$

$$\Rightarrow \quad 58 = 50 + \left(\frac{K\left(\frac{500}{100}\right) - 380}{55}\right) \times 10$$

$$\Rightarrow \quad 8 = \frac{(5K - 380) \times 2}{11} \quad \Rightarrow \quad 88 = 10K - 760$$

$$\Rightarrow \quad K = 84.8$$

$\therefore$ 84.8% workers get less than Rs. 58.

$\therefore$ Percentage of workers getting between Rs. 22 and Rs. 58

$$= 84.8\% - 16.4\% = 68.4\%$$

Alternative Method:

The number of workers getting between Rs. 22 and Rs. 58 can be obtained directly as:

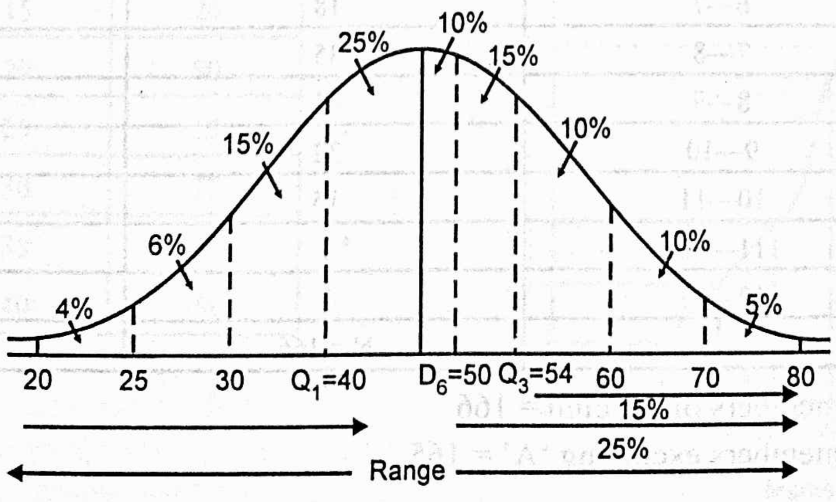
$$= \frac{30-22}{10} \times 85 + 160 + 70 + \frac{58-50}{10} \times 55$$

$$= 68 + 160 + 70 + 44 = 342$$

$$\text{Percentage of workers} = \frac{342}{500} \times 100 = 68.4\%$$

Example 53. For a group of 5,000 workers, the weekly wages vary from Rs. 20 and Rs.80. The wages of 4 per cent of workers are under Rs. 25 and those of 10 per cent are under Rs. 30. 15 per cent of the workers each Rs. 60and over and 5 per cent of them get Rs. 70 and over. The quartile wages are Rs. 40 and Rs. 54. The sixth Decile is Rs. 50. Put this information in a frequency table and find mean wages.

Solution: First we put the information in the form of a normal curve.



Calculation of Mean Wages

Weekly wages (Rs.)	Percentage(%) of workers	No.of workers (f)	Mid-value (m)	fm
20-25	4%	$\frac{4\times5000}{100}=200$	22.5	4500
25-30	6%	$\frac{6\times5000}{100}=300$	27.5	8250
30-40	15%	$\frac{15\times5000}{100}=750$	35	26250
40-50	35%	$\frac{35\times5000}{100}=1750$	45	78750
50-54	15%	$\frac{15\times5000}{100}=750$	52	39000
54-60	10% (balance)	$\frac{10\times5000}{100}=500$	57	28500
60-70	10%	$\frac{10\times5000}{100}=500$	65	32500
70-80	5%	$\frac{5\times5000}{100}=250$	75	18750
		N = 5000		$\Sigma fm = 236500$

$$\overline{X} = \frac{\Sigma fm}{N} = \frac{236500}{5000} = 47\cdot3$$

Example 54. The age distribution of the members of a certain children's club is as follows:

Age as on last birth day (in years):	4	5	6	7	8	9	10	11	12
Number of children:	5	9	18	35	42	32	15	7	3

There is a member 'A' such that there are twice as many members older than 'A' as there are members younger than 'A'. Estimate the age of A.

Solution: As the age on the last birthday is given, we can put the data in the exclusive form of class intervals as under:

Age in years	Number of children	c.f.
4—5	5	5
5—6	9	14
6—7	18	32
7—8	35	67
8—9	42	109
9—10	32	141
10—11	15	156
11—12	7	163
12—13	3	166
	N = 166	

Total members of the club = 166

Total members excluding 'A' = 165

Let total members younger than 'A' = x

The members older than 'A' = $2x$

Thus, $x + 2x = 165 \therefore x = 55$

Thus, 'A's position is at 56th in order of age.

$\therefore$ A's age lies in 7—8 age group

$$\text{A's age} = l_1 + \frac{\text{A's position} - c.f.}{f} \times i$$

$$= 7 + \frac{56-32}{35} \times 1 = 7 + 0.6857 = 7.6857 = 7.6 \text{ years}$$

Graphical Location of Quartiles, Deciles and Percentiles

The various partition values *viz.*, quartiles (Q_1 and Q_3), deciles and percentiles can be easily located graphically with the help of less than cumulative frequency curve or ogive. For first and third quartiles, we mark the values $\frac{N}{4}$ and $\frac{3N}{4}$ on the Y-axis. For rth decile we mark the value $\left(\frac{N}{10}\right)r$ and for rth percentile we mark $\left(\frac{N}{100}\right)r$ on Y-axis. Rest of the steps are the same as explained for drawing median by using the lees than ogive.

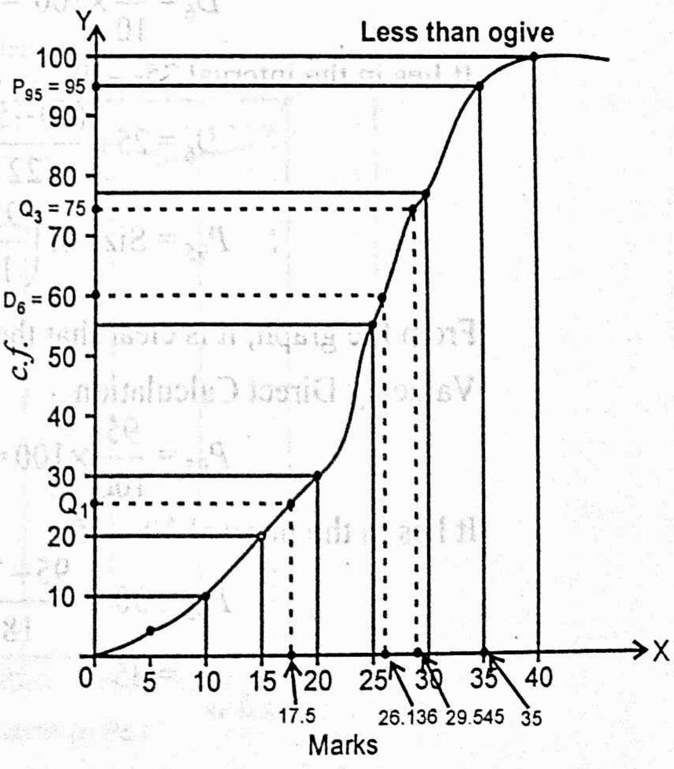
Example 55. The marks obtained by 100 students of a university are given below:

Marks:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
No. of students:	4	6	10	10	25	22	18	5

Draw a less than ogive from the data given above and hence find out Q_1, Q_3, D_6 and P_{95}. Also verify your results by direct formula calculation.

Solution:

Marks	*c.f.*
Less than 5	4
10	10
15	20
20	30
25	55
30	77
35	95
40	100



$$Q_1 = \text{Size of}\left(\frac{N}{4}\right)\text{th item } = \frac{100}{4} = 25\text{th item}$$

From the graph, it is clear that size of 25th item = 17.5.

Value by Direct Calculation

$$Q_1 = \text{Size of}\left(\frac{N}{4}\right)\text{th item } = \frac{100}{4} = 25\text{th item}$$

It lies in the interval 15—20.

$$Q_1 = 15 + \frac{25-20}{10} \times 5 = 17.5$$

$$Q_3 = \text{Size of}\left(\frac{3N}{4}\right)\text{th item} = \frac{3 \times 100}{4} = 75\text{th item.}$$

From the graph, it is clear that size of 75th item = 29.545.

Value by Direct Calculation

$$Q_3 = \text{Size of}\left(\frac{3N}{4}\right)\text{th item} = \frac{3 \times 100}{4} = 75\text{th item.}$$

It lies in the interval 25—30.

$$Q_3 = 25 + \frac{75-55}{22} \times 5 = 29.545$$

$$D_6 = \text{Size of}\left(\frac{6N}{10}\right)\text{th item } \frac{6 \times 100}{10} = 60\text{th item.}$$

From the graph, it is clear that the size of 60th item = 26.136.

Value by Direct Calculation

$$D_6 = \frac{6}{10} \times 100 = 60\text{th item.}$$

It lies in the interval 25—30.

$$D_6 = 25 + \frac{(60-55)}{22} \times 5 = 26.136$$

$$P_{95} = \text{Size of}\left(\frac{95N}{100}\right)\text{th item} = \frac{95 \times 100}{100} = 95\text{th item}$$

From the graph, it is clear that the size of 95th item = 35.

Value by Direct Calculation

$$P_{95} = \frac{95}{100} \times 100 = 95\text{th item.}$$

It lies in the interval 30—35.

$$P_{95} = 30 + \frac{95-77}{18} \times 5$$

$$= 35$$

Example 56. Given below is the pre-tax monthly income of residents of an industrial town:

Pre-tax income (Rs.)	No. of Residents (in thousands)
More than 7,000	2
More than 6,000	8
More than 5,000	10
More than 4,000	15
More than 3,000	25
More than 2,000	40
More than 1,000	55
More than 0	60

Draw a 'less than ogive' and hence find out.

(*i*) the highest income of the lowest 50% of the residents; and

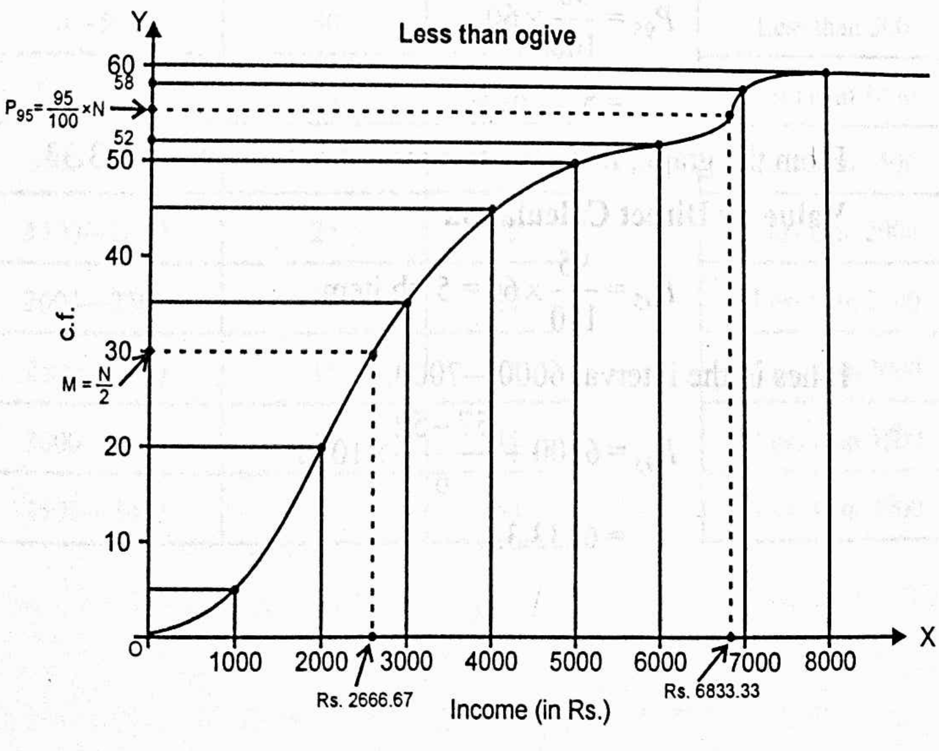
(*ii*) the minimum income earned by the top 5% of the residents.

Also, verify you results by direct formula calculation.

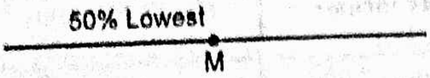
Solution: First we convert the given cumulative frequency distribution into simple frequency distribution and then into less than cumulative frequency distribution.

Income (Rs.)	*f*	*c.f.*
0—1000	5	5
1000—2000	15	20
2000—3000	15	35
3000—4000	10	45
4000—5000	5	50
5000—6000	2	52
6000—7000	6	58
7000—8000	2	60

Income	*c.f.*
Less than 1000	5
Less than 2000	20
Less than 3000	35
Less than 4000	45
Less than 5000	50
Less than 6000	52
Less than 7000	58
Less than 8000	60



(*i*) **Ascending Order Distribution of Income**



We are interested in finding the highest income of the lowest 50% of the residents. So we are find M or D_5 or P_{50}

$$M = \text{Size of}\left(\frac{N}{2}\right)\text{th item} = \frac{60}{2} = 30\text{th item.}$$

From the graph, it is clear that size of 30th item is 2667.

Value by Direct Calculation

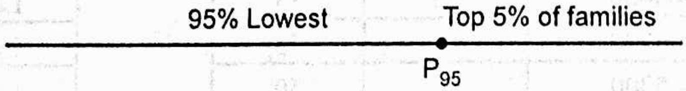
$$\text{Median} = \frac{N}{2} = \frac{60}{2} = 30\text{th item.}$$

It lies in the class interval 2000—3000.

$$M = 2000 + \frac{30-20}{15}\times 1000$$

$$= 2666.67 \approx 2667 \text{ approx.}$$

Ascending Order Distribution of Income



(*ii*) We are interested in finding the income earned by 5% of the families. So we are to find P_{95}.

$$P_{95} = \frac{95}{100}\times 60$$

$$= 57\text{th item.}$$

From the graph, it is clear that size of 57th item is 6833.33.

Value by Direct Calculation

$$P_{95} = \frac{95}{100}\times 60 = 57\text{th item.}$$

It lies in the interval 6000—7000.

$$P_{95} = 6000 + \frac{57-52}{6}\times 1000$$

$$= 6833.33.$$

Example 57. The monthly salary distribution of 250 families in a certain locality in Agra is given below:

Monthly salary	No. of families
More than 0	250
More than 500	200
More than 1,000	120
More than 1,500	80
More than 2,000	55
More than 2,500	30
More than 3,000	15
More than 3,500	5

Draw a 'less than' ogive for the data given above and hence find out:

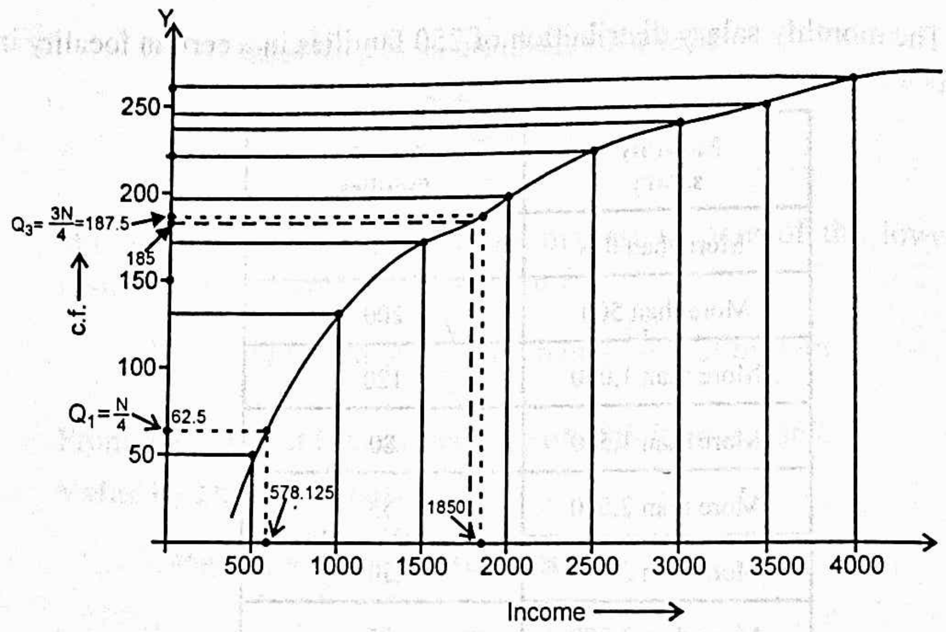
(*i*) Limits of the income of middle 50% of the families; and

(*ii*) If income-tax is to be levied on families, whose income exceeds Rs. 1,800 p.m., calculate the percentage of families, which will be paying income-tax.

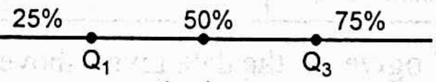
Solution: First we convert the given frequency distribution into simple frequency distribution and then into less than cumulative frequency distribution.

Monthly salary	No. of families (f)	*c.f.*
0—500	50	50
500—1000	80	130
1000—1500	40	170
1500—2000	25	195
2000—2500	25	220
2500—3000	15	235
3000—3500	10	245
3500—4000	5	250

Monthly salary	*c.f.*
Less than 500	50
Less than 1000	130
Less than 1500	170
Less than 2000	195
Less than 2500	220
Less than 3000	235
Less than 3500	245
Less than 4000	250



(*i*) The middle 50% of the families will have income between Q_3 and Q_1 as shown below:



$$Q_1 = \text{Size of}\left(\frac{N}{4}\right)\text{th item} = \frac{250}{4} = 62.5\text{th item.}$$

From the graph, it is clear that $Q_1 = 578.125$.

Value by Direct Calculation

$$Q_1 = \text{Size of}\left(\frac{N}{4}\right)\text{th item} = \frac{250}{4} = 62.5\text{th item.}$$

It lies in the class interval 500—1000.

$$Q_1 = 500 + \frac{62.5 - 50}{80} \times 500 = 578.125$$

$$Q_3 = \text{Size of}\left(\frac{3N}{4}\right)\text{th item}\ \frac{3 \times 250}{4} = 187.5\text{th item.}$$

From the graph, it is clear that $Q_3 = 1850$.

Value by Direct Calculation

$$Q_3 = \text{Size of}\left(\frac{3N}{4}\right)\text{th item}\ \frac{3 \times 250}{4} = 187.5\text{th item.}$$

It lies in the class interval 1500—2000.

$$Q_3 = 1500 + \frac{187.5 - 170}{25} \times 500 = \text{Rs } 1{,}850.$$

Income limit of middle 50% families = 1850 – 578.12 = Rs. 1,271.88

(*ii*) From the graph, it is clear that number of families having income less than equal to 1800 = 185

No. of families having income more than 1800 = 250 – 185 = 65

Percentage of families having income more than 1800 $= \frac{65}{250} \times 100 = 26$

Value by Direct Calculation

Let K% of families getting less or equal to 1800; $\therefore$ $P_K = 1800$

$\therefore$ P_K class is 1500—2000.

Now, $$P_K = 1500 + \frac{\frac{K \times 250}{100} - 170}{25} \times 500$$

$$\Rightarrow \quad 1800 = 1500 + \frac{2.5K - 170}{25} \times 500$$

$$\frac{300 \times 25}{500} = 2.5K - 170$$

$$2.5K - 170 = 15$$

$$K = \frac{185}{2.5} = 74\%.$$

Percentage of families getting less than 1800 = 74%.

Percentage of families getting more than 1800 = 100 – 74 = 26%.

Another Method:

Percentage of families having income more than in 1800

$$= \frac{2000 - 1800}{500} \times 25 + 25 + 15 + 10 + 5$$

$$= 10 + 25 + 15 + 10 + 5 = 65$$

Percentage of families $= \frac{65}{250} \times 100 = 26\%$

EXERCISE 5.6

1. 19 students of B. Com. II class secured following marks in Economics:
   18, 20, 25, 17, 9, 11, 23, 37, 38, 42, 36, 35, 8, 10, 11, 21, 20, 41, 35
   Calculate Q_1, Q_3, D_7 and P_{78} [Ans. $Q_1 = 11, Q_3 = 36, D_7 = 35, P_{78} = 36.6$]

2. From the following data calculate Q_1, Q_3 and P_{60}

Sr.No.	1	2	3	4	5	6	7	8	9	10
Marks:	12	30	20	15	25	10	2	40	4	8

[Ans. $Q_1 = 7, Q_3 = 26.25, P_{60} = 18$]

3. Calculate Median, Q_1 and Q_3, D_6 and P_{89} from the following data:

X:	1	2	3	4	5	6	7	8	9	10	11
f:	5	8	15	22	36	44	28	17	12	9	3

[Ans. $M = 6, Q_1 = 4, Q_3 = 7, D_6 = 6, P_{89} = 9$]

4. Calculate median, lower and upper quartiles, first decile and 68th percentile from the following data:

Size:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
f:	6	7	9	18	25	15	12	8

[Ans. M =22, Q_1=15.83, Q_3=28.33, D_1=7.86, P_{68}=26]

5. Calculate the median and quartiles for the following:

Marks below:	10	20	30	40	50	60	70	80
No. of students:	15	35	60	84	96	127	198	250

[Ans. M=59.35, Q_1=31.04, Q_3=68.52]

6. The first quartile of the following data is given to be 21.5 marks.

Marks:	10—15	15—20	20—25	25—30	30—35	35—40	40—45	45—50
Frequency:	24	?	90	122	?	56	20	33

Find the missing frequencies when N = 460. [Ans. f_1 = 64 and f_2 = 5]

7. For a certain group of 100 saree-weavers of Varanasi, the median and quartile earnings per week are Rs. 44.3, Rs. 43.0 and Rs. 45.9 respectively. The earnings for the group range between Rs. 40 and Rs. 50. 10 per cent of the group earn under Rs. 42.13 per cent earn Rs. 47 and over and 6 per cent Rs. 48 and other. Put these figures in the form of frequency distribution and obtain the value of the mean wage. [Ans. $\overline{X}$ = 44.5]

8. The following series relates to the daily income of working employed in a firm. Compute (*i*) highest income of lowest 50% workers (*ii*) minimum income earned by the top 25% of the workers, and (*iv*) maximum income earned by lowest 25% workers:

Daily Income (Rs.):	10—14	15—19	20—24	25—29	30—34	35—39
No. of Workers:	5	10	15	20	10	5

[**Hint:** Compute Median, Upper Quartile and Lower Quartile]

[Ans. (*i*) Rs. 25.11 (*ii*) Rs. 29.19 (*iii*) Rs. 19.92

9. From the following data of wages, find the percentage of workers getting wages (*i*) between Rs. 125 and Rs. 260; (*ii*) more than Rs. 340.

Weekly wages (Rs.)	50—100	100—150	150—200	200—250	250—300
No. of workers	15	40	35	60	125
Weekly wages (Rs.)	300—350	350—400	400—450	450—500	
No. of workers	100	70	40	15	

[Ans. (*i*) 28% (*ii*) 29%

10. If the quartiles for the following distribution are $Q_1 = 23.125$ and $Q_3 = 43.5$, find the missing frequencies and median of the distribution:

Daily wages:	0—10	10—20	20—30	30—40	40—50	50—60
No. of workers:	5	—	20	30	—	10

[Ans. f_1 =11, f_2 =13.669 ≈ 14 (approx.

11. Find the percentage of workers who earned more than Rs. 840.

Income (in Rs.):	500—600	600—700	700—800	800—900	900—1000	1000—1100
No. of workers:	12	17	22	29	12	6

[Ans. 36.12%]

12. Given below is the distribution of works obtained by 50 students in a class test:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	3	5	9	12	18	3

If 70% students pass the test, find the minimum marks needed by a student to pass the examination.

[**Hint:** Compute P_{70}] [Ans. 28 Marks]

13. Draw a less than ogive from the following data:

Weekly income (Rs.) (equal to or more than)	12,000	11,000	10,000	8,000	6,000	4,000	3,000	2,000	1,000
No. of families	0	6	14	26	42	54	62	70	80

From the graph estimate the number of families in the income range of Rs. 2,500 and Rs. 10,500. Also find maximum income of the lowest 25% of the families. [Ans. (*i*) 56, (*ii*) Rs. 3250]

14. With the help of given data, find (*i*) value of middle 50% class, (*ii*) value of exactly 50% class, (*iii*) the value of P_{40} and D_6, (*iv*) graphically with the help of ogive curves, the values of Q_1, Q_3, median, P_{40} and D_6.

C.I.:	10—14	15—19	20—24	25—29	30—34	35—39
f:	5	10	15	20	10	5

[Ans. $Q_3 - Q_1 = 9.2375$, $M(= Q_2) = 25.125$, $P_{40} = 23.167$, $D_6 = 26.75$]

15. The following table gives the frequency distribution of 800 candidates in an examination:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90	90—100
No. of candidates	10	40	80	140	170	130	100	70	40	20

Draw less than ogive and answer the following questions from the graph:

(*i*) If the minimum works required for passing are 35, what percentage of candidates pass the examination?

(*ii*) If it is decided to allow 80% of the candidates to pass, what should be the minimum marks for passing?

[**Hint:** See Example 122] [Ans. (*i*) 75% (*ii*) 32 approx.]

(4) MODE

Mode is another important measure of central tendency. Mode is defined as the value which occurs most frequently in a series. In other words, it is a value which has the greatest frequency in a distribution. For example, the mode of the series 20, 21, 23, 23, 23, 23, 25, 26, 26 would be 23, since this value occurs most frequently than any of other values. In the words of **Kenny and Keeping**, "*The value of the variable which occurs most frequently in a distribution is called the mode*". According to **A.M. Tuttle**, "*Mode is the value which has the greatest frequency density*". The above definitions indicate that mode is the value around which there is greatest concentration of items. Mode is denoted by the symbol 'Z'.

Calculation of Mode

Individual Series

In case of individual series, mode can be computed by applying any of the two methods:

(1) Inspection Method: This method involves an inspection of the items. One is to simply identify the value that occurs most frequently in a series. Such a value is called mode.

Example 58. Find the mode from the following data:

8, 10, 5, 8, 12, 7, 8, 9, 11, 7

Solution: An inspection of the series shows that the value 8 occurs most frequently in the series. Hence, Mode (Z) = 8.

(2) By changing the Individual series into Discrete Series: When the numbers of items in a series is very large, individual series is first converted into discrete series. Then we identify the value corresponding to which there is highest frequency. Such a value is called mode.

Example 59. Find the mode from the following data:

11.1, 10.9, 10.7, 11.1, 10.6, 11.3, 10.6, 10.7, 10.6, 10.9, 10.6, 10.5, 10.4, 10.6

Solution: First we convert the given series into a discrete series in ascending order as follows:

Size:	10.4	10.5	10.6	10.7	10.9	11.1	11.3
Frequency:	1	1	5	2	2	2	1

The modal value is 10.6 since it appears maximum number of times in the series.

Discrete Series

For calculating mode in discrete series, the following two methods may be used:

(1) Inspection Method

(2) Grouping Method.

(1) Inspection Method: In this method, the value of mode is determined by inspecting the series. The value whose frequency is maximum is mode. Generally, inspection method is used in only those cases where frequency increase upto a point and after reaching the maximum, decline

Example 60. Find out the mode from the following data:

Income (Rs.):	110	120	130	140	150	160
No. of persons:	2	4	8	10	5	4

Solution: An inspection of the series reveals that the value 140 has the maximum frequency, *i.e.*, 10. Thus, mode is 140.

(2) Grouping Method: In some cases, it is possible that value having the highest frequency may not be the modal value. This will specially be so where the difference between the maximum frequency and the frequency preceeding or succeeding it is very small and items are heavily concentrated on either side. Inspection method will also be of no use when the frequencies in the immediate neighbourhood of the highest frequency are very low. In such cases, mode can be determined only by grouping method. Under grouping method, modal value is determined by preparing two tables—(*i*) grouping table and (*ii*) analysis table.

▶ Preparation of Grouping Table

A grouping table has six columns. The various steps in its preparation are as follows:

(*i*) In the column first, the maximum frequency is marked, underlined or put in a circle.

(*ii*) In column second, frequencies are grouped in two's, starting with the first two frequencies of the series.

(*iii*) In the column third, first frequency is left cut and the remaining are grouped in two's.

(*iv*) In fourth column, frequencies are grouped in three's starting with the first three frequencies.

(*v*) In column fifth, leave the first frequency and group the remaining in three's.

(*vi*) In the column sixth, leave the first two frequencies and group the other frequencies in three's.

We mark, underline or circle the maximum frequency in each column. The six columns are to serve as the basis for the preparation of analysis table.

▶ Preparation of Analysis Table

After the preparation of grouping table, the analysis table is prepared. Following steps are followed for its preparation:

(*i*) Put the column numbers on the left hand side.

(*ii*) Put the probable values of the mode on the right hand side.

(*iii*) Now enter into columns the highest frequencies marked in the grouping table.

(*iv*) Take the total of each column to find out the value repeated maximum number of times. This value against which the total is highest is the mode.

Example 61. From the following data, determine the mode by grouping method:

X:	7	8	9	10	11	12	13	14	15	16	17
f:	2	3	6	12	20	24	25	7	5	3	1

Solution:

Grouping Table

X	I (f)	II (1+2)	III (2+3)	IV (1+2+3)	V (2+3+4)	VI (3+4+5)
7	2					
		5				
8	3			11		
			9			
9	6				21	
		18				
10	12					38
			32			
11	20			56		
		44				
12	24				69	
			49			
13	25					56
		32				
14	7			37		
			12			
15	5				15	
		8				
16	3					9
			4			
17	1					

Analysis Table

	Size (X)										
Col. No.	7	8	9	10	11	12	13	14	15	16	17
I							√				
II					√	√					
III						√	√				
IV				√	√	√					
V					√	√	√				
VI						√	√	√			
Total				1	3	5	4	1			

It is clear from the above table that the size 12 occurs the maximum number of times, *i.e.*, 5 times. Thus, the mode (Z) is 12.

- **Continuous Series**

The following steps are taken for the determination of mode in a continuous series:

(*i*) Firstly, modal group is ascertained either by using inspection method or by grouping method. The procedure to be followed will remain the same as in discrete series.

(*ii*) After determining the modal group, mode can be found by using the following formula:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Where, Z = Mode, l_1 = lower limit of the modal class, f_1 = frequency of the modal class, f_0 = frequency of the pre modal class, f_2 = frequency of the post modal class, i = size of the modal group.

Aliter: The above mentioned formula can also be written as:

$$Z = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

Where, l_1=lower limit of the modal class, $\Delta_1 = |f_1 - f_0|$ and $\Delta_2 = |f_1 - f_2|$.

For the calculation of Δ_1 and Δ_2, signs are ignored.

Note 1. If the first class is the modal class, then f_0 is taken as zero. Similarly, if the last class is modal class, then f_2 is taken as zero.

Note 2. If the modal value lies outside the model class, the following formula is used to calculate the mode:

$$Z = l_1 + \frac{f_2}{f_0 + f_2} \times i$$

Note 3. If mode is ill-defined, then we use the following formula $Z = 3M - 2\overline{X}$

where, M = Median, $\overline{X}$ = Mean.

Example 62. Calculate the mode from the following data:

Wages (in Rs.):	0—5	5—10	10—15	15—20	20—25	25—30	30—35
No. of workers:	3	7	15	30	20	10	5

Solution: Since, the series is regular, we may not do grouping for the location of model group. By inspection, modal class is 15—20

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Here, $l_1 = 15, f_1 = 30, f_0 = 15, f_2 = 20, i = 5$

Substituting the values, we get

$$Z = 15 + \frac{30 - 15}{2(30) - 15 - 20} \times 5 = 15 + \frac{15}{25} \times 5 = 18$$

Thus, mode = 18.

Aliter: Mode can also be calculated by using the formula:

$$Z = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

Here, $l_1 = 15$, $\Delta_1 = |f_1 - f_0| = |30 - 15| = 15, \Delta_2 = |f_1 - f_2| = |30 - 20| = 10, i = 5$

$$\therefore \quad Z = 15 + \frac{15}{15+10} \times 5 = 15 + \frac{15}{25} \times 5 = 15 + 3 = 18$$

$$\therefore \quad Z = 18$$

Cumulative Frequency Series

Example 63. Calculate mode from the following data:

Marks between	No. of students
10 and 15	4
10 and 20	12
10 and 25	30
10 and 30	60
10 and 35	80
10 and 40	90
10 and 45	95
10 and 50	97

Solution: Since, this is a cumulative frequency series, we first convert it into simple frequency series:

Marks	f
10—15	4
15—20	8
20—25	18 f_0
25—30	30 f_1
30—35	20 f_2
35—40	10
40—45	5
45—50	2

By inspection, the modal class is 25—30

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$Z = 25 + \frac{30-18}{2(30)-18-20} \times 5 = 25 + \frac{12}{60-18-20} \times 5$$

$$= 25 + \frac{60}{22} = 25 + 2.73 = 27.73$$

$\therefore \quad Z = 27.73$

Aliter: Mode can also be located by using the formula:

$$Z = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

Here, $\Delta_1 = |f_1 - f_0| = |30 - 18| = 12$, $\Delta_2 = |f_1 - f_2| = |30 - 20| = 10$

$$\therefore \quad Z = 25 + \frac{12}{12+10} \times 5 = 25 + \frac{60}{22} = 25 + 2.73 = 27.73$$

$\therefore \quad Z = 27.73$

Inclusive Series

Example 64. Calculate the mode from the following data:

Class:	20—24	25—29	30—34	35—39	40—44	45—49	50—54	55—59
Frequency:	3	5	10	20	12	6	3	1

Solution: Since we are given inclusive class intervals, we first convert it into exclusive one.

Class	Frequency (f)
19.5—24.5	3
24.5—29.5	5
29.5—34.5	10 f_0
34.5—39.5	20 f_1
39.5—44.5	12 f_2
44.5–49.5	6
49.5—54.5	3
54.5—59.5	1

By inspection, modal class is 34.5—39.5

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 34.5 + \frac{20-10}{2 \times 20 - 10 - 12} \times 5$$

$$= 34.5 + \frac{10}{40-22} \times 5 = 34.5 + \frac{50}{18} = 34.5 + 2.77 = 37.27$$

$\therefore \quad Z = 37.27$

Unequal Class Intervals

Example 65. Calculate the mode from the following data:

X:	0—5	5—10	10—20	20—25	25—30	30—40	40—44	44—50	50—70
f:	4	8	10	9	13	30	6	9	12

Solution: Since the class intervals are unequal, we make them equal before calculating th value of mode.

X	f
0—10	$4 + 8 = 12$
10—20	10
20—30	$9 + 13 = 22$ f_1
30—40	30 f_0
40—50	$6 + 9 = 15$ f_2
50—60	6
60—70	6

By inspection, modal class is 30—40

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 30 + \frac{30 - 22}{2 \times 30 - 22 - 15} \times 10$$

$$= 30 + \frac{8}{60 - 22 - 15} \times 10$$

$$= 30 + \frac{80}{23} = 30 + 3.47 = 33.47$$

Mid-value Series

Example 66. Calculate the mode from the following data:

Class:	1	2	3	4	5	6	7	8	9	10
Frequency:	8	6	10	12	20	12	5	3	2	4

Solution: Since we are given mid-values (*i.e.*, central size), first we determine the lower upper limits of the classes by using the formula:

$l_1 = m - i/2 \quad l_2 = m + i/2$ where, i = difference between two mid-valu

m = mid-values (or central size)

Here, $i = 1$, $m = 1$,

$\therefore\ l_1 = 1 - 1/2 = 0.5,\ l_2 = 1 + 1/2 = 1.5$

Hence the first class would be 0.5—1.5

Class	Frequency (f)
0.5—1.5	8
1.5—2.5	6
2.5—3.5	10
3.5—4.5	12
4.5—5.5	$20 = f_1$
5.5—6.5	12
6.5—7.5	5
7.5—8.5	3
8.5—9.5	2
9.5—10.5	4

By inspection, modal class is 4.5—5.5

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Here, $l_1 = 4.5, f_1 = 20, f_0 = 12, f_2 = 12, i = 1$

$$\therefore \quad Z = 4.5 + \frac{20-12}{2\times20-12-12} \times 1$$

$$= 4.5 + \frac{8}{16}$$

$$= 4.5 + 0.5$$

$$= 5.0$$

Hence, $Z = 5$.

Bi-Modal Series

Example 67. Calculate the value of mode from the following data:

Marks:	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—90
No. of students:	4	6	20	32	33	17	8	2

Solution: By inspection, it is difficult to say which is the modal class. Hence we use grouping method to locate the modal class.

Grouping Table

Marks	I (f)	II (1+2)	III (2+3)	IV (1+2+3)	V (2+3+4)	VI (3+4+5)
10—20	4					
		10		30		
20—30	6					
			26			
30—40	20				58	
		52				85
40—50	32					
			65	82		
50—60	33					
		50				
60—70	17				58	
			25			
70—80	8					27
		10				
80—90	2					

Analysis Table

Col. No.	10—20	20—30	30—40	40—50	50—60	60—70	70—80	80—
I					√			
II			√	√				
III				√	√			
IV				√	√	√		
V		√	√	√	√	√	√	
VI			√	√	√			
Total		1	3	5	5	2	1	

As the maximum frequency occurs twice, it is a bi-modal series. Hence, mode is determin applying the formula:

$$\text{Mode} = 3\ \text{Median} - 2\ \text{Mean}$$

Calculation of Mean and Median

Marks	f	M.V. (m)	A = 55 d	$d' = d/10$	fd'	$c.f.$
10—20	4	15	–40	–4	–16	4
20—30	6	25	–30	–3	–18	10
30—40	20	35	–20	–2	–40	30
40—50	32	45	–10	–1	–32	62
50—60	33	55A	0	0	0	95
60—70	17	65	+10	+1	17	112
70—80	8	75	+20	+2	16	120
80—90	2	85	+30	+3	6	122
	$N = 122$				$\Sigma fd' = -67$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

$$= 55 - \frac{67}{122} \times 10 = 55 - \frac{670}{122}$$

$$= 49.51$$

Median item = Size of $\frac{N}{2}$th item = $\frac{122}{2}$ = 61th item which lies in 40—50.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 40 + \frac{61-30}{32} \times 10$$

$$M = 40 + \frac{310}{32} = 40 + 9.69 = 49.69$$

Thus, $\overline{X} = 49.51$, $M = 49.69$

Now, $Z = 3M - 2\overline{X}$

$= 3(49.69) - 2(49.51)$

$\therefore$ $Z = 149.07 - 99.02 = 50.05$

Failure of Formula

Example 68. Calculate mode of the following data:

Wages (Rs.):	25—35	35—45	45—55	55—65	65—75
No.of workers:	4	44	38	28	6
Wages (Rs.):	75—85	85—95	95—105	105—115	
No.of workers:	8	12	2	2	

Solution:

Grouping Table

Wages Rs.	I (*f*)	II (1+2)	III (2+3)	IV (1+2+3)	V (2+3+4)	VI (3+4+5)
25—35	4					
		48				
35—45	44			86		
			82			
45—55	38				110	
		66				
55—65	28					72
			34			
65—75	6			42		
		14				
75—85	8				26	
			20			
85—95	12					22
		14				
95—105	2			16		
			4			
105—115	2					

Analysis Table

Col. No.	25 – 35	35 – 45	45 – 55	55 – 65	65 – 75	75 – 85	85 – 95	95 – 105	105 – 115
I		√							
II			√	√					
III		√	√						
IV	√	√	√						
V		√	√	√					
VI			√	√	√				
Total	1	4	5	3	1				

As the maximum frequency occurs in the class 45 – 55, hence model class is 45 – 55. Appl the formula:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 45 + \frac{38 - 44}{2 \times 38 - 44 - 28} \times 10$$

$$= 45 + \frac{(-6)}{76 - 44 - 28} \times 10$$

$$= 45 - \frac{60}{4} = 45 - 15 = 30$$

The value of the mode lies outside the model class. In such a case, the following alternative formula is used:

$$Z = l_1 + \frac{f_2}{f_0 + f_2} \times i$$

$$= 45 + \frac{28}{44+28} \times 10 = 45 + \frac{28}{72} \times 10$$

$$= 45 + 3.89 = 48.49$$

IMPORTANT TYPICAL EXAMPLES

Example 69. An incomplete distribution families according to their expenditure per week is given below. The median and mode for the distribution are Rs. 25 and Rs. 24 respectively. Calculate the missing frequencies:

Expenditure:	0—10	10—20	20—30	30—40	40—50
No. of families:	14	?	27	?	15

Solution: Two frequencies are missing and let the missing frequencies be x and y respectively.

Expenditure	**No. of families**	***c.f.***
0—10	14	14
10—20	x	$14+x$
20—30	27	$41+x$
30—40	y	$41+x+y$
40—50	15	$56+x+y$
	$N = 56+x+y$	

(*i*) As $M = 25$ $\therefore$ Median class is 20—30

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$\Rightarrow \quad 25 = 20 + \frac{\left(\frac{56+x+y}{2} - 14 - x\right)}{27} \times 10$$

$$\Rightarrow \quad \frac{135}{10} = \frac{28 - x + y}{2} \quad \text{or} \quad 27 = 28 - x + y$$

$$\Rightarrow \quad x - y = 1 \Rightarrow y = x - 1$$

(*ii*) As $Z = 24$ $\therefore$ Modal class is 20—30

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$24 = 20 + \frac{27 - x}{54 - x - y} \times 10$$

$$4(54 - x - y) = (27 - x)10$$

$$2[54 - x - (x - 1)] = 5(27 - x) \qquad \text{[Using } y = x - 1\text{]}$$

$$108 - 2x - 2x + 2 = 135 - 5x$$

$\Rightarrow$ $x = 25$

$\therefore$ $y = x - 1$

$\therefore$ $y = 24$

$\therefore$ The missing frequencies are 25 and 24.

Example 70. The median and mode of the following wage distribution of 230 workers are kn to be Rs. 33.5 and Rs. 34 respectively. Three frequency values from the however are missing. Find the missing values:

Wages in Rs.:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	To
No. of workers:	4	16	—	—	—	6	4	2

Solution: Let the missing frequencies be denoted by x, y and z.

Wages (Rs.)	f	c.f.
0—10	4	4
10—20	16	20
20—30	x	$20+x$
30—40	y	$20+x+y$
40—50	z	$20+x+y+z$
50—60	6	$26+x+y+z$
60—70	4	$30+x+y+z$
	$N = 230$	

Now, $230 = 30 + x + y + z$

$\therefore$ $x + y + z = 230 - 30 = 200$

and $z = 200 - x - y$

Since median and mode are 33.5 and 34, they both lie in the class 30–40

$$\text{Median item} = \text{Size of } \frac{N}{2} \text{ th item}$$

$$= \frac{230}{2} = 115\text{th item.}$$

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$33.5 = 30 + \frac{115-(20+x)}{y} \times 10$$

$$3.5 = \frac{(95-x)}{y} \times 10$$

$$3.5y + 10x = 950 \qquad ...(ii)$$

Now Mode $=34$, which lies in class 30–40

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$34 = 30 + \frac{y-x}{2y-x-z} \times 10$$

$$4 = \frac{y-x}{2y-x-(200-x-y)} \times 10$$

$$4 = \frac{y-x}{2y-x-200+x+y} \times 10$$

$$\Rightarrow \quad 4 = \frac{y-x}{3y-200} \times 10$$

$$4(3y-200) = 10y - 10x$$

$$\Rightarrow \quad 2y + 10x = 800 \qquad ...(iii)$$

Solving (*ii*) and (*iii*),

$$3.5y + 10x = 950$$

$$2y + 10x = 800 \qquad \text{[Subtracting]}$$

$$- \quad - \quad -$$

$$1.5y = 150$$

$$\therefore \quad y = \frac{150}{1.5} = 100$$

Put the value of $y = 100$ in (*ii*)

$$3.5(100) + 10x = 950$$

$$350 + 10x = 950$$

$$10x = 600$$

$$x = 60$$

Now, $z = 200 - x - y$

$$= 200 - 60 - 100 = 40$$

Thus, $x = 60, y = 100, z = 40$.

Example 71. Calculate the mode from the following data:

Class:	100—200	200—300	300—400	400—5000	500—600
f:	27	9	7	3	2

Solution:

Class	f
100—200	27 (f_1)
200—300	9 (f_2)
300—400	7
400-500	3
500—600	2

By inspection, modal class is 100—200

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 100 + \frac{27-0}{2\times27-0-9} \times 100 \quad [\text{Here, } f_0 = 0]$$

$$= 100 + \frac{27}{54-9} \times 100 = 100 + \frac{2700}{45}$$

$$= 100 + 60 = 160$$

Example 72. Find the mode from the following data:

Class:	155—157	157—159	159—161	161—163	163—165
f:	4	8	26	53	89

Solution:

Class	f
155—157	4
157—159	8
159—161	26
161—163	53 (f_0)
163—165	89 (f_1)

By inspection, modal class is 163—165. [Here, $f_2 = 0$]

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 163 + \frac{89-53}{2\times89-53-0} \times 2$$

$$= 163 + \frac{72}{125} = 163.576$$

Graphical Location of Mode

The value of mode in a frequency distribution can be located graphically by means of histogram. It involves the following steps:

(*i*) Present the given data in the form of a histogram.

(*ii*) Find the rectangle whose height is the highest. This will be the modal class.

(*iii*) Join the top corners of the modal rectangle with immediately next corners of the adjacent rectangles.

(*iv*) Locate the point where the joining lines intersect with each other.

(*v*) Then draw a perpendicular line from this point of intersection on the X-axis.

(*vi*) The point where this perpendicular line meets the X-axis gives us the value of the mode.

The following examples would illustrate the procedure.

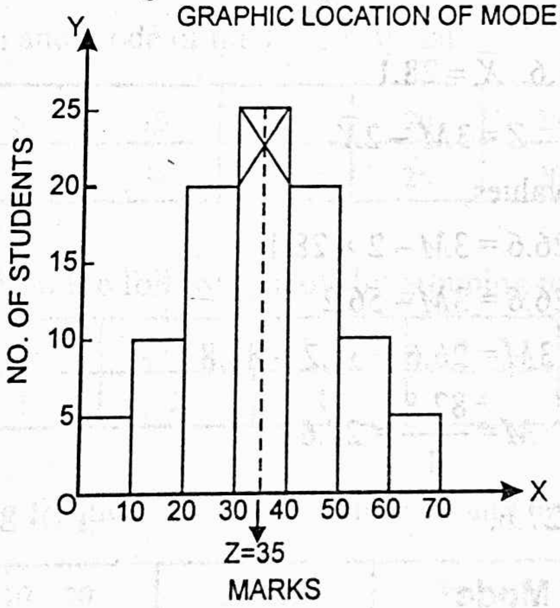
Example 73. Determine the value of mode graphically from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
No. of students:	5	10	20	25	20	10	5

Verify the result by mathematical calculations.

Solution: First we draw the histogram of the given data which is given below:

GRAPHIC LOCATION OF MODE



It is clear from the histogram that the value of the mode is 35.

Calculation of Mode Using Formula

By inspection, mode lies in the class 30—40

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Substituting the values, we get

$$= 30 + \frac{25-20}{50-20-20} \times 10 = 30 + \frac{5}{10} \times 10$$

$$= 35$$

Empirical Relation between Mean, Median and Mode

In a moderately asymmetrical distribution, the difference between $\overline{X}$ and Z is 3 times the difference between $\overline{X}$ and M.

$$\overline{X} - Z = 3\,(\overline{X} - M)$$

This equation can be expressed as

(a) $Z = 3M - 2\overline{X}$, (b) $M = \frac{1}{3}(2\overline{X} - Z)$, and (c) $\overline{X} = \frac{1}{2}(3M - Z)$.

Example 74. In an asymmetrical distribution, the arithmetic mean and median are respectively 30 and 32. Calculate the mode.

Solution: As we know

$$Z = 3M - 2\overline{X}$$

Given: $M = 32$, $\overline{X} = 30$

$$Z = 3 \times 32 - 2 \times 30$$
$$= 96 - 60 = 36$$

$\therefore$ $Z = 26$

Example 75. The mode and mean are 26.6 and 28.1 respectively in a moderately skewed distribution. Find out the median.

Solution: Given: $Z = 26.6$, $\overline{X} = 28.1$

$$Z = 3M - 2\overline{X}$$

Substituting the values,

$$26.6 = 3M - 2 \times 28.1$$

$\Rightarrow$ $26.6 = 3M - 56.2$

$\therefore$ $3M = 26.6 + 56.2 = 82.8$

$\Rightarrow$ $M = \frac{82.8}{3} = 27.6$

Median value is 27.6.

Merits and Demerits of Mode

Merits:

(*i*) It is easy to understand and simple to calculate.

(*ii*) It is not affected by extreme values.

(*iii*) It can be located graphically with the help of histogram.

(*iv*) It can be easily calculated in case of open ended classes.

(*v*) All the frequencies are not needed for its calculation.

(*vi*) It is true representative of frequency distribution, since it is the value which occurs most frequently.

Demerits:

(*i*) It is not suited to algebraic treatment.

(*ii*) There can be bi-modal frequency series.

(*iii*) It is not based on all the items of the series.

(*iv*) It is not rigidly defined.

(*v*) It has no mathematical property.

EXERCISE 5.7

1. Calculate the mode of the following data:
   8, 10, 5, 8, 12, 7, 8, 9, 11, 7

   [Ans. $Z = 8$]
2. Calculate the mode of the following data:
   (*i*) 10, 15, 20, 25, 30, 35
   (*ii*) 7, 8, 10, 15, 10, 22, 20, 26, 20, 34, 20, 6, 10

   [Ans. (*i*) No Mode (*ii*) Bi-modal series]
3. Calculate the mode (by grouping method) of the following data:

Wages (in Rs.):	20	25	30	35	40	45	50	55
No. of workers:	1	3	5	9	14	10	6	4

[Ans. Z = 40]

4. Calculate the median and mode of the following data:

Size:	5	10	15	20	25	30	35	40
Frequency:	3	7	15	25	20	19	16	4

[Ans. M = 25, Z = 20]

5. Calculate the mode from the following table by grouping method:

Value:	0—5	5—10	10—15	15—20	20—25	25—30	30—35
Frequency:	1	2	10	4	10	9	2

[Ans. Z = 24.83]

6. Determine the missing frequencies when mode = 36 and total frequency is = 30.:

C.I :	10 – 20	20 – 30	30 – 40	40 – 50	50 – 60
f:	—	5	12	—	2

[Ans. $f_1 = 3.67 \approx 4, f_2 = 7.33 \approx 7$, the frequencies cannot be in fractions]

7. Calculate Median and Mode from the following:

Income between (Rs.)	No. of Persons
100 and 200	15
100 and 300	33
100 and 400	63
100 and 500	83
100 and 600	100

[Ans. M = 356.67, Z = 354.54]

8. Calculate the median and mode from the following data:

Size:	6—10	11—15	16—20	21—25	26—30
Frequency:	20	30	50	40	10

[Ans. M = 18, Z=18.83

9. Calculate median and mode (by grouping method) from the following series:

Central Size:	5	10	15	20	25	30	35	40	45
Frequency:	7	13	19	24	32	28	17	8	6

[Ans. M=24.68, Z=25.83

10. The frequency distribution of marks obtained by 60 students of a class in a college is given below:

Marks:	30—34	35—39	40—44	45—49	50—54	55—59	60—64
No. of students:	3	5	12	18	14	16	2

(*i*) Draw Histogram for the distribution and find the modal value.

(*ii*) Draw a cumulative frequency curve and find the marks limits of the middle 50% students

[Ans. (*i*) Z=47.5, (*ii*) Q_1 =42.5, Q_3 = 52, $Q_3 - Q_1$ = 9.5

11. The median and mode of the following wage distribution of 230 persons are known to be Rs. 335 and Rs. 340 respectively. Three frequency values from the table are, however, missing. Find out the missing frequencies:

Wages (Rs.):	0—100	100—200	200—300	300—400	400—500	500—600	600—700
No. of persons:	4	16	60	–	–	–	4

[**Hint:** See Example 70] [Ans. 100, 40,

12. Find out missing frequencies in the following incomplete distribution:-

Class Interval :	0—10	10—20	20—30	30—40	40—50
Frequency:	3	?	20	12	?

The values of Median and Mode are 27 and 26 respectively. [Ans. f_1 = 8, f_2 = 7, N =

13. Calculate the mode (by grouping method) from the following data:

Monthly wages:	200—250	250—300	300—350	350—400	400—450	450—500	500—550	550—600
No. of workers:	4	6	20	32	33	17	8	2

[**Hints:** $Z = 3M - 2\overline{X}$] [Ans. Bi-modal series, $\overline{X}$=397.6, M=398.4, Z=4

14. If the mode and mean of a moderately asymmetrical series are 16 and 15.6 respectively, what would be its most probable mean?

[Ans. M=15.

15. If the mean and median of a moderately asymmetrical series are 26.8 and 27.9 respectively, what would be its most probable mode?

[Ans. Z=30

COMBINED EXAMPLES ON MEAN, MEDIAN AND MODE

Example 76. From the data given below, find the mean, median and mode:

Variable	Frequency
10—13	8
13—16	15
16—19	27
19—22	51
22—25	75
25—28	54
28—31	36
31—34	18
34—37	9
37—40	7

Solution: **Calculation of Mean, Median and Mode**

Variable	f	M.V. (m)	d	$d' = \frac{d}{3}$	fd'	c.f.
10—13	8	11.5	–12	–4	–32	8
13—16	15	14.5	–9	–3	–45	23
16—19	27	17.5	–6	–2	–54	50
19—22	51	20.5	–3	–1	–51	101
22—25	75	23.5 A	0	0	0	176
25—28	54	26.5	+3	+1	54	230
28—31	36	29.5	+6	+2	72	266
31—34	18	32.5	+9	+3	54	284
34—37	9	35.5	+12	+4	36	293
37—40	7	38.5	+15	+5	35	300
	N = 300				$\Sigma fd' = 69$	

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times i = 23.5 + \frac{69}{300} \times 3 = 23.5 + 0.69 = 24.19$$

$$\text{Median} = \text{Size of } \frac{N}{2} \text{ th item} = \frac{300}{2} = 150 \text{ th item.}$$

Median lies in the class 22—25

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 22 + \frac{150 - 101}{75} \times 3 = 22 + 1.96 = 23.96$$

By inspection, mode lies in the class 22—25

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 22 + \frac{75 - 51}{150 - 51 - 54} \times 3$$

$$= 22 + \frac{24 \times 3}{45} = 22 + 1.6 = 23.6$$

Thus, $\overline{X} = 24.19$, M = 23.96, Z = 23.6

Inclusive Series

Example 77. From the data given below, find the mean, median and mode:

Marks	No. of students
1—5	7
6—10	10
11—15	16
16—20	30
21—25	24
26—30	17
31—35	10
36—40	5
41—45	1

Solution: The given data is in inclusive form. Firstly we convert it into exclusive form.

Marks	f	M.V. (m)	d	$d' = \frac{d}{5}$	fd'	c.f.
0.5—5.5	7	3	–20	–4	–28	7
5.5—10.5	10	8	–15	–3	–30	17
10.5—15.5	16	13	–10	–2	–32	33
15.5—20.5	30	18	–5	–1	–30	63
20.5—25.5	24	23 = A	0	0	0	87
25.5—30.5	17	28	+5	+1	+17	104
30.5—35.5	10	33	+10	+2	+20	114
35.5—40.5	5	38	+15	+3	+15	119
40.5—45.5	1	43	+20	+4	+4	120
	N = 120				$\Sigma fd' = -64$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 23 + \frac{(-64)}{120} \times 5 = 23 - \frac{320}{120} = 23 - 2.67 = 2033$$

$$M = \text{Size of} \left(\frac{N}{2}\right)\text{th item} = \frac{120}{2} = 60\text{th item.}$$

Median lies in the class 15.5—20.5

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 15.5 + \frac{60 - 33}{30} \times 50$$

$$= 15.5 + \frac{27 \times 5}{30} = 15.5 + \frac{135}{30} = 15.5 + 4.5 = 20$$

By inspection, mode lies in the class 15.5—20.5

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 15.5 + \frac{30 - 16}{60 - 16 - 24} \times 5$$

$$= 15.5 + \frac{70}{20} = 15.5 + 3.5 = 19$$

Thus, $\overline{X} = 20.33$, M = 20, Z = 19.

Example 78. Calculate mean, median and mode for the distribution from the data below:

Central Size	f
35	18
45	37
55	45
65	27
75	15
85	8

Solution: As the central size (mid-value) is given, we have to find the class intervals.

Calculation of Mean, Median and Mode

Classes	f	M.V. (m)	d	$d' = \frac{d}{10}$	fd'	$c.f.$
30—40	18	35	−20	−2	−36	18
40—50	37	45	−10	−1	−37	55
50—60	45	55 A	0	0	0	100
60—70	27	65	+10	+1	27	127
70—80	15	75	+20	+2	30	142
80—90	8	85	+30	+3	24	150
	N = 150				$\Sigma fd' = 8$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 55 + \frac{8}{150} \times 10 = 55 + 0.53 = 55.53$$

$$M = \text{Size of} \left(\frac{N}{2}\right) \text{th item} = \frac{150}{2} = 75\text{th item}$$

Median lies in the class 50—60

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 50 + \frac{75 - 55}{45} \times 10$$

$$= 50 + \frac{200}{45} = 50 + 4.44 = 54.44$$

By inspection, mode lies in the class 50—60

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 50 + \frac{45-37}{90-37-27} \times 10$$

$$= 50 + \frac{8}{26} \times 10 = 50 + \frac{80}{26} = 50 + 3.07 = 53.07$$

Thus, $\overline{X} = 55.53,\ M = 54.44,\ Z = 53.07$

Cumulative Frequency Series

Example 79. The grades of 36 students in an Auditing test are given in the following table:

Grades (Less than):	40	50	60	70	80	90	100
No. of students:	3	7	13	23	29	33	36

Find Mean, Median and Mode.

Solution: The given data is in cumulative frequency form. Firstly we convert it into simple frequency series.

Grades	f	M.V. (m)	d	$d' = \frac{d}{10}$	fd'	c.f.
30—40	3	35	–30	–3	–9	3
40—50	4	45	–20	–2	–8	7
50—60	6	55	–10	–1	–6	13
60—70	10	65	0	0	0	23
70—80	6	75	+10	+1	6	29
80—90	4	85	+20	+2	8	33
90—100	3	95	+30	+3	9	36
	$N = 36$				$\Sigma fd' = 0$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 65 + \frac{0}{36} \times 10 = 65$$

$$\text{Median} = \text{Size of} \left(\frac{N}{2}\right) \text{th item} = \frac{36}{2} = 18\text{th item.}$$

Median lies in the class 60—70

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 60 + \frac{18-13}{10} \times 10 = 60 + 5 = 65$$

By inspection mode lies in the class 60—70.

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 60 + \frac{10-6}{20-6-6} \times 10 = 60 + \frac{40}{8} = 60 + 5 = 65$$

Thus, Mean = 65, Median = 65 and Mode = 65.

IMPORTANT TYPICAL EXAMPLES ON MEAN, MEDIAN AND MODE

Example 80. Calculate mean, median and mode from the following data:

Marks:	10—20	10—30	10—40	10—50	10—60	10—70	10—80	10—90
No. of students:	4	16	56	97	124	137	146	150

Solution: The given data is in cumulative frequency form. It should first be converted into simple frequency.

Marks	***f***	**M.V. (*m*)**	***d***	***d'***	***fd'***	***c.f.***
10—20	4	15	−30	−3	−12	4
20—30	16 − 4 = 12	25	−20	−2	−24	16
30—40	56 − 16 = 40	35	−10	−1	−40	56
40—50	97 − 56 = 41	45	0	0	0	97
50—60	124 − 97 = 27	55	+10	+1	+27	124
60—70	137 − 124 = 13	65	+20	+2	+26	137
70—80	146 − 137 = 9	75	+30	+3	+27	146
80—90	150 − 146 = 4	85	+40	+4	+16	150
	$N = 150$				$\Sigma fd' = 20$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 45 + \frac{20}{150} \times 10 = 45 + 1.33 = 46.33$$

$$\text{Median} = \text{Size of} \left(\frac{N}{2}\right)\text{th item} = \frac{150}{2} = 75\text{th item.}$$

Median lies in the class 40—50

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 40 + \frac{75 - 56}{41} \times 10 = 40 + \frac{19 \times 10}{41} = 40 + \frac{190}{41}$$

$$= 40 + 4.63 = 44.63$$

By inspection mode lies in the class 40—50

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 40 + \frac{41 - 40}{82 - 40 - 27} \times 10 = 40 + \frac{1 \times 10}{15} = 40 + 0.67 = 40.67$$

Thus, $\overline{X} = 46.33$, $M = 44.63$, $Z = 40.67$

Example 81. Calculate mean, median and mode from the following data:

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

Solution: First rewrite the given data in the form of equal classes and then calculate $\overline{X}$, Z and M.

Class Intervals	f	M.V. (m)	d	d'	fd'	c.f.
0—5	1+2+2=5	2.5	−10	−2	−10	5
5—10	3+5=8	7.5	−5	−1	−8	13
10—15	10	12.5	0	0	0	23
15—20	8	17.5	+5	+1	8	31
20—25	4	22.5	+10	+2	8	35
	$\Sigma f = 35$				$\Sigma fd' = -2$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

$$= 12.5 + \frac{(-2)}{35} \times 5 = 12.5 - 0.285 = 12.215$$

By inspection, modal class is 10—15

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 10 + \frac{10-8}{2\times 10 - 8 - 8} \times 5$$

$$= 10 + \frac{2\times 5}{4} = 10 + 2.5 = 12.5$$

$$\text{Median item} = \frac{N}{2} = \frac{35}{2} = 17.5\text{th item.}$$

Median class is 10—15

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 10 + \frac{17.5 - 13}{10} \times 5 = 10 + \frac{4.5}{10} \times 5$$

$$= 10 + 2.25 = 12.25$$

Thus, $\overline{X} = 12.21,\ Z = 12.5,\ M = 12.25$

Example 82. The arithmetic mean, the mode and the median of a group of 75 observations were calculated to be 27, 34 and 29 respectively. It was later discovered that one observation was wrongly read as 43 instead of the correct value 53. Examine to what extent the calculated values of the three averages will be affected by the error.

Solution: We know $\overline{X} = \frac{\Sigma X}{N}$ or $\Sigma X = N\overline{X}$

$N = 75, \overline{X} = 27$

$\therefore \quad \Sigma X = 75 \times 27 = 2025$

Corrected $\Sigma X = 2025 - 43 + 53 = 2035$

Corrected $\overline{X} = \frac{2035}{75} = 27.13$

Thus, the value of mean is affected. But the values of median and mode will not be affected by this error because these are positional averages and the median value is 29 which is far away from the values 43 and 53. Similarly, the value of mode would not be affected by the error since the modal value 34 is far away from the value 43 and 53.

EXERCISE 5.8

1. Calculate the arithmetic mean, median and mode for the data given below:

Weekly earnings (in Rs.):	50—53	53—56	56—59	59—62	62—65	65—68	68—71	71—74	74—77
No. of persons:	3	8	14	30	36	82	16	10	5

[**Ans.** $\overline{X}$=62.47, M=65.36, Z=66.23]

2. An incomplete distribution is given below:

X:	10—20	20—30	30—40	40—50	50—60	60—70	70—80
f:	12	30	?	65	?	25	18

Given that median value is 46 and $N = 229$
Calculate: (*i*) Missing frequencies
(*ii*) Mean, Mode, Quartiles, P_{54} and P_{85} from the completed table.

[**Ans.** $f_1 = 33.5 \simeq 34, f_2 = 45, \overline{X} = 45.82, Z = 46.07,$
$Q_1 = 34.48, Q_3 = 56.72, P_{54} = 47.33, P_{85} = 63.46$]

3. Compute $\overline{X}, M$ and Z from the following data:

Variable:	10—15	10—20	10—25	10—30	10—35	10—40
Frequency:	5	12	25	35	43	50

[**Ans.** $\overline{X}$=25.5, M=25, Z=23.3]

4. Calculate mean, median and mode from the following data:

X:	0.5	1.0	3.0	5—7	7—10	10—15	15—16	16—20	20—25
f:	2	3	2	4	4	12	6	2	5

[**Hint:** Make equal intervals] [**Ans.** $\overline{X}$=12, M=12.08, Z=12.5]

GEOMETRIC MEAN

Geometric mean is a mathematical average. The geometric mean is defined as the nth root of the product of all the n values of the variable. If there are two items in a series, we take the square root. If there are three items, we take the cube root and so on. Symbolically,

$$G.M. = \sqrt[n]{X_1 . X_2 . X_3 \ldots X_n}$$

Where, $X_1, X_2, X_3, \ldots$, etc., are the various values of the series, n-number of items and G.M. the geometric mean.

For example, if there are two items in a series and their values are 4 and 9, then their geometric mean is the square root of the product of the two items.

Thus, $G.M. = \sqrt{4 \times 9} = \sqrt{36} = 6$

Similarly, if there are three items is a series and their values are 2, 4, 8 respectively, their geometric mean will be the cube root of the product of three items. Thus,

$$G.M. = \sqrt[3]{2 \times 4 \times 8} = \sqrt[3]{64} = 4$$

However, if the number of items in a series is very large, it would be difficult to calculate the geometric mean by using the above formula. In order to facilitate the computation of geometric mean, logarithmic (log) values are used. In logarithmic form, geometric mean can be calculated by using the following formula:

$$G.M. = (X_1 . X_2 . X_3 \ldots X_n)^{\frac{1}{n}}$$

Taking logarithms on both sides,

$$\log G.M. = \frac{1}{n}[\log X_1 + \log X_2 + \ldots \log X_n]$$

$$\log G.M. = \frac{\Sigma \log X}{n}$$

$$\therefore \quad G.M. = \text{Antilog}\left[\frac{\Sigma \log X}{n}\right]$$

Note: The values of logs and antilog can be obtained from the log table and antilog tables.

Calculation of Geometric Mean

Individual Series

The formula used for calculating geometric mean in individual series is:

$$G.M. = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right]$$

Steps for Calculation

(*i*) Find the logarithms (logs) of the given values.

(*ii*) Find the sum total of logs, *i.e.*, $\Sigma \log X$

(*iii*) Divide $\Sigma \log X$ by the number of items (N) and calculate its antilog.

The result will give the value of geometric mean.

Example 83. Calculate geometric mean of the following series:

180, 190, 240, 386, 492, 662

Solution:

Calculation of Geometric Mean

X	$\log X$
180	2.2553
190	2.2788
240	2.3802
386	2.5866
492	2.6920
662	2.8209
$N = 6$	$\Sigma \log X = 15.0138$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left[\frac{15.0138}{6}\right]$$

$$\therefore \quad \text{G.M.} = \text{Antilog}\,[2.5023]$$

$$= 317.9$$

Example 84. Calculate geometric mean of the following series:

2574, 475, 75, 5, .8, .08, .005, .0009.

Solution:

Calculation of Geometric Mean

X	$\log X$
2574	3.4106
475	2.6767
75	1.8751
5	0.6990
0.8	$\bar{1}.9031$
0.08	$\bar{2}.9031$
0.005	$\bar{3}.6990$
0.0009	$\bar{4}.9542$
$N = 8$	$\Sigma \log X = 2.1208$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left(\frac{2.1208}{8}\right)$$

$$= \text{Antilog}\ 0.2651 = 1.841$$

Hence, G.M. = 1.841

Example 85. Calculate geometric mean of the following series:
0.9841, 0.3154, 0.0252, 0.0068, 0.0200, 0.0002, 0.5444, 0.4010

Solution:

Calculation of Geometric Mean

X	$\log X$
0.9841	$\bar{1}.9931$
0.3154	$\bar{1}.4989$
0.0252	$\bar{2}.4014$
0.0068	$\bar{3}.8325$
0.0200	$\bar{2}.3010$
0.0002	$\bar{4}.3010$
0.5444	$\bar{1}.7359$
0.4010	$\bar{1}.6031$
	$\Sigma \log X = \overline{11}.6669$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left[\frac{\overline{11}.6669}{8}\right] - \text{Antilog}\left[\frac{\overline{16}+5.6669}{8}\right]$$

$$= \text{Antilog}\,[\bar{2}.7084] = 0.0511$$

Discrete Series

The formula used for calculating geometric mean in discrete series is:

$$\textbf{G.M.} = \textbf{Antilog}\left[\frac{\Sigma f \log X}{N}\right]$$

Steps for Calculation

(1) Find the logarithms (logs) of the given values, *i.e.*, $\log X$.

(2) Multiply the frequency of each item with the log $x$and obtain the total $\Sigma f \log X$.

(3) Divide $\Sigma f \log X$ by the total number of frequency (*i.e.*, $\Sigma f = N$) and calculate its antilog. The result will give the value of geometric mean.

Example 86. From the following data, calculate geometric mean:

Size of items (X):	6	7	8	9	10	11	12
Frequency (f):	8	12	18	26	16	12	8

Solution:

Calculation of Geometric Mean

X	f	$\log X$	$f \log X$
6	8	0.7782	6.2256
7	12	0.8451	10.1412
8	18	0.9031	16.2558
9	26	0.9542	24.8092
10	16	1.0000	16.0000
11	12	1.0414	12.4968
12	8	1.0792	8.6336
	$n = 100$		$\Sigma f \log X = 94.5622$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma f \log X}{N}\right] = \text{Antilog}\left[\frac{94.5622}{100}\right]$$

$$= \text{Antilog } (0.945622) = 8.822$$

$$\therefore \quad \text{G.M.} = 8.822$$

Continuous Series

The formula used for calculating geometric mean in continuous series is:

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

Where, m = mid-value of various classes, N = Total frequency.

Steps for Calculation

(1) Find the logarithms (logs) of the mid-values of the various classes.

(2) Multiply the frequency of each class with log m and obtain the total $\Sigma f \log m$.

(3) Divide $\Sigma f \log m$ by total number of frequency (*i.e.*, $\Sigma f = N$) and calculate its antilog.

Example 87. Calculate geometric mean of the data given below:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	3	4	6	3	2

Solution:

Calculation of Geometric Mean

Marks	M.V. (m)	f	log m	f log m
0—10	5	3	0.6990	2.0970
10—20	15	4	1.1761	4.7044
20—30	25	6	1.3979	8.3874
30—40	35	3	1.5441	4.6323
40—50	45	2	1.6532	3.3064
		$N = 18$		$\Sigma f \log m = 23.1275$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma f \log m}{N}\right] = \text{Antilog}\left[\frac{23.1275}{18}\right]$$

$$= \text{Antilog } [1.28486] = 19.27$$

ombined Geometric Mean

If G_1 and G_2 are the G.M. of two groups having N_1 and N_2 items, then the G.M. of the combined oup is given by:

$$G_{12} = \text{Antilog}\left[\frac{N_1 \log G_1 + N_2 \log G_2}{N_1 + N_2}\right]$$

te: The formula can be extended to three or more groups.

Example 88. The G.M. of a sample of 10 items was found to be 20 and that of a sample of 20 items was found to be 15. Find the combined G.M.

Solution: Here, $N_1 = 10,\ G_1 = 20$

$N_2 = 20,\ G_2 = 15$

$$G_{12} = \text{Antilog}\left[\frac{N_1 \log G_1 + N_2 \log G_2}{N_1 + N_2}\right]$$

$$= \text{Antilog}\left[\frac{10 \times \log 20 + 20 \times \log 15}{10 + 20}\right]$$

$$= \text{Antilog}\left[\frac{10(1.3010) + 20(1.1761)}{30}\right]$$

$$= \text{Antilog}\left[\frac{13.01 + 23.52}{30}\right] = \text{Antilog}\left[\frac{36.53}{30}\right]$$

$$= \text{Antilog}\,[1.2177] = 16.51$$

IMPORTANT TYPICAL EXAMPLES

Example 89. The G.M.of 10 observations on a certain variable was calculated as 16.2. It was later discovered that one of the observations was wrongly recorded as 12.9, in fact, it was 21.9. Apply appropriate correction and calculate the correct G.M.

Solution: Let $x_1\ x_2 \ldots x_{10}$ be ten items.

$$\therefore \quad G.M. = \sqrt[10]{x_1 x_2 \ldots x_{10}} = (x_1 x_2 \ldots x_{10})^{\frac{1}{10}}$$

$$\Rightarrow \quad (16.2)^{10} = x_1 x_2 \ldots x_{10}$$

$$\Rightarrow \quad x_1 x_2 \ldots x_{10} = (16.2)^{10}$$

$$\text{Corrected } x_1 x_2 \ldots x_{10} = (16.2)^{10} \times \frac{\text{Correct Item}}{\text{Incorrect Item}} = (16.2)^{10} \times \frac{21.9}{12.9}$$

$$\text{G.M. (Corrected)} = [\text{Corrected } (x_1 x_2 \ldots x_{10})]^{\frac{1}{10}} = \left[(16.2)^{10} \times \frac{21.9}{12.9}\right]^{\frac{1}{10}}$$

$$\text{G.M. log (Corrected)} = \frac{1}{10}[10 \log 16.2 + \log 21.9 - \log 12.9]$$

$$= \frac{1}{10}[12.09 + 1.3404 - 1.1105]$$

$$= 1.23249$$

$$\therefore \quad \text{Corrected G.M.} = \text{Antilog}\,[1.23249]$$

$$= 17.08$$

Special Uses of Geometric Mean

Geometric mean is specially useful in the following cases:

(1) Geometric mean is generally used to find the average percentage increase in population, prices, sales, production or national income, etc.

(2) As geometric mean measures relative change, it finds special application in index numbers.

(3) This average is most suitable when large weights have to be given to small items and small weights to large items.

The following examples illustrate the uses of geometric mean:

Example 90. Find the average rate of increase in population which in the first decade has increased by 20%, in the second decade by 30% and in the third decade by 40%.

Solution: Suppose the population at the beginning of the decade is 100.

Calculation of Geometric Mean

Decade	Per cent rise	Population at the end of decade (X)	log x
1st	20	120	2.0792
2nd	30	130	2.1139
3rd	40	140	2.1461
$N = 3$			$\Sigma \log x = 6.3392$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left[\frac{6.3392}{3}\right]$$

$$= \text{Antilog } (2.1131) = 129.7$$

The average rate of increase in population is $129.7 - 100 = 29.7$ per cent per decade.

Example 91. The annual rates of growth of an economy over the last five years were: 1.5, 2.7, 3.0, 4.5 and 6.2 per cent respectively. What is the compound rate of growth per annum for the 5 year period?

Solution: Suppose the price at the beginning of the year is 100. Apply the geometric mean:

Years	Annual rate of growth	Growth rate at the end of the year (X)	log X
1	1.5	101.5	2.0064
2	2.7	102.7	2.0116
3	3.0	103.0	2.0128
4	4.5	104.5	2.0191
5	6.2	106.2	2.0261
$N = 5$			$\Sigma \log X = 10.076$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{n}\right] = \text{Antilog}\left[\frac{10.076}{5}\right]$$

$$= \text{Antilog } (2.0152) = 103.5$$

The compound rate of growth per annum $= 103.5 - 100 = 3.5$

Example 92. A machinery depreciates by 40% in the first year, 25% in the second and 10% per annum during the next three years. What is the average rate of depreciation during the whole period?

Solution: Suppose the value of machine in the initial year is 100

Year	Per cent decline	Diminishing value at the end of the year (X)	log X
I	40	100 – 40 = 60	1.7782
II	25	100 – 25 = 75	1.8751
III	10	100 – 10 = 90	1.9542
IV	10	100 – 10 = 90	1.9542
V	10	100 – 10 = 90	1.9542
$N = 5$			$\Sigma \log X = 9.5159$

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right] = \text{Antilog}\left[\frac{9.5159}{5}\right]$$

$$= \text{Antilog}\,[1.9032] = 80.02 = 80$$

Average rate of depreciation = 100 – 80 = 20% per year.

Average Rate of Growth of Population

Geometric mean is also used to compute the average annual per cent increase in population, prices when the values of the variables at the beginning of the first and at the end of the nth period are given. The average annual per cent increase may be computed by applying the formula:

$$P_n = P_0\,(1 + r)^n$$

Where, r = the average rate of growth

n = Number of years

P_0 = Value at the beginning of the period

P_n = Value at the end of the period

Using logarithms, r can be calculated as:

$$r = \sqrt[n]{\frac{P_n}{P_0}} - 1$$

or

$$r = \text{Antilog}\left[\frac{\log P_n - \log P_0}{n}\right] - 1$$

Note : The expression $P_n = P_0(1 + n)^n$ is called **compound interest formula.**

Example 93. The population of a country has increased from 84 million in 1983 to 108 million in 1993. Find the annual rate of growth of population.

Solution: Given, P_0 = 84 million, P_n = 108 million.

Let r be the rate of growth of population. Applying the formula,

$$r = \sqrt[n]{\frac{P_n}{P_0}} - 1 = \sqrt[10]{\frac{108}{84}} - 1$$

$$= \text{Antilog}\left[\frac{\log 108 - \log 84}{10}\right] - 1$$

$$= \text{Antilog}\left[\frac{2.0334 - 1.9243}{10}\right] - 1$$

$$= \text{Antilog}\left[\frac{0.1091}{10}\right] - 1$$

$$= \text{Antilog}\,[0.0109] - 1$$

$$= 1.026 - 1 = .026$$

$$= 2.6\%$$

Thus, the annual growth rate of population is 2.6%.

Example 94. The population of a town was 10,000. At first it increased at the rate of 3% per annum for the first 3 years and then it decreased at the rate of 2% per annum for the next 2 years. What will be the population of the town after 5 years?

Solution: Given, $P_0 = 10{,}000, r_1 = 0.03, r_2 = 0.02, n_1 = 3, n_2 = 2, P_n = ?$

$$P_n = P_0(1+r_1)^3(1-r_2)^2 = 10{,}000(1+0.03)^3(1-0.02)^2$$

$$= 10{,}000(1.03)^3(0.98)^2 = 10{,}000(1.0927)(0.9604)$$

$$= 10{,}000 \times 1.0494 = 10{,}494$$

Example 95. The population of a city was 1,00,000 in 1980 and 1,44,000 in 1990. Estimate the population at the middle of 1980-1990.

Solution:

$$P_n = P_0(1+r)^n$$

Here, $n = 10,\ P_0 = 1{,}00{,}000,\ P_{10} = 1{,}44{,}000$

$$144{,}000 = 1{,}00{,}000(1+r)^{10}$$

$$(1+r)^{10} = \frac{144}{100} = \left(\frac{12}{10}\right)^2$$

$$\Rightarrow \quad \left(\frac{12}{10}\right)^2 = [(1+r)^5]^2$$

$$\Rightarrow \quad (1+r)^5 = \frac{12}{10} \qquad \ldots(i)$$

Now,

$$P_5 = P_0(1+r)^5$$

$$P_5 = 1{,}00{,}000(1+r)^5 \qquad \ldots(ii)$$

Put the value of $(1+r)^5$ from (i) in (ii), we get

$$P_5 = 1{,}00{,}000\left(\frac{12}{10}\right)$$

$$P_5 = 1{,}20{,}000$$

Aliter: Population at middle of the decade as

$$\Rightarrow \quad \text{G.M.} = \sqrt{1{,}44{,}000 \times 1{,}00{,}000}$$

$$= \sqrt{14400000000} = 120{,}000$$

Example 96. The population of a capital of a State was 20 lakh in 1981. It went up to 25 lakhs in 1991. What is the average percentage increase per year? What will be the size of population in 2006 if it continued to increase at the same rate?

Solution:

$$P_n = P_0(1+r)^n$$

Here, $n = 10$, $P_0 = 20$, $P_{10} = 25$

$$25 = 20(1+r)^{10}$$

Taking log on both sides,

$$10 \log(1+r) = \log 1.25$$

$$\log(1+r) = \frac{0.0969}{10} = 0.00969$$

$$1 + r = \text{Antilog}\,[0.00969] = 1.023$$

$$r = 1.023 - 1 = 0.023$$

or $$r = 0.023 \times 100 = 2.3\%$$

Population for 2006, $P_0 = 20, r = .023, N = 25$

$$P_n = P_0(1+r)^{25}$$

$$= 20\,(1 + 0.23)^{25}$$

Taking log on both sides,

$$\log P_n = \log 20 + 25 \log 1.023$$

$$\log P_n = 1.3010 + 25 \times 0.0099$$

$$= 1.3010 + 0.2475 = 1.5485$$

$$P_n = \text{Antilog}\,[1.5485]$$

$$= 35.36 \text{ lakhs}$$

Weighted Geometric Mean

Like weighted arithmetic mean, we can also calculate weighted geometric mean by using the following formula:

$$\textbf{Weighted G.M.} = \textbf{Antilog}\left[\frac{\Sigma W \log X}{\Sigma W}\right]$$

Where, $W_1, W_2 \ldots W_n$ are the weights assigned to different values $X_1, X_2 \ldots X_n$.

ample 97. From the following data, calculate weighted geometric mean:

Items	Index No.	Weights
Food	120	7
Rent	110	5
Clothing	125	3
Fuel	105	2
Others	140	3

lution:

Calculation of Weighted G.M.

Items	X	W	$\log X$	$W \cdot \log X$
Food	120	7	2.0792	14.554
Rent	110	5	2.0414	10.2070
Clothing	125	3	2.0969	6.2907
Fuel	105	2	2.0212	4.0424
Others	140	3	2.1461	6.4383
		$\Sigma W = 20$		$\Sigma W \log X = 41.5328$

$$\text{Weighted G.M.} = \text{Antilog}\left[\frac{\Sigma W \log X}{\Sigma W}\right] = \text{Antilog}\left[\frac{41.5328}{20}\right]$$

$$= \text{Antilog } (2.0766) = 119.3$$

Merits and Demerits of Geometric Mean

erits:

(1) It is rigidly defined.

(2) It is based on all the items of the series.

(3) It is suitable for further algebraic treatment.

(4) It gives less weight to large items.

(5) It possesses the merits of sampling stability.

(6) It is the best measure of ratios of change.

(7) It is least affected by extreme items.

emerits:

(1) It is difficult to calculate as one has to make use of logs and antilogs to calculate geometric mean.

(2) Geometric mean cannot be calculated unless the values of all the items are known.

(3) It may be a value which does not exit in the series.

(4) If value of any one item is zero, then geometric mean will also become zero. Again if one item has negative value, G.M. becomes indeterminate.

EXERCISE 5.9

1. Calculate the geometric mean from the following series:
85, 70, 15, 75, 500, 8, 45, 250, 40, 36
[Ans. G.M.=58.03]

2. Calculate the geometric mean from the following series:

X :	8	10	12	14	16	18
f :	6	10	20	8	5	1

[Ans. G.M.=11.71]

3. Find the geometric mean for the data given below:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of persons:	4	8	10	6	7

[Ans. G.M.=22.06]

4. The price of a commodity increased by 5% from 1985 to 1986, 8% from 1986 to 1987 and 77% from 1987 to 1988. The average increase from 1985 to 1988 is quoted as 26% and not 30%. Explain and verify the result.

5. A machine was purchased for Rs. 10 lakh in 2000. An income tax assessee depreciated the machinery of his factory by 20 per cent in each of the first two years and 40 per cent in the third year. How much average depreciation relief should he claim from the taxation department?
[Ans. 27.32%]

6. The geometric mean of 10 observations is 28.6. It was later discovered that one of the observation was recorded 23.4 instead of 32.4. Apply appropriate correction and calculate the correct Geometric mean.
[Hint: See Example 89]
[Ans. 29.54]

7. In 20 years, the population of a town increased from 10,000 to 20,000. Find the average annual rate of growth of population.
[Ans. 3.5%]

8. The price of a commodity increased by 20% in the first year, decreased by 30% in the second year and increased by 40% in the third year. Find the average increase for these three years.
[Ans. 5.56%]

9. If the price of a commodity doubled in a period of 5 years, what is the average percentage increase per year?
[Ans. 14.8%]

10. The number of bacteria in a certain culture was found to be 4×10^6 at noon of one day. At noon the next day the number was 9×10^6. If the number increased at a constant rate per hour, how many bacteria were there at the intervening mid-night?
[Hint: $\sqrt{4 \times 10^6 \times 9 \times 10^6}$]
[Ans. 6×10^6]

11. In 1950 and 1960 the population of US was 151.3 mw and 179.3 mw respectively (*i*) What was the average percentage increase per year (*ii*) Estimate the population in 1954 (*iii*) If the average percentage increase of population per year from 1960 to 1970 is the same as in (*ii*), what would be the population in 1970.
[Ans. (*i*) 1.7% (*ii*) 161.74 mw (*iii*) 212.71]

HARMONIC MEAN

The harmonic mean is a mathematical average. It is based on the reciprocal of the items. The harmonic mean of a series is defined as the reciprocal of the arithmetic average of the reciprocal of the values of its various items. Symbolically,

$$\text{H.M.} = \text{Reciprocal of } \frac{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots+\frac{1}{X_n}}{N}$$

$$= \frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots+\frac{1}{X_n}} = \frac{N}{\Sigma\left(\frac{1}{X}\right)}$$

Calculation of Harmonic Mean

Individual Series

In individual series, harmonic mean is calculated by using the following formula:

$$\text{H.M.} = \frac{N}{\frac{1}{X_1}+\frac{1}{X_2}+\frac{1}{X_3}+\ldots+\frac{1}{X_n}}$$

Where, $X_1, X_2, X_3, \ldots$, etc. refer to the various items of the variable.

Steps for Calculation

(1) Find out the reciprocals of the values of the series and add to these gets $\Sigma\left(\frac{1}{X}\right)$.

(2) Divide the sum total of reciprocals by the number of items, and find out its reciprocals. This gives the value of harmonic mean.

Example 98. Calculate the H.M. of the following:

2, 4, 7, 12, 19

Solution:

Calculation of H.M.

X	Reciprocal $(1/X)$
2	0.5000
4	0.2500
7	0.1429
12	0.0833
19	0.0526
N = 5	$\Sigma\left(\frac{1}{X}\right)=1.0288$

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{1}{X}\right)} = \frac{5}{1.0288} = 4.86$$

Discrete Series

The formula used for calculating harmonic mean in discrete series is:

$$\text{H.M.} = \frac{N}{\Sigma f \times \frac{1}{X}}$$

Steps for Calculation

(1) Divide each frequency by the respective values of the variable.

(2) Obtain the total $\Sigma\left(f \times \frac{1}{X}\right)$

(3) Substitute the value of N and $\Sigma\left(f \times \frac{1}{X}\right)$ in the above formula.

Example 99. The following table gives the marks obtained by students in a class. Calculate the H.M.

Marks:	18	21	30	45
No. of students:	6	12	9	2

Solution: **Calculation of Harmonic Mean**

Marks X	f	f/X
18	6	6/18 = 0.333
21	12	12/21 = 0.571
30	9	9/30 = 0.3000
45	2	2/45 0.0444
	N = 29	$\Sigma(f/X)$=1.248

$$\text{H.M.} = \frac{N}{\Sigma(f/X)} = \frac{29}{1.248} = 23.237$$

Continuous Series

For calculating harmonic mean in continuous series, the following formula is used:

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

Where, m = mid-value of various classes, N = the total frequency.

Steps for Calculation

(1) Obtain the mid-value of each class and denote it by m.

(2) Divide each frequency by the respective mid-value of the class.

(3) Obtain the total $\Sigma\left(f \times \frac{1}{m}\right)$.

(4) Substitute the value of N and $\Sigma\left(f \times \frac{1}{m}\right)$ in the above formula.

Example 100. Calculate the H.M. for the following:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	4	7	28	12	9

Solution: **Calculation of Harmonic Mean**

Class	f	Mid-value (m)	f/m
0—10	4	5	4/5 = 0.8000
10—20	7	15	7/15 = 0.4667
20—30	28	25	28/25 = 1.1200
30—40	12	35	12/35 = 0.3429
40—50	9	45	9/45 = 0.2000
	$N = 60$		$\Sigma\left(f\times\frac{1}{m}\right) = 2.9296$

$$\text{H.M.} = \frac{N}{\Sigma\left(\frac{f}{m}\right)} = \frac{60}{2.9296} = 20.48$$

Weighted Harmonic Mean

If all the values of the variable are not of equal importance, *i.e.*, are of varying importance, then we calculate weighted H.M. which is given as:

$$\textbf{Weighted H.M.} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$$

Where, $W_1, W_2, \ldots W_n$ are the weights of the corresponding values $X_1, X_2, \ldots X_n$ of the variables.

Example 101. Find the weighted H.M. of the items 4, 7, 12, 19, 25 with weights 1, 2, 1, 1, 1 respectively.

Solution: **Calculation of Weighted H.M.**

X	W	W/X
4	1	0.2500
7	2	0.2857
12	1	0.0833
19	1	0.0526
25	1	0.0400
	$\Sigma W = 6$	$\Sigma\left(\frac{W}{X}\right) = 0.7116$

$$\text{Weighted H.M.} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)} = \frac{6}{0.7116} = 8.4317$$

Uses of Harmonic Mean

H.M. is of very limited use. It is useful in finding averages involving speed, time, price and ratios. The following examples illustrate the uses of H.M.:

Example 102. An aeroplane covers the four sides of a field at speeds of 1000, 2000, 3000 and 4000 km per hour respectively. What is the average speed of the plane in its flight around the field?

Solution: In this question simple H.M. is used which is calculated as:

$$\text{Average Speed} = \frac{4}{\frac{1}{1000}+\frac{1}{2000}+\frac{1}{3000}+\frac{1}{4000}}$$

$$= \frac{4}{\frac{12+6+4+3}{12000}} = \frac{4\times 12000}{25} = 1920 \text{ kms/hrs}$$

Example 103. A cyclist covers first three kms at an average speed of 8 km per hour, another 2 kms at a speed of 3 km per hour and the last two kms at a speed of 2 km per hour. Find the average speed for the entire journey and verify your answer.

Solution: In this case weighted harmonic mean is used with distance covered as weight. Denoting the speed by X and the distance covered by W.

Speed km/hr (X)	Distance (W)	$\frac{W}{X}$
8	3	3/8 = 0.375
3	2	2/3 = 0.667
2	2	2/2 = 1.000
	$\Sigma W = 7$	$\Sigma W/X = 2.042$

$$\text{Average speed} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)} = \frac{7}{2.042} = 3.42 \text{ km/hr.}$$

Thus, the average speed for the entire journey is 3.42 km/hr.

Verification:

Distance (km)	Speed (km. per hr.)	Time = $\frac{\textit{Distance}}{\textit{Speed}}$
3	8	3/8 = 0.375 hrs.
2	3	2/3 = 0.667 hrs.
2	2	2/2 = 1 hrs.
$\Sigma D = 7$		$\Sigma T = 2\cdot 042$

$$\text{Average speed} = \frac{\text{Total Distance Travelled}}{\text{Total Time Taken}} = \frac{7}{2\cdot 042} = 3.42 \text{ km/hr.}$$

Hence, verified.

xample 104. A bus runs 20 km at an average speed of 30 km per hour and then runs for 30 minutes at a speed of 60 km per hour and finally runs for 15 minutes at a speed of 20 km per hour. Find the average speed of the bus over the entire journey.

olution: In 60 minutes bus travels = 60 km

In 1 minute, bus travels $= \frac{60}{60}$

In 30 minutes, bus travels $= \frac{60}{60} \times 30 = 30$ km

Similarly, in 15 minutes, distance travelled will be $= \frac{20}{60} \times 15 = 5$ km

Since, the distance travelled is unequal, so the required average speed will be weighted H.M.:

X (Speed in km)	W (Distance)	$1/X$	$W \times \frac{1}{X}$
30	20	1/30	20/30 = 2/3
60	30	1/60	30/60 = 1/2
20	5	1/20	5/20 = 1/4
	$\Sigma W = 55$		$\Sigma\left(\frac{W}{X}\right) = 17/12$

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

$$= \frac{55}{\frac{17}{12}} = \frac{55}{1} \times \frac{12}{17} = 38.82 \text{ km / hr.}$$

Example 105. A man travelled by car for 3 days. He covered 480 kms each day. On the first day he drove for 10 hours at 48 kms an hour, on the second day he drove for 12 hours at 40 kms an hour and on last day he drove for 15 hours at 32 kms. What was his average speed?

Solution: Since, the distance travelled is constant, *i.e.*, 480 kms each day, the appropriate average is the Harmonic Mean.

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

$$= \frac{3}{\frac{1}{48} + \frac{1}{40} + \frac{1}{32}} = \frac{3}{\frac{37}{480}} = \frac{3 \times 480}{37}$$

$$= 38.92 \text{ km/hr.}$$

Example 106. In a certain factory, a unit of work is completed by A in 4 minutes, by B in 5 minutes, by C in 6 minutes, by D in 10 minutes and by E in 12 minutes. What is their average rate of working? What is the average number of units of work completed per minute? At this rate how many units will they complete in a six hour day?

Solution: (*i*) Average rate of working

$$\text{H.M.} = \frac{5}{\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{10}+\frac{1}{12}} = \frac{5}{\frac{15+12+10+6+5}{60}}$$

$$= 5 \times \frac{60}{48} = 6.25 \text{ minutes per unit.}$$

(*ii*) Average number of unit of work completed per minute $= \frac{1}{6.25} = 0.16$ units.

(*iii*) Total units of work completed within 6 hours (360 minutes) by 5 workers

$$= 360 \times 0.16 \times 5 = 288 \text{ units.}$$

Example 107. Typist A can type a letter in 5 minutes, B in 10 minutes and C in 15 minutes. What is the average number of letters typed per hour per typist?

Solution: Average time to type one letter

$$\text{H.M.} = \frac{3}{\frac{1}{5}+\frac{1}{10}+\frac{1}{15}} = \frac{3}{\frac{6+3+2}{30}} = \frac{3 \times 30}{11} = \frac{90}{11}$$

$$= 8.18 \text{ minutes per letter.}$$

Number of letters typed in 60 minutes $= \frac{60}{8.18} = 7.33$

CHOICE BETWEEN HARMONIC MEAN AND ARITHMETIC MEAN

The harmonic mean, like arithmetic mean, is also used in calculating average price. To explain the method of choosing an appropriate average, consider the following example:

Let the price of a commodity be Rs. 3, 4, and 5 per unit in three successive years. If we take A.M. of these prices, *i.e.*, $\frac{3+4+5}{3} = 4$, then it will denote average price when equal quantities of the commodity are purchased in each year. To verify this, let us assume that 10 units of commodity are purchased in each year.

$\therefore$ Total Expenditure on the commodity in 3 years $= 10 \times 3 + 10 \times 4 + 10 \times 5$

Also, average price $= \frac{\text{Total Expenditure}}{\text{Total Quantity Purchased}} = \frac{10\times3+10\times4+10\times5}{10+10+10} = \frac{3+4+5}{3} = \text{Rs. } 4$

which is arithmetic mean of the prices in three years.

Further, if we take harmonic mean of the given prices, *i.e.*, $\frac{3}{\frac{1}{3}+\frac{1}{4}+\frac{1}{5}} = \text{Rs. } 3.27$, it will denote the average price when equal amounts of money are spent on the commodity in three years. To verify this let us assume that Rs. 100 is spent in each year on the purchase of the commodity.

$$\text{Average Price} = \frac{\text{Total Expenditure}}{\text{Total Quantity Purchased}} = \frac{300}{\frac{100}{3}+\frac{100}{4}+\frac{100}{5}} = \frac{3}{\frac{1}{3}+\frac{1}{4}+\frac{1}{5}} = \text{Rs. } 3.27$$

Next, we consider a situation where different quantities are purchased in the three years. Let us assume that 10, 15 and 20 units of the commodity are purchased at price of Rs. 3, 4, and 5 respectively.

$$\text{Average Price} = \frac{\text{Total Expenditure}}{\text{Total quantity purchased}} = \frac{3\times10+4\times15+5\times20}{10+15+20}$$

which is the weighted arithmetic mean (A.M.) of the prices taking respective quantities as weights. Further, if Rs. 150, 200 and 250 are spent on the purchase of the commodity at price of Rs. 3, 4 and 5 respectively, then,

$$\text{Average Price} = \frac{150+200+250}{\frac{150}{3}+\frac{200}{4}+\frac{250}{5}},$$ where, $\frac{150}{3}, \frac{200}{4}$ and $\frac{250}{5}$ are the quantities purchased in respective situations.

The above average price is equal to the weighted harmonic mean of prices taking money spent as weights.

From the above example, it is clear that arithmetic means is appropriate for calculating average price $\left(\frac{\text{Money}}{\text{Quantity}}\right)$ where quantities purchased in different situations are given. Similarly harmonic mean will be appropriate when sums of money spent in different situations are given. Further, if the equal quantity is purchased or equal sums of money is spent in different situations, we use simple A.M. or simple H.M. otherwise we use weighted A.M. or H.M.

Example 108. A scooterist purchased petrol at the rate of Rs. 24, Rs. 29.50 and Rs. 36.85 per litre during three successive years. Calculate the average price of petrol:

(a) If he purchased 150, 180 and 195 litres of petrol in the respective years.

(b) If he spent Rs. 3,850, Rs. 4,675 and Rs. 5,825 in three years.

Solution:

$$\text{Average Price} = \frac{\text{Money}}{\text{Quantity}}$$

(a) Since, the condition is given in terms of different litres of petrol in three years, therefore, weighted A.M. will be appropriate to calculate the average price.

Price Per Unit (X)	Quantity (W)	WX
24.00	150	3600
29.50	180	5310
36.85	195	7185.75
Total	525	16,095.75

$$\text{Average price} = \frac{\text{Money Spent}}{\text{Quantity}} = \frac{\Sigma WX}{\Sigma W}$$

$$= \frac{3600+5310+7185.75}{525} = \frac{16095.75}{525} = \text{Rs. } 30.65$$

(b) Since, the condition is given in terms of different sums of money spent in three years, therefore, weighted H.M. will be appropriate to calculate the average price.

Price Per Unit (X)	Money Spent (W)	$\frac{W}{X}$
24.00	3850	$\frac{3850}{24}$
29.50	4675	$\frac{4675}{29.5}$
36.85	5825	$\frac{5825}{36.85}$

$$\text{Average price} = \frac{\text{Money Spent}}{\text{Quantity}} = \frac{\Sigma W}{\Sigma\left(\frac{W}{X}\right)}$$

$$= \frac{3850+4675+5825}{\frac{3850}{24}+\frac{4675}{29.5}+\frac{5825}{36.85}} = \frac{14350}{160.41+158.47+158.07}$$

$$= \frac{14350}{476.96} = \text{Rs. } 30.09$$

Example 109. Kapil purchases oranges from one shop @ Rs. 2 per kg, from second shop @ Rs. 2.50 per kg, from third shop @ Rs. 3 per kg and from the fourth shop @ Rs. 3.50 per kg. Find the average price per kg if

(a) He purchases 5 kg oranges from each shop.

(b) He purchases oranges of Rs. 50 from each shop.

Solution: $\text{Average price} = \frac{\text{Money}}{\text{Quantity}}$

(a) Since, the condition is given the terms of same quantity from different shops, therefore, simple A.M. will be appropriate to be calculate average price.

Price Per Unit (X)	Quantity (W)
2.00	5
2.50	5
3.00	5
3.50	5

$$\text{Average price} = \frac{2+2.50+3+3.50}{4}$$

$$= \text{Rs. } 2.75$$

(b) Since, the condition is given in terms of equal sum of money spent, therefore, simple H.M. will be appropriate to calculate the average price.

Price per unit (X)	Money spent (W)
2.00	50
2.50	50
3.00	50
3.50	50

$$\text{Average Price} = \frac{4}{\frac{1}{2}+\frac{1}{2.50}+\frac{1}{3}+\frac{1}{3.50}} = \text{Rs. } 2.64$$

Example 110. An individual purchases three qualities of pencil. The relevant data are given below:

Quality	Price per pencil (Rs.)	Money spent
A	1.00	50
B	1.50	30
C	2.00	20

Calculate the average price per pencil.

Solution: Since, different sums of money spent in various situations are given, we shall calculate weighted H.M. to calculate average price.

Price per unit (X)	Money spent (W)
1.00	50
1.50	30
2.00	20
	$\Sigma W = 100$

$$\text{Weighted H.M.} = \frac{50+30+20}{\frac{50}{1.00}+\frac{30}{1.50}+\frac{20}{2.00}} = \frac{50+30+20}{50+20+10} = \text{Rs. } 1.25$$

Merits and Demerits of Harmonic Mean

Merits:
(1) It is rigidly defined.
(2) It is based on all the observations in the given data.
(3) It is capable of further algebraic treatment.
(4) In problems relating to time and rates, it gives better results than other averages.

Demerits:
(1) It is difficult to understand and calculate.
(2) It gives more weight to small items. This is not a desirable feature.
(3) Its value cannot be computed when there are both positive and negative items in a series or when one or more items are zero.

EXERCISE 5.10

1. Find the harmonic mean of the following series:
   3, 5, 6, 6, 7, 10, 12
   [Ans. 5.9]
2. The following table gives marks obtained by a group of students in a test. Calculate the harmonic mean of the series:

Marks obtained:	20	21	22	23	24	25
No. of students:	4	2	7	1	3	1

[Ans. 21.9]

3. Calculate the H.M. for the following:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	4	7	28	12	9

[Ans. 20.48]

4. A cyclist covers first 30 km at an average speed of 80 km/hr, another 20 km at 30 km/hr and the last 20 km at 20 km/hr. Find the average speed for the entire journey and verify your answer. [Ans. 34.3 km/hr]
5. A cyclist covers successive quarters of a mile at the speed of 12, 10, 8 and 7 km/hr respectively. Find the average speed. [Ans. 8.8 km/hr]
6. A train runs 25 km at an average speed of 30 km/hr, another 50 km at a speed of 40 km/hr, then due to repair of the track travels for 6 minutes at a speed of 10 km/hr and finally covers the remaining distance of 24 km at a speed of 24 km/hr. What is the average speed in km/hr? [Ans. 31.41km/hr]
7. A man starts from rest and travels successive quarters of miles at an average speeds of 12, 16, 24 and 48 kms per hour. The average speed over the whole distance is 19.2 km/hr. and not 25 km./hr. Explain and show how you can verify the arithmetic.
8. Four persons A, B, C and D are working in a factory. A produces a unit of production in 12 minutes, B in 16 minutes, C in 24 and D in 48 minutes. Find the average rate of working. What is the average number of units completed per minute. At this rate how many units will they complete in 6 hours a day. [Ans. (*i*) 19.2 minutes per unit (*ii*) 0.052 units (*iii*) 74.88 units]
9. The interest paid on each of the three different sums of money yielding 10%, 12% and 15% simple interest p.a. is the same. What is the average yield per cent on the sum invested? **[Hint:** Use H.M.] [Ans. 12%]
10. A man having to drive 90 km wishes to achieve an average speed of 30 km per hour. For the first half of the journey, he averages only 20 km per hour. What must be his average speed for the second half of the journey if his overall average speed is to be 30 km per hour? [Ans. X_2 = 60 m.p.h.]
11. A man buys mangoes from one shop at the rate of Rs. 20 per kg, from the second shop at the rate of Rs. 25 per kg, from the third shop at the of Rs. 30 per kg and from the fourth shop at the rate of Rs. 35 per kg. Find the average rate in rupees per kg if he buys mangoes of Rs. ... from each shop. [Ans. Rs. 26.33 per Kg]
12. A person buys kerosene at Rs. 0.80, Rs. 1.20, Rs. 1.80 and Rs. 2.80 per litre for four successive years. What was the average cost of oil when he spends Rs. 1000 each year and when he buys 1,000 litres every year? [Ans. Rs. 1.335 per litre, Rs. 1.65 per litre]

◘ RELATIONSHIP BETWEEN A.M., G.M. AND H.M.

(*i*) **For any two positive numbers, G.M. = $\sqrt{\text{A.M.} \times \text{H.M.}}$**. This result can verified as

Let a and b be two positive numbers.

$$\text{A.M.} = \frac{a+b}{2}, \ \text{G.M.} = \sqrt{ab}, \ \text{H.M.} = \frac{2ab}{a+b}$$

Now, $$\text{A.M.} \times \text{H.M.} = \frac{a+b}{2} \times \frac{2ab}{a+b} = ab = (\text{G.M.})^2$$

$\therefore$ $$\text{G.M.} = \sqrt{\text{A.M.} \times \text{H.M.}}$$

Hence, the result is proved.

(*ii*) When all the values of the series differ in size, A.M. is greater than G.M. and G.M. is greater than harmonic mean, *i.e.*,

$$\text{A.M.} > \text{G.M.} > \text{H.M.}$$

(*iii*) If all the values of the series are equal, then A.M. is equal to G.M. and G.M. is equal to H.M., *i.e.*,

$$\text{A.M.} = \text{G.M.} = \text{H.M.}$$

Example 111. If the A.M. of two numbers is 10 and their G.M. is 8, find their harmonic mean and two numbers.

Solution: Given: A.M. = 10, G.M. = 8

We also know that

$$\sqrt{(\text{A.M.})(\text{H.M.})} = (\text{G.M.})$$

$$\sqrt{(10)(\text{H.M.})} = (8)$$

Squaring both means, we get

$$\Rightarrow \quad \text{H.M.} = \frac{64}{10} = 6.4$$

Let the two numbers be X_1 and X_2. We are given that

$$\text{A.M.} = \frac{X_1 + X_2}{2} = 10 \quad \text{and} \quad \text{G.M.} = \sqrt{X_1 . X_2} = 8$$

$$\Rightarrow \quad X_1 + X_2 = 20 \ ...(i) \qquad \Rightarrow \quad X_1 . X_2 = 64 \qquad ...(ii)$$

We can write,

$$(X_1 - X_2)^2 = (X_1 + X_2)^2 - 4X_1 . X_2$$

$$= 400 - 256 = 144$$

$$\Rightarrow \quad X_1 - X_2 = 12 \qquad ...(iii)$$

Adding (*i*) and (*iii*), we get

$$2X_1 = 32 \quad \therefore \quad X_1 = 16$$

Also $X_2 = 4$

Example 112. Using the values 2, 4 and 8, verify that

$$A.M. > G.M. > H.M.$$

Solution:

$$A.M. = \frac{2+4+8}{3} = 4.67 \text{ approx.}$$

$$G.M. = 3\sqrt{2\times4\times8} = (64)^{1/3} = 4$$

$$H.M. = \frac{3}{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}} = \frac{3}{\frac{3}{8}} = \frac{3\times8}{7} = \frac{24}{7} = 3.43$$

Thus, $\quad A.M. > G.M. > H.M.$

Example 113. Comment on the following:

$$A.M. = 25,\ G.M. = 20 \text{ and } H.M. = 21$$

Solution: The statement is wrong because H.M. cannot be greater than G.M.

RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE

The relationship between $\overline{X}$, M and Z depends on the shape of the frequency distribution which is discussed below:

(a) In a perfectly symmetrical distribution, mean, median and mode are all equal, *i.e.*,

$$\overline{X} = M = Z$$

(b) In a moderately asymmetrical distribution, mean, mode and median are not equal.

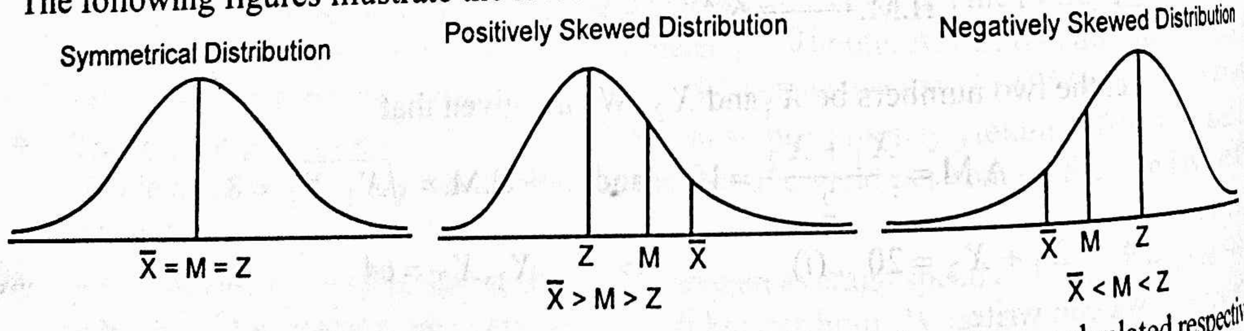
(*i*) when the distribution is positively skewed, *i.e.*, skewed to right, then

$$\overline{X} > M > Z$$

(*ii*) when the distribution is negatively skewed, *i.e.*, skewed to the left, then

$$\overline{X} < M < Z$$

The following figures illustrate the idea:



Example 114. In a negatively skewed distribution, mean, median and mode are calculated respectively as:

$$\overline{X} = 25, M = 28, Z = 22$$

Do you agree? Comment.

Solution: Given that in a nagatively skewed distribution,

$$\overline{X} = 25, M = 28, Z = 22$$

But we do not agree with the statement because in a negatively skewed distribution,

$$\overline{X} < M < Z$$

ISCELLANEOUS SOLVED EXAMPLES

ample 115. A firm declared bonus according to respective salary group as given below:

Salary group	Rate of Bonus	No.of Employees
60—75	60	3
75—90	70	4
90—105	80	5
105—120	90	5
120—135	100	7
135—150	110	6

Calculate the Arithmetic Mean of salary and Geometric Mean of the bonus to the employees.

ition: **Calculation of Arithmetic Mean of Salary**

Salary group	Mid-value	No. of employees (f)	$A = 112.5$ $d = X - A$	d'	fd'
60—75	67.5	3	–45	–3	–9
75—90	82.5	4	–30	–2	–8
90—105	97.5	5	–15	–1	–5
105—120	112.5	5	0	0	0
120—135	127.5	7	+15	1	7
135—150	142.5	6	+30	2	12
		30			–3

$$\overline{X} = A + \frac{\Sigma fd'}{\Sigma f} \times i$$

$$= 112.5 + \frac{(-3)}{30} \times 15 = 112.5 - 1.5$$

$$= 111 \text{ rupees.}$$

Rate of bonus (X)	Frequency (f)	$\log X$	$f \log X$
60	3	1.7782	5.3346
70	4	1.8451	7.3804
80	5	1.9031	9.5155
90	5	1.9542	9.7710
100	7	2.0000	14.0000
110	6	2.0414	12.2484
			58.2499

$$\log G = \frac{\Sigma f \log X}{\Sigma f} = \frac{58.2499}{30} = 1.9416$$

$$\text{G.M.} = \text{Antilog } [1.9416]$$
$$= 87.42$$

Example 116. The rate of certain commodity in the first week of January 1987 is 0.4 kg per rupee; is 0.6 kg per week in the second week and 0.5 kg per rupee in the third week. Therefore is it correct to say that the average price is 0.5 kg per rupee? Verify.

Solution: The answer given is not correct as it is based on arithmetic mean whereas in ca[se] appropriate average is harmonic mean.

$$\text{H.M.} = \frac{N}{\frac{1}{a}+\frac{1}{b}+\frac{1}{c}} = \frac{3}{\frac{1}{0.4}+\frac{1}{0.6}+\frac{1}{0.5}}$$

$$= \frac{3}{\frac{0.3+0.2+0.24}{0.12}} = \frac{3}{\frac{0.74}{0.12}} = \frac{3\times 0.12}{0.74}$$

$$= 0.486$$

Hence, the average price is 0.486 kg per rupee and not 0.5 kg per rupee.

Example 117. In a class of 50 students 10 have failed and their average of marks is 2.5. The total mar[ks] secured by the entire class were 281. Find the average marks those who have passed.

Solution: $N = 50$, failed = 10

Mean marks of those who failed = 2.5

Total marks of 10 students who failed = $2.5 \times 10 = 25$

Total marks secured by entire class = 281

Total marks obtained by those who passed = $281 - 25 = 256$

Average marks obtained by those who passed = $\frac{256}{40} = 6.4$

Example 118. B.Com. (Pass) III year has three Sections A, B and C with 50, 40, 60 stud[ents] respectively. The mean marks for the three sections were determined as 85, 60 and [65] respectively. However, marks of a student of Section A were wrongly recorded a[s 50] instead of zero. Determine the mean marks of all the three sections put together.

Solution:

Section	A	B	C
Number (N)	50	40	60
Mean ($\overline{X}$)	85	60	65
ΣX	4250	2400	3900
Correct ΣX	(4250 – 50)	2400	3900

Correct mean of all the three sections = $\frac{4200+2400+3900}{150} = 70$

xample 119. An incomplete distribution is given below:

Variable:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	Total
Frequency:	10	20	—	40	—	25	15	=170

You are given that the median value is 35. Find out missing frequency.

lution: Let the missing frequencies be denoted by x and y.

Variable	f	c.f.
0—10	10	10
10—20	20	30
20—30	x	30+x
30—40	40	70+x
40—50	y	70+x+y
50—60	25	95+x+y
60—70	15	110+x+y
	N=170	

Now, $110+x+y=170$

$\therefore \quad x+y=60$

or $\quad y=60-x$

Median = 35

$\therefore$ Median class is 30—40

$$\text{Median item} = \frac{N}{2} = \frac{170}{2} = 85 \text{ th item.}$$

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$35 = 30 + \frac{85-(30+x)}{40} \times 10$$

$$5 = \frac{85-30-x}{4} \times 10 = \frac{55-x}{4}$$

$$\therefore \quad 20 = 55 - x$$

$$\therefore \quad x = 35$$

$$y = 60 - x = 60 - 35 = 25$$

Hence, the missing frequency of the class 20—30 is 35 and the missing frequency of the class 40—50 is 25.

ample 120. In 500 small-scale industrial units the return on investment ranged from 0 to 30 per cent, no unit sustaining any loss. 5 per cent of the units had returns ranging from 0 per cent upto (and including) 5 per cent and 15 per cent of the units earned returns exceeding 5 per cent but not exceeding 10 per cent. The median rate of return was

15 per cent and the upper quartile 20 per cent. The uppermost layer of the returns exceeding 25 per cent was earned by 50 units.

Present this information in the form of a frequency table with intervals of 5 per cent as follows:

Exceeding 0 per cent but not exceeding 5 per cent.

Exceeding 5 per cent but not exceeding 10 per cent.

Exceeding 10 per cent but not exceeding 15 per cent.

Exceeding 15 per cent but not exceeding 20 per cent.

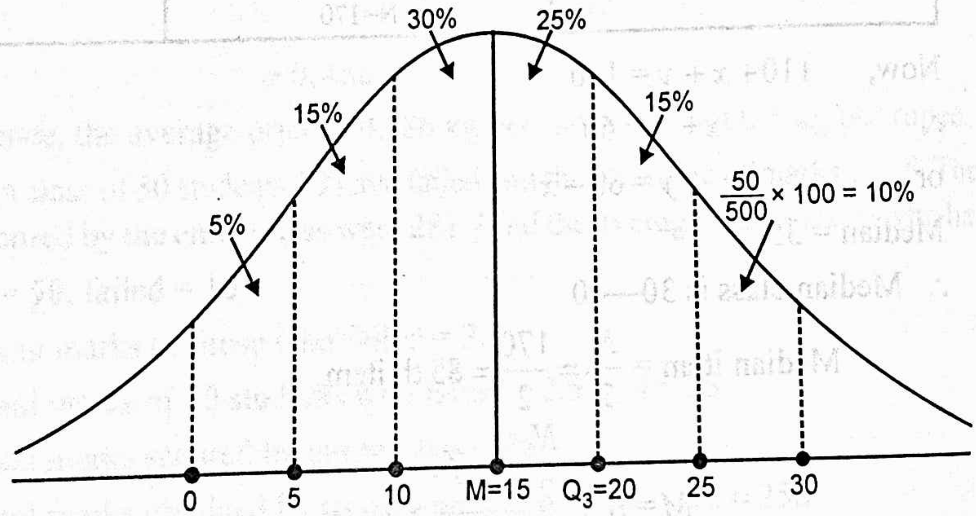
Exceeding 20 per cent but not exceeding 25 per cent.

Exceeding 25 per cent but not exceeding 30 per cent.

Use $\frac{N}{4}, \frac{2N}{4}, \frac{3N}{4}$ as the ranks of the lower, middle and upper quartiles respectively.

Find rate of return round which there is maximum concentration of the units.

Solution: First we put the information in the form of a normal curve.



The information given above can be summarised as follows:

Rate of Return on Investment	Firms 5% of total	Number of firms (f)
Exceeding 0 but not exceeding 5	5	25
" 5 " " " 10	15	$75-f_0$
" 10 " " " 15	30	$150-f_1$
" 15 " " " 20	25	$125-f_2$
" 20 " " " 25	15 (Balance)	75
" 25 " " " 30	10	50
	100	500

The rate of return around which there is maximum concentration is the modal class.

By inspection, mode lies in the classs 10—15.

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

$$= 10 + \frac{150-75}{300-75-125} \times 5$$

$$= 10 + \frac{75}{100} \times 5 = 10 + 3.75$$

$$= 13.75$$

Hence, the rate of return around which there is maximum concentration of the units is 13.75%.

Example 121. Find the missing frequency of the group 20—30 when the Median is 28.

X:	0—10	10—20	20—30	30—40	40—50
f:	5	8	?	16	6

Solution: **Calculation of Missing Frequency**

X	*f*	*c.f.*
0—10	5	5
10—20	8	13
20—30	f_1	$13+f_1$
30—40	16	$29+f_1$
40—50	6	$35+f_1$
	$N=35+f$	

Since, Median = 28, it lies in the class 20—30

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$28 = 20 + \frac{\frac{35+f_1}{2} - 13}{f_1} \times 10$$

$$8 = \frac{\frac{35+f_1}{2} - 13}{f_1} \times 10$$

$$\Rightarrow \quad 8f_1 = \left(\frac{35+f_1}{2} - \frac{13}{1}\right)10 = \left(\frac{35+f_1-26}{2}\right)10$$

$$16f_1 = 350 + 10f_1 - 260$$

$$\Rightarrow \quad 6f_1 = 90$$

$$\therefore \quad f_1 = \frac{90}{6} = 15$$

Thus, the missing frequency is 15.

Example 122. The following table gives the frequency distribution of the marks of 400 candidates in an examination.

Marks	No. of Candidates	Marks	No. of candidates
0—10	5	50—60	65
10—20	20	60—70	50
20—30	40	70—80	35
30—40	70	80—90	20
40—50	85	90—100	10

(*i*) If the minimum marks required for passing are 35, what percentage of candidates pass the examination?

(*ii*) It is decided to allow 80% of the candidates to pass, what should be minimum marks for passing?

Solution:

Marks	f	$c.f.$
0—10	5	5
10—20	20	25
20—30	40	65
30—40	70	135
40—50	85	220
50—60	65	285
60—70	50	335
70—80	35	370
80—90	20	390
90—100	10	400
	$N = 400$	

(*i*) Let us assume that marks are given in whole numbers, the students getting marks upto 34 will fail.

∴ The percentage of candidates passing the examination

$$= \left[\left(\frac{5}{10} \times 70\right) + 85 + 65 + 50 + 35 + 20 + 10\right] \times \frac{100}{400}$$

$$= [(35) + 265] \times \frac{100}{400} = 300 \times \frac{100}{400} = 75\%$$

(*ii*) Since 80% candidates are passed and 20% have failed. The minimum pass marks are given by P_{20}.

$$P_{20} = \text{Size of } \frac{20(N)}{100}\text{th item} = \frac{20(400)}{100} = 80\text{th item}$$

It lies in class 30—40.

Applying the formula

$$P_{20} = 30 + \frac{80-65}{70} \times 10 = 30 + 2.1 = 32.1 \approx 32$$

Example 123. During a period of decline in stock market price, a stock sold at Rs. 50 per share on one day, Rs.40 on the next day and Rs. 25 on the third.

(*i*) If an investor bought 100, 120 and 180 shares on the respective three days, find the average price paid per share.

(*ii*) If an investor bought Rs. 1000 worth of shares on each of three days, find the average price paid per share.

Solution: (*i*) Since different quantities of shares are bought, we use weighted A.M.

Price (M) (X)	Quantity (W)	WX
50	100	5,000
40	120	4,800
25	180	4,500
	$\Sigma W = 400$	$\Sigma WX = 14,300$

$$\text{Average price} = \frac{\Sigma WX}{\Sigma W} = \frac{14,300}{400} = 35.75$$

(*ii*) Since equal sum of money is spent, we use simple H.M.

Price (X)	Money spent (Rs.)
50	1000
40	1000
25	1000

$$\text{Average price} = \frac{3}{\frac{1}{50} + \frac{1}{40} + \frac{1}{25}} = 35.29$$

Example 124. A man gets three annual increment in salary. At the end of the first year, he gets an increase of 4%, at the end of second an increase of 6% on his salary and at the end of the third year an increase of 9% on his salary. What is the average percentage increase?

Solution: Average rate of increase in the salary would be obtained by applying geometric mean. Suppose the salary at the beginning of the year is 100.

Year	Annual increase	Salary at the end of year (X)	log *X*
I	4%	104	2.0170
II	6%	106	2.0253
III	9%	109	2.0374
$N = 3$			$\Sigma \log X = 6.0797$

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

$$= \text{Antilog}\left(\frac{6.0797}{3}\right)$$

$$= \text{Antilog } [2.0265]$$

$$= 106.3$$

The average percentage increase = 106.3 – 100 = 6.3%

Example 125. Find the missing information in the following table:

	A	B	C	Combined
Number	10	8	–	24
Mean	20	–	6	15
Geometric Mean	10	7	–	8.397

Solution: Finding missing information:

Number: For C missing information shall be

$$24 - (10 + 8) = 6$$

Mean: Let x be the mean of B

Then, $(20 \times 10) + (8 \times x) + (6 \times 6) = (15 \times 24)$

$$200 + 8x + 36 = 360$$

$$\Rightarrow \quad 8x = 360 - 200 - 36 = 124$$

$$\therefore \quad x = \frac{124}{8}$$

$$= 15.5$$

Hence, mean of B = 15.5.

Geometric Mean : Let x be the geometric mean of C.

$$(10)^{10} \times (7)^8 \times x^6 = (8.397)^{24}$$

$$10 \log 10 + 8 \log 7 + 6 \log x = 24 \log 8.397$$

$$10 + (8 \times 0.8451) + 6 \log x = 24(0.9241)$$

$$10 + 6.7608 + 6 \log x = 22.1784$$

$$\Rightarrow \quad 6 \log x = 22.1784 - 10 - 6.7608 = 5.4176$$

$$\log x = 0.9029$$

$$x = \text{Antilog } 0.9029$$

$$= 7.997$$

Hence, geometric mean of C is 7.997.

IMPORTANT FORMULAE

1. Arithmetic Mean

For Individual Series:

(i) $\overline{X} = \frac{\Sigma x}{N}$ — Direct Method

(ii) $\overline{X} = A + \frac{\Sigma d}{N}$ — Short-cut Method

(iii) $\overline{X} = A + \frac{\Sigma d'}{N} \times i$ — Step Deviation Method

For Discrete and Continuous Series:

(i) $\overline{X} = \frac{\Sigma fx}{N}$ or $\frac{\Sigma fm}{N}$ — Direct Method

(ii) $\overline{X} = A + \frac{\Sigma fd}{N}$ — Short-cut Method

(iii) $\overline{X} = A + \frac{\Sigma fd'}{N} \times i$ — Step Deviation Method

2. Weighted Arithmetic Mean

$$\overline{X}_w = \frac{\Sigma WX}{\Sigma W}$$

3. Combined Mean

$$\overline{X}_{123} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2 + N_3\overline{X}_3}{N_1 + N_2 + N_3}$$

4. Median

For Individual and Discrete Series:

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

For Continuous Series:

$$M = \text{Size of} \left(\frac{N}{2}\right) \text{th item}$$

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

5. Mode

For Continuous Series:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

6. Emprical Mode

Mode (Z) = 3 Median – 2 Arithmetic Mean

or $$Z = 3M - 2\overline{X}$$

7. Geometric Mean

For Individual Series

$$\text{G.M.} = \text{Antilog}\left[\frac{\Sigma \log X}{N}\right]$$

For Discrete and Continuous Series

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

Compound Interest Formula

$$P_n = P_0\,(1+r)^n$$

or $$r = \text{Antilog}\left[\frac{\log P_n - \log P_0}{n}\right] - 1$$

8. Harmonic Mean

For Individual Series

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

For Discrete and Continuous Series

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

Weighted Harmonic Mean

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

QUESTIONS

1. What is meant by central tendency? State important measures of central tendency.
2. What is a statistical average? What are the properties of an ideal average?
3. Explain the relative merits and demerits of arithmetic mean, median and mode as measures of central tendency.

   Or

   Define arithmetic mean, median and mode and discuss their relative merits and demerits.
4. Discuss in brief merits and demerits of various measures of central tendency.
5. Give different measures of central tendency with their formulae. Also state the situations where these measures can be used.
6. Give relationship between A.M., G.M. and H.M.
7. Give the empirical relationship between $\overline{X}$, M and Z for a perfectly symmetrical distribution and a moderately skewed distribution.
8. What are the essentials of a good average?

   Or

   What are the characteristics of a good measure of central tendency?
9. Explain the mathematical properties of Arithmetic Mean. What is the relationship among Mean, Median and Mode?
10. What are the desirable properties of an average? Which of the average you know possess most of them?

6 Measures of Dispersion

INTRODUCTION

The various measures of central tendency or averages discussed in the previous chapter give only one single figure that represents the entire set of data. But the average alone cannot describe the set of observations fully. It does not reveal the degree of spread out or extent of variability of individual observations in a series. There can be a number of series whose average may be the same but still there can be wide disparities in the formation of the series. In such a case, it becomes necessary to study the variability or dispersion of the observations. Measures of dispersion help to study variability of the items, *i.e.*, the extent to which the items vary from one another and from the central value.

MEANING OF DISPERSION

The term dispersion is generally used in two senses: (1) Firstly, dispersion refers to the variations of the items among themselves. If the value of all the items of a series is the same, there will be no variation among the various items and the dispersion will be zero. On the other hand, the greater the variation among different items of a series, the more will be the dispersion. (2) Secondly, dispersion refers to the variation of the items around an average. If the difference between the value of items and the average is large, the dispersion will be high and on the other hand if the difference between the value of the items and average is small, the dispersion will be low. Thus, dispersion is defined as scatterness or spreadness of the individual items in a given series.

DEFINITION OF DISPERSION

Some important definitions of dispersion are given below:

1. *Dispersion is a measure of the variations of the items.* —Bowley
2. *Dispersion is a measure of the extent to which the individual items vary.* —Connor
3. *The degree to which numerical data tend to spread about average is called variation or dispersion of data.* —Spiegel

The above definitions make it clear that dispersion refers to the extent to whch the items vary from one another and from the central value.

OBJECTIVES OF MEASURING DISPERSION

The measures of dispersion are helpful in statistical investigation. Some of the main objectives of dispersion are as under:

(1) To determine the reliability of an average: The measures of dispersion help in determining the reliability of an average. It points out as to how far an average is representative of a statistical series. If the dispersion or variation is small, the average will closely represent the individual values and it is highly representative. On the other hand, if the dispersion or variation is large, the average will be quite unreliable.

(2) To compare the variability of two or more series: The measures of dispersion helps in comparing the variability of two or more series. It is also useful to determine the uniformity or consistency of two or more series. A high degree of variation would mean less consistency or less uniformity as compared to the data having less variation.

(3) For facilitating the use of other statistical measures: Measures of dispersion serves the basis of many other statistical measures such as correlation, regression, testing of hypothesis, etc. These measures are based on measures of variation of one kind or another.

(4) Basis of Statistical Quality Control: The measures of dispersion is the basis of statistical quality control. The extent of the dispersion gives indication to the management as to whether the variation in the quality of the product is due to random factors or there is some defect in the manufacturing process. On the basis of this analysis, the management may take suitable measures to control the cause of variation in the quality of the product.

PROPERTIES OF A GOOD MEASURE OF DISPERSION

A good measure of dispersion should possess the following properties:

(1) It should be easy to understand.

(2) It should be simple to calculate.

(3) It should be uniquely defined.

(4) It should be based on all observations.

(5) It should not be unduly affected by the extreme items.

(6) It should be capable of further algebraic treatment.

ABSOLUTE AND RELATIVE MEASURES OF DISPERSION

Measures of dispersion may be either **absolute** or **relative**.

Absolute Measure of Dispersion: Absolute measure of dispersion is expressed in the same unit in which data of the series are expressed. They are expressed in same statistical unit, *e.g.*, rupees, kilogram, tons, years, centimeters, etc.

Relative Measure of Dispersion: Relative measure of dispersion refers to the variability stated in the form of ratio or percentage. Thus, relative measure of dispersion is independent of unit of measurement. It is also called **coefficient of dispersion**. These measures are used to compare two series expressed in different units.

METHODS OF MEASURING DISPERSION

The following are the main methods of measuring dispersion:

(1) Range

(2) Interquartile Range and Quartile Deviation

(3) Mean Deviation

(4) **Standard Deviation**

(5) **Coefficient of Variation**

(6) **Lorenz Curve.**

(1) RANGE

It is the simplest measures of dispersion. It is defined as the difference between the largest a smallest values in the series. Its formula is:

$$R = L - S$$

Where, R = Range, L = Largest value in the series, S = Smallest value in the series.

The relative measure of range, also called **coefficient of range**, is defined as:

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

The following examples illustrate the calculation of range:

Calculation of Range

Individual Series

Example 1. Five students obtained the following marks in statistics:

20, 35, 25, 30, 15

Find the range and coefficient of Range.

Solution: Here, L = 35, and S = 15

$$\text{Range} = L - S$$

$$\therefore \quad \text{Range} = 35 - 15 = 20$$

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

$$= = \frac{35-15}{35+15} = \frac{20}{50} = \frac{2}{5} = +\,0.40$$

Discrete Series

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

Marks:	10	20	30	40	50	60	70
No. of students:	15	18	25	30	16	10	9

Solution: Here, L = 70, and S = 10

$$\text{Range} = L - S$$

$$\therefore \quad \text{Range} = 70 - 10 = 60$$

$$\text{Coefficient of Range} = \frac{L-S}{L+S} = \frac{70-10}{70+10} = \frac{60}{80} = \frac{3}{4} = +\,0.75$$

Continuous Series

Example 3. Find out range and the coefficient of range of the following series:

Size:	5—10	10—15	15—20	20—25	25—30
Frequency:	4	9	15	30	40

Solution:

$$\text{Range} = L - S$$

Here, L = Upper limit of the largest class = 30, S = Lower limit of the smallest class = 5

$$\therefore \quad \text{Range} = 30 - 5 = 25$$

$$\text{Coefficient of Range} = \frac{30-5}{30+5} = \frac{25}{35} = \frac{5}{7}$$

Note: Since the maximum and minimum of the observations are not identifiable for a continuous series, the range is defined as the difference between the upper limit of the largest class and the lower limit of the smallest class.

Example 4. Find out range and coefficient of range of the following series:

Marks:	20—29	30—39	40—49	50—59	60—69
No. of students:	8	12	20	7	3

Solution: This is an inclusive series. For the calculation of range, the series must be converted into exclusive series as:

Marks	**No. of students**
19.5—29.5	8
29.5—39.5	12
39.5—49.5	20
49.5—59.5	7
59.5—69.5	3

Here, L = 69.5, S = 19.5

$$\therefore \quad \text{Range} = 69.5 - 19.5 = 50$$

$$\text{Coefficient of Range} = \frac{69.5-19.5}{69.5+19.5} = \frac{50}{89} = 0.562$$

Merits and Demerits of Range

erits:

(*i*) It is simple to understand.

(*ii*) It is easy to calculate.

(*iii*) It is widely used in statistical quality control. Range charts are useful in controlling the quality of the product.

merits:

(*i*) It cannot be calculated in case of open ended distribution.

(*ii*) It is not based on all observations of the series.

(*iii*) It is affected by sampling fluctuations.

(*iv*) It is affected by extreme values in the series.

EXERCISE 6.1

1. 5 students obtained following marks in statistics:

   20, 35, 25, 30, 15

   Find out range and coefficient of range. [Ans. R=20, Coeff. of R=

2. Calculate the range and its coefficient from the following data:

Marks:	10	20	30	40	50	60	70
No. of students:	15	18	25	30	16	10	9

[Ans. R=60, Coeff. of Range = 0

3. Calculate the range and its coefficient from the following data:

Marks:	10-20	20-30	30-40	40-50	50-60
No. of students:	8	10	12	8	4

[Ans. R=50, Coeff. of R=0.

4. Find the value of the smallest item of data if the coefficient of range is 0.6 and largest value is

   [Ans. S

◘ (2) INTERQUARTILE RANGE AND QUARTILE DEVIATION

Interquartile range and quartile deviation are another measures of dispersion. The differ between the upper quartile (Q_3) and the lower quartile (Q_1) is called interquartile r Symbolically,

$$\textbf{Interquartile Range} = Q_3 - Q_1$$

The interquartile ranges covers dispersion of middle 50% of the items of the series.

Quartile deviation, also called **Semi-interquartile Range** is half of the difference betwe upper and lower quartiles, *i.e.*, half of the interquartile range. Its formula is as:

$$\textbf{Quartile Deviation (Q.D.)} = \frac{Q_3 - Q_1}{2}$$

The relative measure of quartile deviation also called the **coefficient of quartile devi** defined as:

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

▶ Steps for Calculation

(*i*) Find out the lower quartile (Q_1).

(*ii*) Find out the upper quartile (Q_3).

(*iii*) Put the values of Q_1 and Q_3 in the formula of quartile deviation and coefficient of deviation.

The following examples illustrate the calculation of quartile deviation:

Individual Series

Example 5. Find interquartile range, quartile deviation and coefficient of quartile deviation from the following data:

28, 18, 20, 24, 27, 30, 15

Solution: First arrange the data in ascending order:

15, 18, 20, 24, 27, 28, 30

$$Q_1 = \text{Size of}\left(\frac{N+1}{4}\right)\text{th item} = \text{Size of}\left(\frac{7+1}{4}\right)\text{th item}$$

$$= \text{Size of 2nd item} = 18 \text{ marks}$$

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)\text{th item} = \text{Size of } 3\left(\frac{7+1}{4}\right)\text{th item}$$

$$= \text{Size of 6th item} = 28 \text{ marks}$$

$$\text{Interquartile Range} = Q_3 - Q_1 = 28 - 18 = 10$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{28-18}{2} = \frac{10}{2} = 5$$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{28-18}{28+18} = \frac{10}{46} = 0.217$$

Discrete Series

Example 6. Calculate interquartile range, quartile deviation and the coefficient of quartile deviation from the following data:

Wages (Rs.):	10	20	30	40	50	60
No. of workers:	2	8	20	35	42	20

Solution:

Calculation of Q.D.

(X)	f	$c.f.$
10	2	2
20	8	10
30	20	30
40	35	65
50	42	107
60	20	127
	$N = 127$	

$$Q_1 = \text{Size of } \frac{N+1}{4} \text{ th item} = \text{Size of}\left(\frac{127+1}{4}\right)\text{th item}$$

$$= \text{Size of 32nd item} = 40$$

$$\therefore \quad Q_1 = 40$$

$$Q_3 = \text{Size of } 3\left(\frac{N+1}{4}\right)\text{th item} = \frac{3\times 128}{4} = \text{Size of 96th item} = 50$$

$$\text{Interquartile Range} = Q_3 - Q_1 = 50 - 40 = 10$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{50-40}{2} = \frac{10}{2} = 5$$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{50-40}{50+40} = \frac{10}{90} = 0.11$$

○ Continuous Series

Example 7. Calculate interquartile range, quartile deviation and coefficient of quartile deviation from the following data:

Age (years):	0—20	20—40	40—60	60—80	80—100
Persons:	4	10	15	20	11

Solution:

Calculation of Q.D. and Coefficient of Q.D.

Age (yrs.)	f	$c.f.$
0—20	4	4
20—40	10	14
40—60	15	29
60—80	20	49
80—100	11	60
	$N = 60$	

Calculation of Q_1

$$\frac{N}{4} = \frac{60}{4} = 15\text{th item} \qquad \therefore Q_1 \text{ lies in the class 40—60}$$

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 40 + \frac{15-14}{15} \times 20$$

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

Calculation of Q_3

$$\frac{3N}{4} = \frac{3}{4} \times 60 = 45 \text{ th item.} \quad \therefore Q_3 \text{ lies in the class 60—80}$$

$$Q_1 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i$$

$$= 60 + \frac{45-29}{20} \times 20$$

$$= 60 + 16 = 76$$

$$\text{Interquartile Range} = Q_3 - Q_1 = 76 - 41.33 = 34.67$$

$$\text{Quartile Deviation} = \frac{Q_3 - Q_1}{2} = \frac{76 - 41.33}{2} = 17.33$$

$$\text{Coefficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{76 - 41.33}{76 + 41.33} = \frac{34.67}{117.33} = 0.29$$

Example 8. Find the range which contains the middle 50% of the items and coefficient of quartile deviation from the following data:

X:	11—20	21—30	31—40	41—50	51—60
f:	4	8	20	12	6

Solution: Since, we are given inclusive series, we first convert it into exclusive one:

Calculation of Q.D. and Coefficient of Q.D.

X	*f*	*c.f.*
10.5—20.5	4	4
20.5—30.5	8	12
30.5—40.5	20	32
40.5—50.5	12	44
50.5—60.5	6	50
	$N = 50$	

Calculation of Q_1

$$\frac{N}{4} = \frac{50}{4} = 12.5\text{th item.} \qquad \therefore Q_1 \text{ lies in the class } 30.5 - 40.5$$

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 30.5 + \frac{12.5 - 12}{20} \times 10$$

$$= 30.5 + \frac{0.5}{20} \times 10 = 30.5 + 0.25 = 30.75$$

Calculation of Q_3

$$\frac{3N}{4} = \frac{3 \times 50}{4} = \frac{150}{4} = 37.5\text{ th item.} \quad \therefore Q_3 \text{ lies in the class } 40.5 - 50.5$$

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 40.5 + \frac{37.5 - 32}{12} \times 10$$

$$= 40.5 + 4.58 = 45.08$$

To find the middle 50% of the items, compute $Q_3 - Q_1$, *i.e.*, Interquartile range.

$$\text{Interquartile Range} = Q_3 - Q_1 = 45.08 - 30.75 = 14.33$$

$$\text{Coefficient of } Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{45.08 - 30.75}{45.08 + 30.75} = \frac{14.33}{75.83} = 0.188 = 0.19$$

Merits and Demerits of Quartile Deviation

Merits:

(*i*) It is easy to compute and simple to understand.

(*ii*) It is less affected by extreme values.

(*iii*) It can be computed in open ended classes.

(*iv*) It is superior and more reliable than the range.

(*v*) It is useful when dispersion of middle 50% items is to be calculated.

Demerits:

(*i*) It gives 50% of the items, *i.e.*, the first 25% and the last 25%.

(*ii*) It is not capable for further algebraic treatment.

(*iii*) It is affected by sampling fluctuations.

(*iv*) It is not a good measure of dispersion particularly for series in which variation is considerable.

EXERCISE 6.2

1. From the following data, compute Q.D. and Coefficient of Q.D.:

X:	4	8	10	7	11	15	18	14	12	16

[**Ans.** Q.D.=3.75, Coeff. of Q.D.= 0.326]

2. From the following data, calculate Q.D. and its coefficient:

Height:	58	59	60	61	62	63	64	65	66
No. of students:	15	20	32	35	33	22	20	10	8

[**Ans.** Q.D.=1.5, Coeff. of Q.D. = 0.024

3. Calculate Quartile deviation and its coefficient from the following data:

Size:	less than 32	32—34	34—36	36—38	38—40	40—42	over 42
Frequency:	12	18	16	14	12	8	6

[**Ans.** Q.D. = 2.85, Coeff. of Q.D. = 0.07

4. Calculate the range of marks obtained by middle 50% of the students from the followi data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	5	8	12	8	7

[**Hint:** Calculate $Q_3 - Q_1$] [**Ans.** *IQR* =

5. Calculate Q.D. and its coefficient from the following data:

X:	31—40	41—50	51—60	61—70	71—80
f:	40	60	20	40	25

[**Ans.** Q.D. = 7.087, Coeff. of Q.D. = 0.0

6. Calculate the range of marks obtained by middle 80% of the students:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60170
No. of students:	8	16	22	30	24	12	6

[**Hint:** Calculate $P_{90} - P_{10}$] [**Ans.** Percentile Range $= 42.79$]

7. If the first quartile is 142 and the semi-interquartile range is 18, find the median (assuming the distribution to be symmetrical). [**Ans.** Q_2 (M) = 160]

(3) MEAN DEVIATION OR AVERAGE DEVIATION

Mean deviation is another measure of dispersion. It is also known as **average deviation.** Mean eviation is defined as the arithmetic average of the deviation of the various items of a series omputed from some measures of central tendency say mean or median. In taking deviation of the arious items, algebraic signs '+' and '–' are not taken into consideration. Although mean deviation in be computed either from the mean or median, but theoretically median is preferred because **the im of the deviations of the items taken from median is minimum when signs are ignored.** The rmulae of calculating mean deviation are:

$$\textbf{M.D. from Median} = \frac{\Sigma|X - M|}{N} \text{ or } \frac{\Sigma|d_M|}{N}$$

$$\textbf{M.D. from Mean} = \frac{\Sigma|X - \bar{X}|}{N} \text{ or } \frac{\Sigma|d_{\bar{x}}|}{N}$$

The relative measure of mean deviation, also called the **coefficient of mean deviation** is obtained dividing mean deviation by the particular average used in computing mean deviation. Thus,

$$\textbf{Coefficient of } M.D._M = \frac{M.D._M}{Median}$$

$$\textbf{Coefficient of } M.D._{\bar{x}} = \frac{M.D._{\bar{x}}}{Mean}$$

Calculation of Mean Deviation

ndividual Series

n case of individual series, the following formula is used for calculating mean deviation:

$$M.D._{\bar{X}} = \frac{\Sigma|d_{\bar{X}}|}{N}, \text{ from mean}$$

$$M.D._M = \frac{\Sigma|d_M|}{M}, \text{ from median.}$$

teps for Calculation

i) Calculate either the mean or median.

i) Find the deviations of the items from $\bar{X}$ or M ignoring signs and denote these deviations by $|d_{\bar{X}}|$ or $|d_m|$.

i) Find the total of the deviations, *i.e.*, $\Sigma|d_{\bar{X}}|$ or $\Sigma|d_M|$.

) Divide the total obtained in step (*iii*) by the number of item (*N*). This gives us the mean deviation.

Coefficient of Mean Deviation

The formulae are:

$$\text{Coefficient of } M.D._{\bar{X}} = \frac{M.D._{\bar{X}}}{\bar{X}}$$

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

Example 9. Calculate the mean deviation from mean as well as from median and coefficient of mean deviation from the following data:

Marks: 20, 22, 25, 38, 40, 50, 65, 70, 75

Solution: **Calculation of Mean Deviation**

Marks (X)	Deviation from Mean 45 $\lvert d_{\bar{X}} \rvert$	Deviations from Median 40 $\lvert d_M \rvert$
20	25	20
22	23	18
25	20	15
38	7	2
40	5	0
50	5	10
65	20	25
70	25	30
75	30	35
$N=9,\ \Sigma X=405$	$\Sigma\lvert d_{\bar{X}}\rvert=160$	$\Sigma\lvert d_M\rvert=155$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{405}{9} = 45$$

$$\text{M.D. from Mean} = \frac{\Sigma\lvert d_{\bar{X}}\rvert}{N} = \frac{160}{9} = 17.78$$

$$\text{Coeff. of } M.D._{\bar{X}} = \frac{M.D._{\bar{X}}}{\bar{X}} = \frac{17.78}{45} = 0.39$$

$$\text{Median} = \text{Size of} \left(\frac{N+1}{2}\right)\text{th item} = \text{Size of} \left(\frac{9+1}{2}\right)\text{th item}$$

$$= \text{Size of 5th item} = 40$$

$$\text{M.D. from Median} = \frac{\Sigma\lvert d_M\rvert}{N} = \frac{155}{9} = 17.22$$

$$\text{Coeff. of } M.D._{M} = \frac{M.D._{M}}{M} = \frac{17.22}{40} = 0.43$$

Discrete Series

In case of discrete series, the following formula is used for calculating mean deviation:

$$MD._{\overline{X}} = \frac{\Sigma f \mid d_{\overline{X}} \mid}{N}, \text{ from mean}$$

$$M.D._M = \frac{\Sigma f \mid d_M \mid}{M}, \text{ from median.}$$

Steps for Calculation

(i) Calculate the mean or median from the given series.

(ii) Find the deviations of the items from $\overline{X}$ ignoring '+' or '–' signs and denote these deviations by $\mid d_{\overline{X}} \mid$ or $\mid d_M \mid$.

(iii) Multiply these deviations by their respective frequencies and obtain the total $\Sigma f \mid d \mid$.

(iv) Finally, divide this total by the number of items. This will give the mean deviation.

(v) To calculate the coefficient of mean deviation, we divide the mean deviation by the particular average used in computing mean deviation.

Example 10. Calculate the mean deviation from median and mean and their coefficients from the following table:

X:	20	30	40	50	60	70
f:	8	12	20	10	6	4

Solution: (i) **Calculation of Mean Deviation from Median**

X	f	$c.f.$	M = 40 $\mid d_{\overline{M}} \mid = \mid X - M \mid$	$f \mid d_M \mid$
20	8	8	20	160
30	12	20	10	120
40	20	40	0	0
50	10	50	10	100
60	6	56	20	120
70	4	60	30	120
	$N = 60$			$\Sigma f \mid d_M \mid = 620$

$$M = \text{Size of} \left(\frac{N+1}{2} \right) \text{th item} = \text{Size of} \left(\frac{60+1}{2} \right) \text{th item}$$

$$= \text{Size of 30.5th item} = 40$$

$$\therefore \quad M = 40$$

$$\text{M.D. from Median } (M.D._M) = \frac{\Sigma f \mid d_M \mid}{N} = \frac{620}{60} = 10.33$$

$$\text{Coefficient of M.D. from Median} = \frac{M.D._M}{M} = \frac{10.33}{40} = 0.258$$

(ii) **Calculation of Mean Deviation from Mean**

X	f	fX	$\bar{X}=41$ $\lvert d_{\bar{X}}\rvert=\lvert X-\bar{X}\rvert$	$f\lvert d_{\bar{X}}\rvert$
20	8	160	21	168
30	12	360	11	132
40	20	800	1	20
50	10	500	9	90
60	6	360	19	114
70	4	280	29	116
	$N=60$	$\Sigma fX=2460$		$\Sigma f\lvert d_{\bar{X}}\rvert=640$

$$\bar{X}=\frac{\Sigma fX}{N}=\frac{2460}{60}=41$$

$$\text{M.D. from Mean } (M.D._{\bar{X}})=\frac{\Sigma f\lvert d_{\bar{X}}\rvert}{N}=\frac{640}{60}=10.67$$

$$\text{Coefficient of M.D. from Mean}=\frac{M.D_{\bar{X}}}{\bar{X}}=\frac{10.67}{41}=0.26$$

Continuous Series

For calculating the mean deviation in continuous series, the procedure remains the same as discussed above. The only difference is that here we have to obtain the mid-values of the various classes and then take deviations from these values as before. The formulae are:

$$M.D._{\bar{X}}=\frac{\Sigma f\lvert d_{\bar{X}}\rvert}{N}, \text{ where } d_{\bar{X}}=m-\bar{X}$$

$$M.D_{M}=\frac{\Sigma f\lvert d_{M}\rvert}{N}, \text{ where } d_{M}=m-M$$

$$\text{Coefficient of } M.D_{\bar{X}}=\frac{M.D_{\bar{X}}}{\bar{X}}$$

$$\text{Coefficient of } M.D_{M}=\frac{M.D_{M}}{M}$$

Example 11. Calculate the mean deviation from mean and its coefficient from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No. of students:	5	8	15	16	6

Solution: **Calculation of Mean Deviation from Mean**

Marks	f	M.V. (m)	fm	$\bar{X}=27$ $\lvert d_{\bar{X}}\rvert=\lvert m-27\rvert$	$f\lvert d_{\bar{X}}\rvert$
0—10	5	5	25	22	110
10—20	8	15	120	12	96
20—30	15	25	375	2	30
30—40	16	35	560	8	128
40—50	6	45	270	18	108
	$N=50$		$\Sigma fm=1350$		$\Sigma f\lvert d_{\bar{X}}\rvert=472$

$$\bar{X} = \frac{\Sigma fm}{N} = \frac{1350}{50} = 27$$

$$\text{M.D. from Mean} = \frac{\Sigma f|d_{\bar{X}}|}{N} = \frac{472}{50} = 9.44$$

$$\text{Coefficient of } M.D._{\bar{X}} = \frac{M.D._{\bar{X}}}{\bar{X}} = \frac{9.44}{27} = 0.349$$

Example 12. Calculate mean deviation from median and its coefficient from the following data:

Size:	100—120	120—140	140—160	160—180	180—200
Frequency:	4	6	10	8	5

Solution:

Calculation of Mean Deviation from Median

Size	f	$c.f.$	$M.V.$ (m)	M = 153 $\lvert d_M\rvert = \lvert m - M\rvert$	$f\lvert d_M\rvert$
100—120	4	4	110	43	172
120—140	6	10	130	23	138
140—160	10	20	150	3	30
160—180	8	28	170	17	136
180—200	5	33	190	37	185
	$N = 33$				$\Sigma f\lvert d_M\rvert = 661$

N/2 = 33/2 = 16.5th item ∴ Median lies in 140 – 160.

Applying the formula

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 140 + \left(\frac{16.5 - 10}{10}\right) \times 20 = 140 + 13 = 153$$

$$\text{Mean Deviation from Median} = \frac{\Sigma f|d_M|}{N} = \frac{661}{33} = 20.03$$

$$\text{Coefficient of Mean Deviation from Median} = \frac{M.D._M}{M} = \frac{20.03}{153} = 0.1309$$

Thus, $M.D._M = 20.03$, and Coefficient of $M.D._M = 0.1309$.

Short-cut Method for Mean Deviation

If value of the average comes out to be in fractions, the calculation of M.D. by $\frac{\Sigma f|d|}{N}$ would become quite tedious. In such a case, the following formula is used:

$$M.D. = \frac{(\Sigma fX)_A - (\Sigma fX)_B - [(\Sigma f)_A - (\Sigma f)_B]\ \bar{X} \text{ or } M}{N}$$

Where, $\bar{X}$ or M is the average about which M.D. is to be calculated. In this formula, suffixes A and B denote the sums corresponding to the values of $X > \bar{X}$ or M and $X < \bar{X}$ or M respectively.

This formula can also be used for an individual series, by taking 'f' equal to 1 for each X, in the series. In this case, the formula reduces to

$$M.D. = \frac{(\Sigma X)_A - (\Sigma X)_B - [(N)_A - (N)_B]\,\overline{X} \text{ or } M}{N}$$

Where, $(N)_A$ and $(N)_B$ are the number of items whose values are greater than $\overline{X}$ or M and less than $\overline{X}$ or M respectively.

Note: If short-cut method is to be used to find M.D. $(\overline{X})$, then it is advisable to use *direct method* to find $\overline{X}$, because we would be needing $(\Sigma fX)_A$ and $(\Sigma fX)_B$ in the calculation of M.D.$(\overline{X})$.

Now, we shall take some examples to illustrate this method:

Example 13. Using short-cut method, calculate the mean deviations from mean and median from the following data:

7, 9, 13, 13, 15, 17, 19, 21, 23

Solution:

X	Taking $\overline{X}$	Taking M
7	$\Sigma X_B = 57$	$\Sigma X_B = 42$
9	$N_B = 5$	$N_B = 4$
13		
13		
15		M = 15
	$\overline{X} = 15.22$	
17	$\Sigma X_A = 80$	$\Sigma X_A = 80$
19	$N_A = 4$	$N_A = 4$
21		
23		
$\Sigma X = 137$ $N = 9$		

$$\therefore \quad \overline{X} = \frac{137}{9} = 15.22$$

$$M = \text{Size of}\left(\frac{N+1}{2}\right)\text{th} = \frac{9+1}{2} = \text{5th item} = 15$$

From Mean:

$$M.D_{\overline{X}} = \frac{\Sigma X_A - \Sigma X_B - (N_A - N_B)(\overline{X})}{N}$$

$$= \frac{80 - 57 - (4-5)(15.22)}{9}$$

$$= \frac{23 + 15.22}{9} = \frac{38.22}{9} = 4.25$$

From Median:

$$M.D_M = \frac{\Sigma X_A - \Sigma X_B - (N_A - N_B)(M)}{N}$$

$$= \frac{80 - 42 - (4-4)(15)}{9}$$

$$= \frac{38}{9} = 4.22$$

Example 14. Calculate the mean deviation and its coefficient from mean and median from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50
No.of students:	6	28	51	11	4

Use short-cut method.

Solution: (*i*) **Calculation of M.D. from Mean**

Marks	f		M.V. (X)	fX	
0—10	6	$(\Sigma f)_B = 34$	5	30	$(\Sigma fX)_B = 450$
10—20	28		15	420	
					$\overline{X} = 22.9$
20—30	51	$(\Sigma f)_A = 66$	25	1275	$(\Sigma fX)_A = 1840$
30—40	11		35	385	
40—50	4		45	180	
	$N = 100$			$\Sigma fX = 2290$	

$$\overline{X} = \frac{\Sigma fX}{N} = \frac{2290}{100} = 22.9$$

$$M.D_{\overline{X}} = \frac{(\Sigma fX)_A - (\Sigma fX)_B - [(\Sigma f)_A - (\Sigma f)_B]\overline{X}}{N}$$

$$= \frac{1840 - 450 - [66 - 34][22.9]}{100}$$

$$= \frac{1390 - (32 \times 22.9)}{100}$$

$$= \frac{1390 - 732.8}{100}$$

$$= \frac{657.2}{100} = 6.572$$

$$\text{Coefficient of } M.D_{\overline{X}} = \frac{M.D_{\overline{X}}}{\overline{X}} = \frac{6.572}{22.9} = 0.287$$

(ii) **Calculation of M.D. from Median**

Marks	f		$c.f.$	M.V. (X)	fX	
0—10	6	$(\Sigma f)_B = 34$	6	5	30	$(\Sigma fX)_B$
10—20	28		34	15	420	
20—30	51	$(\Sigma f)_A = 66$	85	25	1275	$M =$
30—40	11		96	35	385	$(\Sigma fX)_A$
40—50	4		100	45	180	
	$N = 100$					

$$\frac{N}{2} = \frac{100}{2} = 50\text{th item}$$

Median lies in the class 20—30

$$\therefore \qquad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 20 + \frac{50-34}{51} \times 10$$

$$= 20 + \frac{160}{51} = 20 + 3.14 = 23.14$$

$$M.D_M = \frac{(\Sigma fX)_A - (\Sigma fX)_B - [(\Sigma f)_A - (\Sigma f)_B](M)}{N}$$

$$= \frac{1840 - 450 - (66-34)(23.14)}{100}$$

$$= \frac{1390 - 32 \times 23.14}{100}$$

$$= \frac{1390 - 740.48}{100} = \frac{649.52}{100} = 6.4952$$

$$\text{Coefficient of } M.D_M = \frac{M.D_M}{M} = \frac{6.4952}{23.14} = 0.281$$

Merits and Demerits of Mean Deviation

Merits:

(i) It is simple to understand and easy to compute.

(ii) It is based on all the observations.

(iii) It is less affected by the extreme items.

(iv) It is very useful in various fields such as economics, commerce and social fields.

(v) Comparision about formation of different series can be easily made as deviations ar from a central value.

Demerits:

(i) Ignoring '±' signs are not correct from mathematical point of view.
(ii) It is not an accurate method when it is calculated from mode.
(iii) It is difficult to compute when the value of mean or median comes in fractions.
(iv) It is not capable of further algebraic treatment.
(v) It is not used in statistical conclusion.

EXERCISE 6.3

1. The monthly salary of five families are given below:
852, 635, 792, 836, 750
Find the mean deviation from the mean. [Ans. $M.D._{\overline{X}} = 64.4$]

2. (i) Find mean deviation and its coefficient from the mean for the following data:

X:	5	10	15	20	25	30	35	40
f:	8	16	18	22	14	9	6	7

[Ans. $M.D._{\overline{X}} = 7.57$, Coefficient of $M.D._{\overline{X}} = 0.37$]

(ii) Calculate the average deviation from median and its coefficient for the following data:

X:	6	12	18	24	30	36	42
f:	4	7	9	18	15	10	5

[Ans. $M = 24$, $M.D._M = 7.94$, Coeff. of $M.D._M = 0.330$]

3. Calculate mean deviation from mean and median for the following data:

Marks:	140—150	150—160	160—170	170—180	180—190	190—200
No. of students:	4	6	10	18	9	3

[Ans. $M.D._{\overline{X}} = 10.56, M.D._M = 10.24$]

4. Calculate mean deviation from the median and its coefficient for the following data:

X:	10—19	20—29	30—39	40—49	50—59
f:	3	4	6	5	2

[Ans. $M = 34.5$, $M.D._M = 9.5$, Coefficient of $M.D_M = 0.275$]

5. Calculate the average deviation from mean and its coefficient from the following data:

Profit (Rs.):	0—10	10—20	20—30	30—40	40—50
No. of Shops:	5	10	15	20	25

[Ans. $\overline{X} = 31.66, M.D._{\overline{X}} = 10.67$, Coeff. of $M.D._{\overline{X}} = 0.34$]

6. Calculate the mean deviation from the median and its coefficient for the following data:

Class:	0—5	5—10	10—15	15—20	20—25	25—30
f:	7	10	16	32	24	18

Use short-cut method. [Ans. $M = 18.2, M.D. = 4.617$, Coeff. of $M.D. = 0.0308$]

7. Find out the mean deviation from the mean and its coefficient from the following d

Class:	0—3	3—6	6—9	9—12	12—15	15—18	1
f:	2	7	10	12	9	6	1

Use short-cut method. [Ans. $\overline{X} = 10.68, M.D. = 3.8184$, Coeff. of M.D.=

(4) STANDARD DEVIATION

Standard deviation is the most important and widely used measure of dispersion. It w used by Karl Pearson in 1893. Standard deviation is also called as **root mean square de** Standard deviation is defined as the square root of the arithmetic mean of the squares deviation of the values taken from the mean. Standard deviation is denoted by the small Grea σ(read as sigma) and is computed as follows:

$$\sigma = \sqrt{\frac{\Sigma(X-\overline{X})^2}{N}} \quad \text{or} \quad \sqrt{\frac{\Sigma x^2}{N}} \qquad \text{where, } x = X - \overline{X},\ \sigma = S.D.$$

The relative measure of standard deviation, called the **coefficient of S.D.** is obtai dividing the standard deviation by the arithmetic average. Thus,

$$\textbf{Coefficient of S.D.} = \frac{\sigma}{\overline{X}}$$

Difference between Mean Deviation and Standard Deviation

Both these measures of dispersion are based on each and every item of the series. But the in the following respects:

(1) Algebraic signs of deviations (+ or –) are ignored while calculating mean deviation in the calculation of standard deviation signs of deviations are not ignored, *i.e.*, they are tak account.

(2) Mean deviation can be computed either from mean, median or mode. The standard de on the other hand, is always computed from the mean because the sum of the squares deviations taken from the mean is minimum.

Calculation of Standard Deviation

Individual Series

In case of individual series, standard deviation can be computed by applying any of t methods:

(1) Actual Mean Method

When deviations are taken from the actual mean, the following formula is used:

$$\sigma = \sqrt{\frac{\Sigma(X-\overline{X})^2}{N}} \quad \text{or} \quad \sqrt{\frac{\Sigma x^2}{N}} \quad \text{where, } x = X - \overline{X}$$

Steps for Calculation

(i) Calculate the actual mean ($\bar{X}$) of the series.

(ii) Then take the deviations of the items from the mean, *i.e.*, find $X-\bar{X}$ and denote these deviations by x.

(iii) Square these deviations and obtain the total, *i.e.*, Σx^2.

(iv) Divide Σx^2 by the total number of items, *i.e.*, N and take the square root of it. The result will give the value of standard deviation.

Example 15. Calculate the standard deviation from the following data:

X: 16, 20, 18, 19, 20, 20, 28, 17, 22, 20

Solution:

Calculation of Standard Deviation

X	$\bar{X}=20$ $x=X-\bar{X}$	x^2
16	–4	16
20	0	0
18	–2	4
19	–1	1
20	0	0
20	0	0
28	8	64
17	–3	9
22	2	4
20	0	0
$N=10,\ \Sigma X=200$	$\Sigma x=0$	$\Sigma x^2=98$

$$\bar{X}=\frac{\Sigma X}{N}=\frac{200}{10}=20$$

$$\sigma=\sqrt{\frac{\Sigma x^2}{N}}=\sqrt{\frac{98}{10}}=\sqrt{9.8}=3.13$$

(2) Assumed Mean Method

When the actual mean is not a whole number but in fraction, then it becomes difficult to take deviations from mean and then obtain the squares of these deviations. To save time and labour, we use assumed mean method or called **short-cut method.** When deviations are taken from assumed mean, the following formula is used:

$$\sigma=\sqrt{\frac{\Sigma d^2}{N}-\left(\frac{\Sigma d}{N}\right)^2} \quad \text{where, } d=X-A$$

Steps for Calculation

(i) Any one of items in the series is taken as assumed mean, A.

(ii) Take the deviations of the items from the assumed mean, *i.e.*, X – A and denote the deviations by '*d*'. Sum up these deviations to obtain Σd.

(iii) Then square these deviations taken from assumed mean and obtain the total, *i.e.*, Σd^2.

(iv) Substitute the value of Σd^2, Σd and N in the above formula. The result will give the value of standard deviation.

Example 16. Calculate the standard deviation of the following series:

7, 10, 12, 13, 15, 20, 21, 28, 29, 35

Use assumed mean method.

Solution: **Calculation of Standard Deviation**

X	A = 20 $d = X - A$	d^2
7	–13	169
10	–10	100
12	–8	64
13	–7	49
15	–5	25
20 = A	0	0
21	1	1
28	8	64
29	9	81
35	15	225
$N = 10$	$\Sigma d = -10$	$\Sigma d^2 = 778$

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

$$= \sqrt{\frac{778}{10} - \left(\frac{-10}{10}\right)^2} = \sqrt{77.8 - 1} = \sqrt{76.8} = 8.76$$

$\therefore$ $\sigma = 8.76$

▶ (3) Method Based on Use of Actual Data

When number of observations are few, standard deviation can be calculated by using the actual data. When this method is used, the following formula is used:

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

Steps for Calculation

(*i*) First, we find the sum of the items, *i.e.*, ΣX.

(*ii*) Then, the values of X are squared up and added to get ΣX^2.

(*iii*) Substitute the values in the above formula. The result will give the value of standard deviation.

Example 17. Calculate the standard deviation from the following series:

X: 16, 20, 18, 19, 20, 20, 28, 17, 22, 20

Solution:

Calculation of Standard Deviation

X	X^2
16	256
20	400
18	324
19	361
20	400
20	400
28	784
17	289
22	484
20	400
$N=10,\ \Sigma X=200$	$\Sigma X^2=4098$

$$\sigma=\sqrt{\frac{\Sigma X^2}{N}-\left(\frac{\Sigma X}{N}\right)^2}=\sqrt{\frac{4098}{10}-\left(\frac{200}{10}\right)^2}$$

$$=\sqrt{409.8-400}=\sqrt{9.8}=3.13$$

Aliter: σ can be calculated by using the formula:

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

Here, $$\overline{X}=\frac{\Sigma X}{N}=\frac{200}{10}=20,\ \Sigma X^2=4098$$

$\therefore$ $$\sigma=\sqrt{\frac{4098}{10}-(20)^2}=\sqrt{409.8-400}$$

$$=\sqrt{9.8}=3.13$$

Discrete Series

For calculating standard deviation in discrete series, the following three methods may be used:

(1) Actual Mean Method

(2) Assumed Mean Method (or Short-cut Method)

(3) Step Deviation Method.

(1) Actual Mean Method

Under this method, deviations of the items are taken from actual mean, *i.e.*, we find $X-\overline{X}$ and denote these deviations by x. Then these deviations are squared and multiplied by their respective frequencies. The following formula is used:

$$\sigma = \sqrt{\frac{\Sigma fx^2}{N}} \qquad \text{where, } x = X - \overline{X}$$

However, this method is rarely used in practice because if the actual mean is in fraction, the calculations become tedious and time consuming.

(2) Assumed Mean Method

When this method is applied, the following formula is used:

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} \qquad \text{where, } d = X - A$$

Steps for Calculation

(*i*) Take the deviations of the items from an assumed mean instead of actual mean, *i.e.*, $X-A$ and denote these deviations by d.

(*ii*) Then multiply these deviations by their respective frequencies and find Σfd.

(*iii*) Now we square up the deviations, *i.e.*, calculate d^2.

(*iv*) Multiply the squared deviations d^2 by the respective frequencies and find Σfd^2.

(*v*) Substitute the values of Σfd^2, Σfd and N in the above formula.

Example 18. Calculate the standard deviation from the data given below:

X:	3	4	5	6	7	8	9
f:	7	8	10	12	4	3	2

Solution:

Calculation of Standard Deviation

X	f	A=6 $d = X - 6$	fd	fd^2
3	7	−3	−21	63
4	8	−2	−16	32
5	10	−1	−10	10
6	12	0	0	0
7	4	+1	+4	4
8	3	+2	+6	12
9	2	+3	+6	18
	$N = 46$		$\Sigma fd = -31$	$\Sigma fd^2 = 139$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{139}{46} - \left(\frac{-31}{46}\right)^2}$$

$$= \sqrt{3.0217 - 0.4541} = \sqrt{2.5676} = 1.602$$

(3) Step Deviation Method

This method is used to simplify calculations. Under it, we divide the deviations taken from assumed mean (d) by the common factor and get step deviation d', *i.e.*, $d' = \frac{d}{i}$. The remaining process remains as such mentioned in assumed mean method. The formula for calculating standard deviation is:

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

Where, $d' = \frac{X - A}{i}$ and i = common factor.

Example 19. Calculate standard deviation from the following data:

X:	10	20	30	40	50	60	70
f:	3	5	7	9	8	5	3

Solution: **Calculation of Standard Deviation**

X	f	A = 40 $d = X - 40$	$d' = \frac{d}{10}$	fd'	fd'^2
10	3	–30	–3	–9	27
20	5	–20	–2	–10	20
30	7	–10	–1	–7	7
40 = A	9	0	0	0	0
50	8	+10	+1	+8	8
60	5	+20	+2	+10	20
70	3	+30	+3	+9	27
	$N = 40$			$\Sigma fd' = +1$	$\Sigma fd'^2 = 109$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{109}{40} - \left(\frac{1}{40}\right)^2} \times 10$$

$$= \sqrt{2.725 - 0.000625} \times 10 = \sqrt{2.724375} \times 10$$

$$= 1.65 \times 10 = 16.5$$

Continuous Series

In continuous series, we can use any of the three methods discussed above for discrete series because the classes are represented by their mid-values. However, in practice, it is the **step deviation method** which is most commonly used. The formula for calculating standard deviation is

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

Where, $d' = \frac{m - A}{i}$ and i = Size of class interval or common factor, m = mid-value of the classes.

► **Steps for Calculation**

(*i*) Find out the mid-point of the various classes.

(*ii*) Then take the deviations of these mid-points from assumed mean and denote these deviations by '*d*'.

(*iii*) If the class intervals are equal, then divide with common factor (*i*) and get $d' = \frac{d}{i}$

(*iv*) Find fd' and fd'^2.

(*v*) Obtain $\Sigma fd'$ and $\Sigma fd'^2$.

(*vi*) Substitute the values of $\Sigma fd'^2$, $\Sigma fd'$ and N in the above formula.

Thus, the only difference in the procedure in case of continuous series is to find the mid-values of the various classes.

Example 20. Calculate mean and standard deviation for the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80
No. of students:	5	10	20	40	30	20	10	4

Solution: **Calculation of Mean and Standard Deviation**

Marks	f	M.V. (m)	$d = m - 35$	$d' = \frac{d}{10}$	fd'	fd'^2
0—10	5	5	−30	−3	−15	45
10—20	10	15	−20	−2	−20	40
20—30	20	25	−10	−1	−20	20
30—40	40	35 = A	0	0	0	0
40—50	30	45	+10	+1	+30	30
50—60	20	55	+20	+2	+40	80
60—70	10	65	+30	+3	+30	90
70—80	4	75	+40	+4	+16	64
	$N = 139$				$\Sigma fd' = 61$	$\Sigma fd'^2 = 3$

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times i = 35 + \frac{61}{139} \times 10$$

$$= 35 + 4.38 = 39.38$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{369}{139} - \left(\frac{61}{139}\right)^2} \times 10$$

$$= \sqrt{2.6546 - 0.1925} \times 10 = \sqrt{2.4621} \times 10$$

$$= 1.569 \times 10 = 15.69$$

Example 21. Find the mean and standard deviation for the following data:

Age (under):	10	20	30	40	50	60
No. of persons:	15	32	51	78	97	109

Solution: Since the cumulative frequencies are given, firstly we find the simple frequencies:

Age	f	M.V. (m)	d	$d' = \frac{d}{10}$	fd'	fd'^2
0—10	15	5	−30	−3	−45	135
10—20	17	15	−20	−2	−34	68
20—30	19	25	−10	−1	−19	19
30—40	27	35 A	0	0	0	0
40—50	19	45	+10	+1	19	19
50—60	12	55	+20	+2	24	48
	$N = 109$				$\Sigma fd' = -55$	$\Sigma fd'^2 = 289$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 35 - \frac{55}{109} \times 10 = 29.95$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{289}{109} - \left(\frac{-55}{109}\right)^2} \times 10$$

$$= \sqrt{2.6514 - 0.2546} \times 10 = \sqrt{2.3968} \times 10 = 1.548 \times 10 = 15.48$$

IMPORTANT TYPICAL EXAMPLE

Example 22. A welfare organisation introduced an educational scholarship scheme for the school going children of a backward village. The rates of scholarship were fixed as per table:

Age group (in completed years):	5—7	8—10	11—13	14—16	17—19
Amount of scholarship p.m. (in Rs.):	30	40	50	60	70

The ages of 30 school going children were noted as: 11, 8, 10, 5, 7, 12, 7, 17, 5, 13, 9, 8, 10, 15, 7, 12, 6, 7, 8, 11, 14, 18, 6, 13, 9, 10, 6, 15, 13, 5 years respectively. Calculate mean and S.D. of monthly scholarship.

Solution: Firstly we classify data into different groups:

Age group	Tally Bars	No. of children (f)	Amount of Scholarship (X)
5—7	~~IIII~~ ~~IIII~~	10	30
8—10	~~IIII~~ III	8	40
11—13	~~IIII~~ II	7	50
14—16	III	3	60
17—19	II	2	70
		$N = 30$	

Calculation of Mean and Standard Deviation

Amt. of Scholarship (X)	No. of Children (f)	$A = 50$ $d = X - 50$	$d' = \frac{d}{10}$	fd'	fd'^2
30	10	−20	−2	−20	40
40	8	−10	−1	−8	8
50	7	0	0	0	0
60	3	+10	+1	3	3
70	2	+20	+2	4	8
	$N = 30$			$\Sigma fd' = -21$	$\Sigma fd'^2 = 59$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i$$

$$= 50 - \frac{21}{30} \times 10 = 50 - \frac{210}{30} = 50 - 7 = 43$$

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

$$= \sqrt{\frac{59}{30} - \left(\frac{-21}{30}\right)^2} \times 10 = \sqrt{1.966 - 0.49} \times 10$$

$$= \sqrt{1.476} \times 10 = 1.214 \times 10 = 12.14$$

Variance

Variance is another measure of dispersion. The term variance was first used by R.A. Fisher in 1918. **Variance is the square of the standard deviation.** Symbolically,

$$\textbf{Variance} = (S.D.)^2 = \sigma^2$$

Calculation of Variance

(*i*) $\text{Variance} = \frac{\Sigma f(X - \overline{X})^2}{N}$ (Actual Mean Method)

(*ii*) $\text{Variance} = \frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2$ (Assumed Mean Method)

(*iii*) $\text{Variance} = \left[\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2\right] \times i^2$ (Step Deviation Method)

Example 23. Calculate the mean and variance from the data given below:

Daily wages:	0—10	10—20	20—30	30—40	40—50
No. of workers:	2	7	10	5	3

Solution:

Calculation of Mean and Variance

Daily wages	f	M.V. (m)	$A = 25$, $d = m - 25$	$d' = \frac{d}{10}$	fd'	fd'^2
0—10	2	5	−20	−2	−4	8
10—20	7	15	−10	−1	−7	7
20—30	10	25 = A	0	0	0	0
30—40	5	35	+10	+1	+5	5
40—50	3	45	+20	+2	+6	12
	$N = 27$				$\Sigma fd' = 0$	$\Sigma fd'^2 = 32$

$$\overline{X} = = A + \frac{\Sigma fd'}{N} \times i = 25 + \frac{0}{7} \times 10 = 25$$

$$\text{Variance} = \left[\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2\right] \times i^2 = \left[\frac{32}{27} - \left(\frac{0}{27}\right)^2\right] \times 10^2$$

$$\Rightarrow \qquad \sigma^2 = 1.185 \times 100 = 118.51$$

$$\therefore \qquad \overline{X} = 25, \ \sigma^2 = 118.51$$

EXERCISE 6.4

1. Calculate the standard deviation from the following data:

X:	63	67	64	59	61	67	68	66	63	61	68	61

[**Ans.** $\sigma = 3$]

2. Calculate the standard deviation from the following data using assumed mean method:

X:	48	75	54	60	63	69	72	51	57	56

[**Ans.** $\sigma = 8.62$]

3. Calculate the mean and standard deviation for the following data:

Size:	10	20	30	40	50	60	70
Frequency:	6	8	16	15	33	11	12

[**Ans.** $\overline{X} = 44.059, \sigma = 16.36$]

4. Calculate mean and standard deviation of the following series:

Daily wages:	0—10	10—20	20—30	30—40	40—50
No. of workers:	2	7	10	5	3

[**Ans.** $\overline{X} = 25, \sigma = 10.88$]

Calculate median and S.D. from the following data:

Variable:	21—25	26—30	31—35	36—40	41—45	46—50	51—55
Frequency:	5	15	28	42	15	12	3

[**Ans.** $M = 36.928, \sigma = 6.735$]

6. Calculate the mean and the standard deviation of the following series:

Marks (Above):	0	10	20	30	4-0	50	60	70
No. of students :	100	90	75	50	25	15	5	0

[Ans. $\overline{X} = 31, \sigma = 15.9$]

7. Calculate the mean and standard deviation from the following data:

Class Interval:	–40 to – 30	–30 to –20	–20 to –10	–10 to 0	0 to 10	10 to 20	20 to 30
Frequency:	10	28	30	42	65	180	10

[Ans. $\overline{X} = 4.29, \sigma = 14.75$]

8. The following table gives the marks obtained by a group of 80 students in an examination. Calculate the mean and variance.

Marks obtained	No. of students	Marks obtained	No. of students
10—14	2	34—38	10
14—18	4	38—42	8
18—22	4	42—46	4
22—26	8	46—50	6
26—30	12	50—54	2
30—34	16	54—58	4

[Ans. $\overline{X} = 33.5, \sigma^2 = 110.144$]

9. A charitable organisation decided to give old age pensions to people over 60 years of age. The scale of pension were fixed as follows:

Age group:	60—65	65—70	70—75	75—80	80—85
Pension per month (Rs.):	20	25	30	35	40

The ages of 25 persons who secured the pensions' rights are as given below:
74, 62, 84, 72, 61, 83, 72, 81, 64, 71, 63, 61, 60, 67, 74, 64, 79, 73, 75, 76, 69, 68, 78, 66, 67.
Calculate mean and S.D. of monthly pension. [Ans. $\overline{X} = 28.20, \sigma = 6.765$]

Combined Standard Deviation

Just as it is possible to calculate combined mean of two or more groups, similarly the combined standard deviation of two or more groups can be calculated. The combined standard deviation of two groups is denoted by σ_{12} and is computed as follows:

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

Where, σ_{12} = combined standard deviation;

σ_1 = standard deviation of the first group;

σ_2 = standard deviation of the second group;

$d_1 = \overline{X}_1 - \overline{X}_{12}$, $d_2 = \overline{X}_2 - \overline{X}_{12}$

The above formula can be extended to calculate the standard deviation of three or more groups. For example, combined S.D. of three groups is given by:

$$\sigma_{123} = \sqrt{\frac{N_1\,\sigma_1^2 + N_2\,\sigma_2^2 + N_2\,\sigma_3^2 + N_1\,d_1^2 + N_2\,d_2^2 + N_3\,d_3^2}{N_1 + N_2 + N_3}}$$

Where, $d_1 = \overline{X}_1 - \overline{X}_{123}$; $d_2 = \overline{X}_2 - \overline{X}_{123}$; $d_3 = \overline{X}_3 - \overline{X}_{123}$

Example 24. Two samples of size 100 and 150 respectively have means 50 and 60 and standard deviations 5 and 6. Find the mean and standard of the combined sample of size 250.

Solution: Given, $N_1 = 100 \quad \overline{X}_1 = 50, \quad \sigma_1 = 5$

$N_2 = 150 \quad \overline{X}_2 = 60, \quad \sigma_2 = 6$

Now, $$\overline{X}_{12} = \frac{N_1\,\overline{X}_1 + N_2\,\overline{X}_2}{N_1 + N_2} = \frac{100\times 50+150\times 60}{100+150}$$

$$= \frac{100\times 50+150\times 60}{100+150}$$

$$= \frac{5000+9000}{250} = \frac{14000}{250} = 56$$

$$d_1 = \overline{X}_1 - \overline{X}_{12} = 50-56 = -6$$

$$d_2 = \overline{X}_2 - \overline{X}_{12} = 60-56 = +4$$

$$\sigma_{12} = \sqrt{\frac{N_1\,\sigma_1^2 + N_2\,\sigma_2^2 + N_1\,d_1^2 + N_2\,d_2^2}{N_1 + N_2}}$$

$$= \sqrt{\frac{100\times 25+150\times 36+100(-6)^2+150(4)^2}{100+150}}$$

$$= \sqrt{\frac{2500+5400+3600+2400}{250}}$$

$$= \sqrt{\frac{13900}{250}} = 7.46$$

Hence, the combined mean is 56 and standard deviation is 7.46.

Example 25. For a group containing 100 observations, the arithmetic mean and standard deviation are 8 and $\sqrt{10.5}$. For 50 observations selected from the 100 observations, the mean and standard deviations are 10 and 2 respectively. Find the arithmetic mean and the standard deviations of the other half.

Solution: Given: $N = 100, \quad \overline{X}_{12} = 8 \quad \sigma_{12} = \sqrt{10.5}$

$N_1 = 50, \quad \overline{X}_1 = 10 \quad \sigma_1 = 2$

$N_2 = 100 - N_1 = 100 - 50 = 50$

We know that:

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

$$8 = \frac{50(10)+50(\overline{X}_2)}{100}$$

$$800 = 500 + 50\,\overline{X}_2$$

$$300 = 50\,\overline{X}_2$$

$$\therefore \quad \overline{X}_2 = \frac{300}{50} = 6$$

$$d_1 = \overline{X}_1 - \overline{X}_{12} = 10 - 8 = 2 \Rightarrow d_1^2 = 4$$

$$d_2 = \overline{X}_2 - \overline{X}_{12} = 6 - 8 = -2 \Rightarrow d_2^2 = 4$$

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

Substituting the values, we get

$$\sqrt{10.5} = \sqrt{\frac{50\times 4 + 50\sigma_2^2 + 50\times 4 + 50\times 4}{100}}$$

Squaring both sides,

$$10.5 = \frac{200 + 50\sigma_2^2 + 200 + 200}{100}$$

$$10.5 \times 100 = 600 + 50\sigma_2^2$$

$$1050 = 600 + 50\sigma_2^2$$

$$\therefore \quad 50\sigma_2^2 = 450$$

$$\therefore \quad \sigma_2^2 = \frac{450}{50} = 9$$

$$\Rightarrow \quad \sigma_2 = 3$$

Thus, $\overline{X}_2 = 6$, $\sigma_2 = 3$

Example 26. Find the missing information from the following:

	Group I	Group II	Group III	Combined
Number	50	—	90	200
Standard Deviation	6	7	—	7.746
Mean	113	—	115	116

Solution: We are given:

$$N = N_1 + N_2 + N_3 = 200 \quad N_1 = 50,\ N_3 = 90$$

$$\therefore \quad N_2 = N - (N_1 + N_3) = 200 - 140 = 60$$

Now, $$\overline{X}_{123} = \frac{N_1\,\overline{X}_1 + N_2\,\overline{X}_2 + N_3\,\overline{X}_3}{N_1 + N_2 + N_3}$$

We are given: $\overline{X}_1 = 113$, $\overline{X}_3 = 115$, $\overline{X}_{123} = 116$

Substituting the values, we get

$$116 = \frac{(50)(113) + (60)(\overline{X}_2) + (90)(115)}{200}$$

$$116 \times 200 = 50 \times 113 + 60\overline{X}_2 + 90 \times 115$$

$$23200 = 5650 + 60\overline{X}_2 + 10350$$

$$\therefore \quad 60\overline{X}_2 = 23200 - 5650 - 10350 = 7200$$

$$\therefore \quad \overline{X}_2 = \frac{7200}{60} = 120$$

$$d_1 = \overline{X}_1 - \overline{X}_{123} = 113 - 116 = -3 \Rightarrow d_1^2 = 9$$

$$d_2 = \overline{X}_2 - \overline{X}_{123} = 120 - 116 = 4 \Rightarrow d_2^2 = 16$$

$$d_3 = \overline{X}_3 - \overline{X}_{123} = 115 - 116 = -1 \Rightarrow d_3^2 = 1$$

$$\sigma_{123} = \sqrt{\frac{N_1\,\sigma_1^2 + N_2\,\sigma_2^2 + N_2\,\sigma_3^2 + N_1\,d_1^2 + N_2\,d_2^2 + N_3\,d_3^2}{N_1 + N_2 + N_3}}$$

We are given: $\sigma_{123} = 7.745$, $\sigma_1 = 6$, $\sigma_2 = 7$

Substituting the values, we get

$$7.746 = \sqrt{\frac{50(36) + 60(49) + 90\sigma_3^2 + 50(9) + 60(16) + 90(1)}{50 + 60 + 90}}$$

$$7.746 = \sqrt{\frac{1800 + 2940 + 90\sigma_3^2 + 450 + 960 + 90}{200}}$$

$$7.746 = \sqrt{\frac{6{,}240 + 90\sigma_3^2}{200}}$$

Squaring both sides,

$$(7.746)^2 = \frac{6{,}240 + 90\sigma_3^2}{200}$$

$$12000 = 6240 + 90\sigma_3^2$$

$$\Rightarrow \quad 90\sigma_3^2 = 12000 - 6240 = 5760$$

$$\Rightarrow \quad \sigma_3^2 = \frac{5760}{90} = 64$$

$$\Rightarrow \quad \sigma_3 = \sqrt{64} = 8$$

Thus, $N_2 = 60$, $\overline{X}_2 = 120$, $\sigma_3 = 8$

IMPORTANT TYPICAL EXAMPLE

Example 27. The mean weight of 150 students is 60 kg. The mean weight of boys is 70 kg with a standard deviation of 10 kg. For the girls, the mean weight is 55 kg and the standard deviation is 15 kg. Find the number of boys and girls and the combined standard deviation..

Solution: Given: $N = N_1 + N_2 = 150,\ \bar{x}_{12} = 60$

$$\bar{X}_1 = 70,\ \sigma_1 = 10,\ \bar{X}_2 = 55,\ \sigma_2 = 15$$

We have to determine the number of boys

$$\therefore \quad N_2 = 150 - N_1$$

Here, N_2 will be the number of girls and N_1 will be the number of boys.

We know, $$\bar{X}_{12} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N_1 + N_2}$$

Substituting the values, we get

$$60 = \frac{N_1(70) + (150 - N_1)(55)}{150}$$

$$60 \times 150 = 70\,N_1 + 8250 - 55\,N_1$$

$$9000 = 70\,N_1 + 8250 - 55\,N_1$$

$$9000 = 8250 + 15\,N_1$$

$$\Rightarrow \quad 15N_1 = 9000 - 8250 = 750$$

$$\Rightarrow \quad N_1 = \frac{750}{15} = 50$$

Hence, $$N_2 = 150 - 50 = 100$$

Thus, the number of boys and girls are 50 and 100 respectively.

$$\sigma_{12} = \sqrt{\frac{N_1\,\sigma_1^2 + N_2\,\sigma_2^2 + N_1\,d_1^2 + N_2\,d_2^2}{N_1 + N_2}}$$

Here, $N_1 = 50,\ \sigma_1 = 10,\ N_2 = 100,\ \sigma_2 = 15$

$$d_1 = \bar{X}_1 - \bar{X}_{12} = 70 - 60 = 10 \Rightarrow d_1^2 = 100$$

$$d_2 = \bar{X}_2 - \bar{X}_{12} = 55 - 60 = -5 \Rightarrow d_2^2 = 25$$

Substituting the values, we get

$$\sigma_{12} = \sqrt{\frac{50(100) + 100(225) + 50(100) + 100(25)}{50 + 100}}$$

$$\sigma_{12} = \sqrt{\frac{5000 + 22500 + 5000 + 2500}{150}} = \sqrt{\frac{35000}{150}} = \sqrt{233.33} = 15.28.$$

Thus, combined S.D. is 15.28.

EXERCISE 6.5

1. Two samples of size 40 and 60 respectively have means 20 and 25 and standard deviations 5 and 6 respectively. Find the combined mean and standard deviation of size 100. [Ans. $\bar{X}_{12} = 23, \sigma_{12} = 6.13$]

2. For two groups of observations the following results were available :

Group I	Group II
$\Sigma(X-5) = 8$	$\Sigma(X-8) = -10$
$\Sigma(X-5)^2 = 40$	$\Sigma(X-8)^2 = 70$
$N_1 = 20$	$N_2 = 25$

Find mean and standard deviation of both the groups taken together. [Hint: See Example 45] [Ans. $\bar{X}_{12} = 6.62, \sigma_{12} = 1.864$]

3. The mean height of the students in a class is 152 cm. The mean height of boys is 158 cm with a standard deviation of 5 cm. And the mean height of girls is 148 cm with a standard deviation of 4 cm. Find the percentage of boys in the class and also the S.D of heights of all the students in the class. [Ans. Percentage of boys = 40%, $\sigma_{12} = 6.603$]

4. The first of two subgroups has 100 items with mean 15 and standard deviation 3. If the whole group has 250 items with mean 15.6 and standard deviation $\sqrt{13.44}$, find the mean and standard deviation of the second such group. [Ans. $\bar{X}_2 = 16, \sigma_2 = 4$]

Correcting Incorrect Values of Mean and Standard Deviation

In certain cases, mean and standard deviation are calculated by using one or two incorrect alues of the variable. Just as we can correct an incorrect mean, similarly, there is a procedure of orrecting an incorrect standard deviation.

teps: The various steps in the calculation of correct S.D. are as follows:

(i) Find out incorrected sum of the squared values of the variable, *i.e.*, find ΣX^2. This is to be found by using the following formula which involves incorrected $\bar{X}$ and σ.

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

or
$$\sigma^2 = \frac{\Sigma X^2}{N} - (\bar{X})^2$$

$\therefore$ Incorrected $\Sigma X^2 = N[\sigma^2 + (\bar{X})^2]$

(ii) Find corrected ΣX^2. To do so, we subtract the square of the incorrect item from incorrected ΣX^2 and add the square of correct item to incorrected ΣX^2. Thus,

$$\text{Corrected } \Sigma X^2 = \text{Incorrected } \Sigma X^2 - (\text{Incorrect value})^2 + (\text{Correct value})^2$$

(iii) Apply the following formula:

$$\text{Corrected } \sigma = \sqrt{\frac{\text{Corrected } \Sigma X^2}{N} - (\text{Corrected } \bar{X})^2}$$

Example 28. For a group of 100 observations, the mean and standard deviation were found to be 60 and 5 respectively. Later on it was discovered that a correct item 50 was wrongly copied as 30. Find the correct mean and standard deviation.

Solution: Given: $N = 100, \bar{X} = 60, \sigma = 5$

Calculation of Correct Mean

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

$$60 = \frac{\Sigma X}{100}$$

or $$\Sigma X = 6000$$

Incorrected $\Sigma X = 6000$

Corrected $\Sigma X = 6000 +$ Correct item $-$ Incorrect item

$$= 6000 + 50 - 30 = 6020$$

Hence, Corrected $\bar{X} = \frac{6020}{100} = 60.20$

Calculation of Correct S.D.

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

Putting the values, we get

$$5 = \sqrt{\frac{\Sigma X^2}{100} - (60)^2}$$

Squaring both sides, we get

$$25 = \frac{\Sigma X^2}{100} - 3600$$

$$25 + 3600 = \frac{\Sigma X^2}{100}$$

$\therefore$ Corrected $\Sigma X^2 = 100[25 + 3600] = 362500$

Corrected $\Sigma X^2 = 362500 + (\text{Correct item})^2 - (\text{Incorrect item})^2$

$$= 362500 + (50)^2 - (30)^2$$

$$= 362500 + 2500 - 900$$

$\therefore$ Corrected $\Sigma X^2 = 364100$

Corrected $\sigma = \sqrt{\frac{364100}{100} - (60.20)^2}$

$$= \sqrt{3641.00 - 3624.04}$$

$$= \sqrt{16.96} = 4.12$$

$\therefore$ Corrected $\bar{X} = 60.20$, Corrected $\sigma = 4.12$

IMPORTANT TYPICAL EXAMPLE

Example 29. The mean, standard deviation and range of a symmetrical distribution of weights of a group of 20 boys are 40 kgs, 5 kgs, and 6 kgs respectively. Find the mean and standard deviation of the group if the lightest and heaviest boys are excluded.

Solution: Since, the distribution is given to be symmetrical, the mean will lie at the middle of the range.

Therefore, the weight of the heaviest boy = 40 + 3 = 43 kgs and

the weight of the lightest boy = 40 – 3 = 37 kgs.

We are given that $\bar{X} = 40, \sigma = 5$ and $N = 20$

$$\bar{X} = \frac{\Sigma X}{N} \quad \text{or} \quad 40 = \frac{\Sigma X}{20} \Rightarrow \Sigma X = 800$$

$$\text{Corrected } \Sigma X = 800 - 43 - 37 = 720.$$

$$\therefore \quad \text{Corrected } \bar{X} = \frac{720}{18} = 40.$$

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

$$5 = \sqrt{\frac{\Sigma X^2}{20} - (40)^2}$$

Squaring both sides,

$$(5)^2 = \frac{\Sigma X^2}{20} - (40)^2$$

$$\therefore \quad \Sigma X^2 = 20[25 + 1{,}600] = 20 \times 1625 = 32{,}500$$

$$\text{Corrected } \Sigma X^2 = 32{,}500 - 43^2 - 37^2 = 29{,}282$$

$$\text{Corrected } \sigma = \sqrt{\frac{29{,}282}{18} - (40)^2} = \sqrt{26.777} = 5.17$$

$$\therefore \quad \text{Corrected } \sigma = 5.17$$

EXERCISE 6.6

1. A student obtained the mean and standard deviation of 100 observations as 40 and 5 respectively. It was later found that one observation was wrongly copied as 50, the correct figure being 40. Find the correct mean and standard deviation. **[Ans.** $\bar{X} = 39.3, \sigma = 4.9$**]**
2. During nine days in a festival the highest sale of a shop was on Sunday and Rs 90 more than the average sale for other days. If the standard deviation of the sale during the festival is 33.33, find the standard deviation leaving that the highest sale.
[Hint: See Example 49] **[Ans.** $\sigma = 15.4$**]**

3. The mean age and standard deviation of a group of 200 persons (grouped in intervals 0—5, 5—10, ... etc.) were found to be 40 and 15. Later on it was discovered that the age 43 was misread as 53. Find the correct mean and standard deviation.
[Hint: See Example 53] [Ans. $\overline{X} = 39.95, \sigma = 14.97$]

4. The mean and standard deviation of 20 items is found to be 10 and 2 respectively. At the time of checking, it was found that one item 8 was incorrect. Calculate mean and standard deviation if: (*i*) the wrong item is omitted. (*ii*) it is replaced by 12.
[Ans. (*i*) $\overline{X} = 10.11, \sigma = 1.997$; (*ii*) $\overline{X} = 10.2, \sigma = 1.99$]

Determination of Missing Values

In certain situations, the values of one or more items may be missing from the given information. The method of computing missing values is explained with the help of the following examples:

Example 30. The mean of 5 observations is 4.4 and the variance is 8.24. If three of the observations are 4, 6 and 9, find the other two.

Solution: Let the missing observations be x_1 and x_2

X	X^2
4	16
6	36
9	81
x_1	x_1^2
x_2	x_2^2
$\Sigma X = x_1 + x_2 + 19$	$\Sigma X^2 = 133 + x_1^2 + x_2^2$

Here we are given $N = 5, \overline{X} = 4.4, \sigma^2 = 8.24$

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

$$\therefore \quad 4.4 = \frac{x_1 + x_2 + 19}{5} \quad \text{or} \quad x_1 + x_2 + 19 = 22$$

$$\therefore \quad x_1 + x_2 = 3 \Rightarrow x_2 = 3 - x_1$$

Now, $$(S.D.)^2 = Variance = \frac{\Sigma X^2}{N} - (\overline{X})^2$$

$$8.24 = \frac{133 + x_1^2 + x_2^2}{5} - (4.4)^2$$

$$\therefore \quad 133 + x_1^2 + x_2^2 = 5[8.24 + 19.36]$$

$$\Rightarrow \quad 133 + x_1^2 + x_2^2 = 138$$

$$\Rightarrow \quad x_1^2 + x_2^2 = 5$$

From (*i*) and (*ii*), $x_1^2 + (3 - x_1)^2 = 5$

$\Rightarrow x_1^2 + 9 + x_1^2 - 6x_1 = 5 \Rightarrow 2x_1^2 - 6x_1 + 4 = 0$

$\Rightarrow \quad x_1^2 - 3x_1 + 2 = 0 \quad \Rightarrow \quad (x_1 - 1)(x_1 - 2) = 0$

$\therefore \quad x_1 = 1, 2$

If $x_1 = 1, x_2 = 2$ and if $x_1 = 2, x_2 = 1$.

Example 31. Mean and standard deviation of the following continuous series are 135.3 and 9.6 respectively. The distribution after taking step deviations is as follows:

d:	–4	–3	–2	–1	0	1	2	3
f:	2	5	8	18	22	13	8	4

Determine the actual class intervals.

Solution: Here, d is identical to d' which is referred as $d' = \dfrac{X - A}{i}$.

In order to ascertain the class intervals, we need two values—size of the class interval (*i*) and assumed mean (A). From the formula of S.D., we can determine the size of class interval (*i*) and from the formula of mean, we can determine the assumed mean (A).

Calculations for Determining *i* and *A*

d	*f*	*fd*	fd^2
–4	2	–8	32
–3	5	–15	45
–2	8	–16	32
–1	18	–18	18
0	22	0	0
1	13	13	13
2	8	16	32
3	4	12	36
	$N = 80$	$\Sigma fd = -16$	$\Sigma fd^2 = 208$

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

$$9.6 = \sqrt{\frac{208}{80} - \left(\frac{-16}{80}\right)^2} \times i$$

$$\Rightarrow \quad 9.6 = \sqrt{2.6 - 0.04} \times i$$

$$\Rightarrow \quad 9.6 = \sqrt{2.56} \times i$$

$$9.6 = 1.6 \times i$$

$$i = \frac{9.6}{1.6} = 6$$

$$\overline{X} = A + \frac{\Sigma fd}{N} \times i$$

$$135.3 = A + \frac{(-16)}{80} \times 6 = A - 1.2$$

$$\therefore \quad A = 136.5$$

Using the values of A, i and d, we can write the mid-values as given below:

$$d = \frac{X - A}{i} \Rightarrow di = X - A$$

$$\Rightarrow \quad X = A + id$$

Here, $A = 136.5$, $i = 6$.

d:	–4	–3	–2	–1	0	1	2	3
M.V. $(X) = A + id$	136.5 + (6) (–4) = 112.5	118.5	124.5	130.3	136.5	142.5	148.5	154.5

The various class intervals shall be obtained by using the formula:

$$m \pm \frac{i}{2}$$

The various class intervals are:

$$112.5 \pm \frac{6}{2}, \quad 118.5 \pm \frac{6}{2}, \quad 124.5 \pm \frac{6}{2}, \quad 130.5 \pm \frac{6}{2},$$

$$136.5 \pm \frac{6}{3}, \quad 142.5 \pm \frac{6}{3}, \quad 148.5 \pm \frac{6}{3}, \quad 154.5 \pm \frac{6}{2}$$

i.e., $109.5 - 115.5$, $115.5 - 121.5$, $121.5 - 127.5$, $127.5 - 133.5$,
$133.5 - 139.5$, $139.5 - 145.5$, $145.5 - 151.5$, $151.5 - 157.5$

Thus,

Class Intervals	f	Class Intervals	f
109.5—115.5	2	133.5—139.5	22
115.5—121.5	5	139.5—145.5	13
121.5—127.5	8	145.5—151.5	8
127.5—133.5	18	151.5—157.5	4

EXERCISE 6.7

1. The mean of 5 observations is 4.4 and variance is 8.24. If three of five observations are 1, 2, 6 find the other two. [Ans. If $x_1 = 9, x_2 = 4$ and if $x_1 = 4, x_2 = 9$]
2. Mean and S.D. of the following continuous series are 31 and 15.9. The distribution after taking step deviations is as follows:

d:	–3	–2	–1	0	1	2	3
f:	10	15	25	25	10	10	5

Determine the actual class intervals.

[Ans. 0—10, 10—20, 20—30, 30—40, 40—50, 50—60, 60—

Mathematical Properties of Standard Deviation

The important mathematical properties of standard deviation are as follows:

(1) The standard deviation of first n natural numbers can be found from the following formula:

$$\sigma = \sqrt{\frac{1}{12}\cdot(n^2-1)}$$

For example, the standard deviation of the first 5 natural numbers is given as:

$$\sigma = \sqrt{\frac{1}{12}\cdot(5^2-1)} = \sqrt{\frac{24}{12}} = \sqrt{2} = 1.414$$

(2) The combined S.D. of two or more groups can be found by using the following formula:

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1d_1^2 + N_2d_2^2}{N_1+N_2}} \quad \text{where, } d_1 = \overline{X}_1 - \overline{X}_{12},\ d_2 = \overline{X}_2 - \overline{X}_{12}$$

(3) The sum of the squares of the deviations of the items taken from arithmetic mean is least. That is why standard deviation is computed from the A.M.

(4) If a constant amount 'a' is added or subtracted from each item of a series, then S.D. remains unaffected, *i.e.*, S.D. is independent of the change of origin.

(5) If each item of a series is multiplied or divided by a constant 'a', then S.D. is affected by the same amount, *i.e.*, S.D. is not independent of the change of scale.

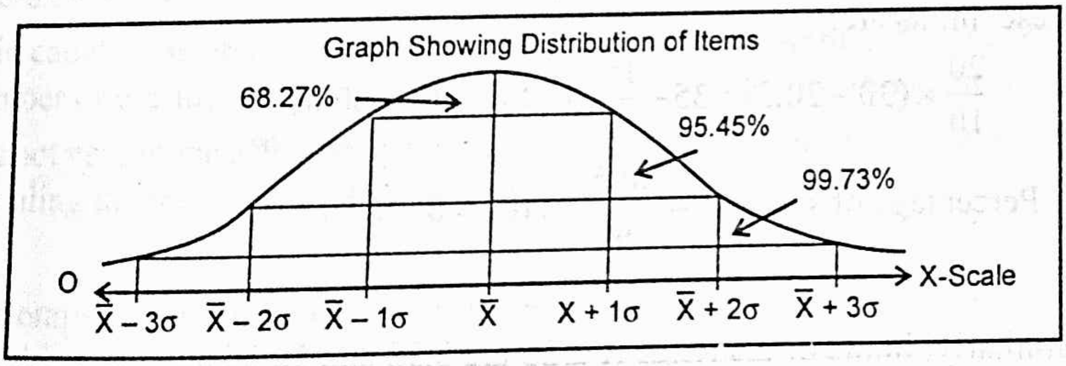
(6) The standard deviation has the following relation to the arithmetic mean in a symmetrical distribution:

$\overline{X} \pm 1\sigma$ includes 68.27% of the items.

$\overline{X} \pm 2\sigma$ includes 95.45% of the items.

$\overline{X} \pm 3\sigma$ includes 99.73% of the items.

The following figure illustrate the relationship:



(7) The standard deviation has the following relation to quartile deviation (Q.D.) and mean deviation (M.D.) in a symmetrical (or normal) distribution:

$$Q.D. = \frac{2}{3}\sigma,\quad M.D. = \frac{4}{5}\sigma,\quad Q.D : M.D : S.D :: 10 : 12 : 15$$

IMPORTANT TYPICAL EXAMPLES

Example 32. The following table gives the distribution of marks obtained by 90 students in an examination:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of Students:	4	10	20	35	15	6

Calculate (*i*) Mean, (*ii*) Standard deviation and (*iii*) Percentage of students lying within the range (*a*) $\bar{X} \pm 1\sigma$ and (*b*) $\bar{X} \pm 2\sigma$.

Solution:

Calculation of $\bar{X}$ and σ

Classes	f	M.V. (m)	$d = m - A$	$d' = \frac{d}{i}$	fd'	fd'^2
0—10	4	5	–30	–3	–12	36
10—20	10	15	–20	–2	–20	40
20—30	20	25	–10	–1	–20	20
30—40	35	35 = A	0	0	0	0
40—50	15	45	+10	+1	15	15
50—60	6	55	+20	+2	12	24
	$N = 90$				$\Sigma fd' = -25$	$\Sigma fd'^2 = 135$

(*i*) $\bar{X} = A + \frac{\Sigma fd'}{N} \times i = 35 + \frac{(-25)}{90} \times 10 = 32.22$

(*ii*) $\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{135}{90} - \left(\frac{-25}{90}\right)^2} \times 10 = 11.92$

(*iii*) $\bar{X} \pm 1\sigma = 32.22 \pm 11.92$

The limit of the range $\bar{X} \pm \sigma$ are 20.3 and 44.14. Under the assumption the observations in a class are uniformly distributed, the number of students lying within these limits are:

$$\frac{20}{10} \times (30 - 20.3) + 35 + \frac{15}{10}(44.14 - 40) = 60.61$$

$\therefore$ Percentage of students $= \frac{60.61}{90} \times 100 = 67.34\%$

$$\bar{X} \pm 2\sigma = 32.22 \pm 2 \times 11.92$$

Similarly, limits of the range $\bar{X} \pm 2\sigma$ are 8.38 and 56.06 and the number of students lying within these limits are:

$$\frac{4}{10} \times (10 - 8.38) + 10 + 20 + 35 + 15 + \frac{6}{10} \times (56.06 - 50) = 84.29$$

$\therefore$ Percentage of students $= \frac{84.29}{90} \times 100 = 93.66\%$

Example 33. You are the incharge of the rationing department of a state affected by food shortage. The following information is received from your local investigators :

Area	Mean Calories	Standard Deviation of Calories
X	2,500	500
Y	2,200	300

The estimated requirement of an adult is taken at 3,000 calories daily and absolute minimum at 1,250. Comment on the reported figures and determine which area needs more urgent action.

Solution: We shall compute the 3-sigma limits $\overline{X}\pm3\sigma$ for each area, which will include approximately 99.73% of the population observation [assuming that the distribution is approximately normal].

	3-σ Limits = $\overline{X}\pm3\sigma$
Area X	$2500 \pm 3 \times 500 = 2500 \pm 1500 = (1000, 4000)$
Area Y	$2200 \pm 3 \times 300 = 2200 \pm 900 = (1300, 3100)$

The absolute daily minimum calories requirement for a person is 1250. From the above figures we observe that almost all the persons in the area Y are getting more than the minimum calories requirement as the lower limit in this area is 1300. However, since in the area X, the lower 3-σ limit is 1000 which is less than 1250, quite a number of people in area X are not getting the minimum requirement of 1250 calories. Hence, as the incharge of the rationing department, it becomes my duty to take urgent action for the people of area X.

Merits and Demerits of Standard Deviation

erits

(*i*) It is a rigidly defined.

(*ii*) It is based on all the observations.

(*iii*) It is capable of being treated mathematically. For example, if standard deviations of a number of groups are known, their combined standard deviation can be computed.

(*iv*) It is not very much affected by the fluctuations of sampling and, therefore, is widely used in sampling theory and test of significance.

merits

(*i*) As compared to the quartile deviation and range, etc., it is difficult to understand and difficult to calculate.

(*ii*) It gives more importance to extreme observations.

(*iii*) Since, it depends upon the units of measurement of the observations, it cannot be used to compare the dispersions of the distributions expressed in different units.

EXERCISE 6.8

1. Calculate the S.D. of the first 7 natural numbers. [Ans. σ =
2. If mean and standard deviation of 75 observations is 40 and 8 respectively, find the ne mean and standard deviation if :
   (*i*) Each observation is multiplied by 5.
   (*ii*) 7 is added to each observation.
   [Hint: See Example 50] [Ans. (*i*) New mean = 200, New S.D. = 4
   (*ii*) New mean = 47, New S.D. = 8
3. 5 observations of a series are 4, 6, 8, 12 and 15. Their mean and standard deviation are 9 an 4 respectively. Make such alternations in the terms of the series that new standard deviatio is 20 and mean is 50. [Ans. 25, 35, 45, 65 and 8
   [Hint: See Example 51]
4. The following table gives the length of life of 300 persons:

Age (X):	0—9	10—19	20—29	30—39	40—49	50—59	60—69	70—79	80—89
No. of persons:	6	15	33	39	45	27	18	10	7

Calculate (*i*) Mean (*ii*) Standard deviation (*iii*) The percentage of persons whose length o life falls within $\overline{X} \pm 2\sigma$. [Ans. $\overline{X} = 41.85, \sigma = 18.5, 95\%$

5. From the following figures determine the percentage of cases which lie outside the range:
   $\overline{X} \pm \sigma, \overline{X} \pm 2\sigma, \overline{X} \pm 3\sigma$.
   115, 117, 121, 125, 116, 120, 118, 117, 119, 116, 122, 124, 123, 118, 120, 118, 126, 127, 122, 123.
   [Ans. $\overline{X} = 120.35, \sigma = 3.45$, 3.5%, 0%, 0%]
6. A collar manufacturer is considering the production of a new style of collar to attract young men. The following statistics of neck circumference are available based on measurements of a typical group of college students. Compute the SD and use the criterion ($\overline{X} \pm 3\sigma$), to determine the largest and smallest sizes of collars, he should make in order to meet the needs of practically all his customers, bearing in mind, that collars are worn, on average $\frac{3}{4}$ inch larger than neck size.

Mid-points:	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0	16.5
f:	4	19	30	63	66	29	18	1	1

[Hints: See Example 52] [Ans. $\overline{X} = 14.232, \sigma = 0.719, \overline{X} \pm 3\sigma + \frac{3}{4} = 12.825$ to 17.139]

◘ (5) COEFFICIENT OF VARIATION

Coefficient of variation is an important relative measure of dispersion. It was developed by Karl Pearson and is widely used in comparing the variability of two or more series. Coefficient of variation is denoted by C.V. and is given by:

$$\textbf{Coefficient of Variation (C.V.)} = \frac{\sigma}{\overline{X}} \times 100$$

Steps for Calculation

(i) First of all calculate $\overline{X}$.
(ii) Calculate σ.
(iii) Put the value of $\overline{X}$ and σ in the above formula.

Uses of Coefficient of Variation

Coefficient of variation is used to compare the variability, homogenity, stability, consistency and uniformity of two or more series. The series having less value of the coefficient of variation is considered more consistent in comparison to a series having a higher value of the coefficient of variation.

Example 34. From the prices of shares of X and Y given below, state which share is more stable in value:

X:	41	44	43	48	45	46	49	50	42	40
Y:	91	93	96	92	90	97	99	94	98	95

Solution: For finding out which share is more stable in value, we have to compare the coefficient of variation.

Calculation of C.V.

X	*A* = 45 *dx*	dx^2	*Y*	*A* = 95 *dy*	dy^2
41	−4	16	91	−4	16
44	−1	1	93	−2	4
43	−2	4	96	1	1
48	3	9	92	−3	9
45 = A	0	0	90	−5	25
46	1	1	97	2	4
49	4	16	99	4	16
50	5	25	94	−1	1
42	−3	9	98	3	9
40	−5	25	95 = A	0	0
$N = 10$ $\Sigma X = 448$	$\Sigma dx^2 = -2$	$\Sigma dx^2 = 106$	$\Sigma Y = 945$	$\Sigma dy = -5$	$\Sigma dy^2 = 85$

Share X: $\overline{X} = \dfrac{\Sigma X}{N} = \dfrac{448}{10} = 44.8$

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

$$= \sqrt{\frac{106}{10} - \left(\frac{-2}{10}\right)^2} = 3.25$$

$$\therefore C.V._X = \frac{3.25}{44.8} \times 100 = 7.25\%$$

Share Y: $Y = \dfrac{\Sigma Y}{N} = \dfrac{945}{10} = 94.5$

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

$$= \sqrt{\frac{85}{10} - \left(\frac{-5}{10}\right)^2} = 2.87$$

$$\therefore C.V._Y = \frac{2.87}{94.5} \times 100 = 3.03\%$$

Since, the coefficient of variation is less for share Y, hence share Y is more stable in price.

Example 35. The scores of two batsmen A and B in ten innings during a certain match are:

A:	32	28	47	63	71	39	10	60	96	14
B:	19	31	48	53	67	90	10	62	40	80

Find out who is a better scorer and who is more consistent batsman.

Solution: For finding out which of the two batsman is a better scorer, we have to compare the arithmetic means and for finding out which batsman is more consistent, we have to compare the coefficient of variation.

Calculation of $\overline{X}$ and C.V.

X	$\overline{X}=46$ $x=X-\overline{X}$	x^2	Y	$\overline{Y}=50$ $y=Y-\overline{Y}$	y^2
32	−14	196	19	−31	961
28	−18	324	31	−19	361
47	+1	1	48	−2	4
63	+17	289	53	+3	9
71	+25	625	67	+17	289
39	−7	49	90	+40	1600
10	−36	1296	10	−40	1600
60	+14	196	62	+12	144
96	+50	2500	40	−10	100
14	−32	1024	80	+30	900
$N=10$ $\Sigma X=460$	$\Sigma x=0$	$\Sigma x^2=6500$	$\Sigma Y=500$	$\Sigma y=0$	$\Sigma y^2=5968$

Batsman A: $\overline{X}=\dfrac{\Sigma X}{N}=\dfrac{460}{10}=46,$ **Batsman B:** $\overline{Y}=\dfrac{\Sigma Y}{N}=\dfrac{500}{10}=50$

Since the arithmetic mean is higher for batsman B, hence batsman B is a better scorer.

$\sigma=\sqrt{\dfrac{\Sigma x^2}{N}}=\sqrt{\dfrac{6500}{10}}=25.495$ $\sigma=\sqrt{\dfrac{\Sigma y^2}{N}}=\sqrt{\dfrac{5968}{10}}=24.43$

$\therefore\ C.V.=\dfrac{\sigma}{\overline{X}}=\dfrac{25.49}{46}\times 100=55.41\%$ $\therefore\ C.V.=\dfrac{\sigma}{\overline{X}}=\dfrac{24.43}{50}\times 100=48.86\%$

Since, the coefficient variation is less for batsman B, hence batsman B is more consistent.

Example 36. Goals scored by two teams A and B in a football session were as follows:

No. of goals scored:	0	1	2	3	4
No. of matches by A:	27	9	8	5	4
No. of matches by B:	17	9	6	5	3

By calculating the coefficient of variation in each case, find which team may be considered more consistent.

Solution: **For team A:**

X (goals)	f (No. of matches)	$A = 2$ $d = X - A$	fd	fd^2
0	27	–2	–54	108
1	9	–1	–9	9
2 = A	8	0	0	0
3	5	+1	+5	5
4	4	+2	+8	16
	$N = 53$		$\Sigma fd = -50$	$\Sigma fd^2 = 138$

$$\overline{X} = A + \frac{\Sigma fd}{N} = 2 - \frac{50}{53} = 2 - 0.94 = 1.06$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{138}{53} - \left(\frac{-50}{53}\right)^2}$$

$$= \sqrt{2.603 - 0.889} = \sqrt{1.74} = 1.309$$

$$\text{C.V. for Team A} = \frac{\sigma}{\overline{X}} \times 100 = \frac{1.309}{1.06} \times 100 = 123.49\%$$

For team B:

X (goals)	f (No. of matches)	$A = 2$ $d = X - A$	fd	fd^2
0	17	–2	–34	68
1	9	–1	–9	9
2	6	0	0	0
3	5	+1	5	5
4	3	+2	6	12
	$N = 40$		$\Sigma fd = -32$	$\Sigma fd^2 = 94$

$$\overline{X} = A + \frac{\Sigma fdx}{N} = 2 - \frac{32}{40} = 2 - 0.8 = 1.2$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{94}{40} - \left(\frac{-32}{40}\right)^2}$$

$$= \sqrt{2.35 - 0.64} = \sqrt{1.14} = 1.307$$

$$\text{C.V. for Team B} = \frac{\sigma}{\overline{X}} \times 100 = \frac{1.307}{1.2} \times 100 = 108.9\%$$

Since, the coefficient of variation of Team B is less than team A, so team B is more consistent.

Note: Normally, the value of C.V. does not exceed 100% but in Z-shaped distribution, the value of C.V. exceeds 100%

Example 37. You are given below the daily wages paid to workers in two factories X and Y.

Daily wages	No. of workers	
	Factory X	Factory Y
12—13	15	25
13—14	30	40
14—15	44	60
15—16	60	35
16—17	30	12
17—18	14	15
18—19	7	5

Using appropriate measures, answer the following:

(*i*) Which factory pays higher average wages?

(*ii*) Which factory has a more consistent wage structure?

Solution: For finding out which factory pays higher wages, we have to compute the arithmetic means and for finding out which factory has a more consistent wage structure, we have to compare the coefficient of variation:

Calculation of $\overline{X}$ and C.V.

Wages	M.V. (m)	$d = m - A$	Factory X			Factory Y		
			f	fd	fd^2	f	fd	fd^2
12—13	12.5	− 3	15	− 45	135	25	− 75	225
13—14	13.5	− 2	30	− 60	120	40	− 80	160
14—15	14.5	− 1	44	− 44	44	60	− 60	60
15—16	15.5	0	60	0	0	35	0	0
16—17	16.5	+ 1	30	+30	30	12	+ 12	12
17—18	17.5	+ 2	14	+28	56	15	+ 30	60
18—19	18.5	+ 3	7	+21	61	5	+ 15	45
			$N = 200$	$\Sigma fd = -70$	$\Sigma fd^2 = 448$	$N = 192$	$\Sigma fd = -158$	$\Sigma fd^2 = 562$

(*i*) **Factory X** **Factory Y**

$$\overline{X} = A + \frac{\Sigma fd}{N} = 15.5 - \frac{70}{200} = 15.15 \qquad \overline{Y} = A + \frac{\Sigma fd}{N} = 15.5 - \frac{158}{192} = 14.67$$

Since, the arithmetic mean is higher for factory X, hence factory X pays higher average wage.

(ii) **Factory X**

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

$$= \sqrt{\frac{448}{200} - \left(\frac{-70}{200}\right)^2}$$

$$= \sqrt{2.24 - 0.1225} = 1.445$$

$$C.V._X = \frac{\sigma}{\overline{X}} \times 100$$

$$= \frac{1.455}{15.15} \times 100 = 9.60\%$$

Factory Y

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

$$= \sqrt{\frac{562}{192} - \left(\frac{-158}{192}\right)^2}$$

$$\sigma = \sqrt{2.93 - 0.677} = 1.504$$

$$C.V._Y = \frac{\sigma}{\overline{Y}} \times 100$$

$$= \frac{1.504}{14.67} \times 100 = 10.25\%$$

Since, the coefficient of variation is less for factory X, hence factory X has more consistent wage structure.

Example 38. An analysis of the monthly wages paid to workers in firm A and B belonging to the same industry gives the following results:

	Firm A	**Firm B**
No. of workers:	500	600
Average monthly wage (Rs.):	186	175
Variance of distribution of wages (Rs.)	81	100

(i) Which firm pays a larger wage bill?

(ii) In which firm is there greater variability in individual wages?

(iii) Find the combined mean and standard deviation of wages of the two firms taken together.

Solution: (i) **Total wage bill of firm A**

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

$\therefore$ Total wage (ΣX) bill of firm A $= \overline{X} \times N = 186 \times 500 =$ Rs. 93,000.

Total wage bill of firm B

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

Total wage (ΣY) bill of firm B $= \overline{Y} \times N = 175 \times 600 =$ Rs. 1,05,000.

Hence, firm B pays larger wage bill.

(*ii*) To determine the firm in which there is greater variability in individual wages, we shall compare the coefficient of variation.

Firm A

$$C.V. = \frac{\sigma}{\overline{X}} \times 100$$

Given: $\sigma^2 = 81 \Rightarrow \sigma = \sqrt{81} = 9$, $\overline{X} = 186$

$$\therefore C.V._A = \frac{9}{186} \times 100 = 4.84\%$$

Firm B

$$C.V. = \frac{\sigma}{\overline{Y}} \times 100$$

Given: $\sigma^2 = 100 \Rightarrow \sigma = 10, \overline{Y} = 175$

$$\therefore C.V._B = \frac{10}{175} \times 100 = 5.71\%$$

Since, the coefficient of variation is greater in case of firm B, there is greater variability in individual wages of firm B.

(*iii*) **Combined Mean and Standard Deviation.**

$$\overline{X}_{12} = \frac{N_1 \overline{X}_1 + N_2 \overline{X}_2}{N_1 + N_2} = \frac{500 \times 186 + 600 \times 175}{500 + 600}$$

$$\overline{X}_{12} = \frac{93{,}000 + 1{,}05{,}000}{1{,}100} = \frac{1{,}98{,}000}{1{,}100} = \text{Rs. } 180$$

$$d_1 = \overline{X}_1 - \overline{X}_{12} = 186 - 180 = 6 \quad \Rightarrow d_1^2 = 36$$

$$d_2 = \overline{X}_2 - \overline{X}_{12} = 175 - 180 = -5 \Rightarrow d_2^2 = 25$$

$$\sigma_{12} = \sqrt{\frac{N_1 \sigma_1^2 + N_2 \sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

$$= \sqrt{\frac{500 \times 81 + 600 \times 100 + 500 \times 36 + 600 \times 25}{500 + 600}}$$

$$= \sqrt{\frac{40500 + 60000 + 18000 + 15000}{1100}}$$

$$= \sqrt{\frac{133500}{1100}} = \sqrt{121.36} = 11.01$$

Example 39. Given: sum of squares of items = 2430, $\overline{X} = 7, N = 12$, find the coefficient of variation.

Solution: Given : $\Sigma X^2 = 2430$, $\overline{X} = 7$, $N = 12$

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

$$\sigma = \sqrt{\frac{2430}{12} - (7)^2} = \sqrt{153.5} = 12.38$$

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100$$

$$= \frac{12.38}{7} \times 100 = 176.85\%$$

EXERCISE 6.9

1. Batsmen X and Y scored following runs in different innings they played in a test series. Which of the two is a better scorer? Who is more consistent batsman?

X:	12	115	6	73	7	19	119	36	84	29
Y:	47	12	76	42	4	51	37	48	13	0

[**Ans.** $\overline{X} = 50, \sigma_X = 41.83, C.V._X = 83.66; \overline{Y} = 33, \sigma_Y = 23.37, C.V_Y = 70.81$
Batsman X is a better scorer; Batsman Y is a consistent batsman]

2. The following is the record number of bricks laid each day for 10 days by two layers A and B. Calculate the coefficient in each case and discuss the relative consistency of the two brick-layers.

A:	700	675	725	625	650	700	650	700	600	650
B:	550	600	575	550	650	600	550	525	625	600

If each of the values in respect of worker A is decreased by 10 and each of the values for worker B is increased by 50, how will it affect the results obtained earlier?
[**Hint:** See Example 43] [**Ans.** $\overline{X}_A = 667.5, \sigma_A = 37.15$, C.V.$_A$ = 5.56%
$\overline{X}_B = 582.5, \sigma_B = 37.15$, C.V.$_B$ = 6.38%]

3. Goals scored by two teams A and B in a footfall session were as follows:

No. of goals scored:	0	1	2	3	4	5
No. of matches by A:	15	10	7	5	3	2
No. of matches by B:	20	10	5	4	2	1

Find out which team is more consistent.
[**Ans.** C.V. for A = 102.06%, C.V. for B = 124.6%, Team A is more consistent]

4. A factory produces two types of electric bulbs A and B. In an experiment relating to their life, the following results were obtained:

Length of life (in hrs.)	**No. of lamps (A)**	**No. of lamps (B)**
500—700	5	4
700—900	11	30
900—1100	26	12
1100—1300	10	8
1300—1500	8	6

Which type of electric lamp do you prefer? Give reasons.
[**Ans.** C.V.(A) = 21.6, C.V.(B) = 23.4
As C.V. of A is less, so lamp A is preferred]

5. For two firms A and B belonging to the same industry, the following data is given:

	Firm A	Firm B
No. of wage earners:	586	648
Average monthly wage (Rs.):	52.5	47.5
Standard deviation:	10	11

(*i*) Which firm A or B pays larger amount as weekly wages?

(*ii*) Which firm shows greater variability in the wage rate?

(*iii*) Find the mean and S.D. of all workers in the two factories taken together.

[Ans. (*i*) Firm B, (*ii*) In firm B, there is greater variability, (*iii*) $\bar{X}_{12} = 49.87, \sigma_{12} = 10.83$]

6. From the following data, find out Range, Quartile Deviation, Mean Deviation and Coefficient of Variation when mean of the distribution is 37.4.

X:	0—10	10—20	20—30	30—40	40—50	50—60	60—70
f:	2	7	9	11	?	8	6

[Ans. Missing Frequency = 7, R = 70, Q_1 = 23.809, Q_3 = 59.88, Q.D. = 13.995, S.D. = 17.04, C.V. = 45.56%]

7. If 20 is substracted from every observation in a data set, then the coefficient of variation of the resulting set is 20%. If 40 is added to every observation of the same data, then the coefficient of variation of the resulting set of data is 10%. Find the $\bar{X}$ and σ of the original set of data.

[Hint: See Similar Example 57, $20 = \dfrac{\sigma \times 100}{\bar{X} - 20}$, $10 = \dfrac{\sigma \times 100}{\bar{X} + 40}$] **[Ans.** $\bar{X}$ = 80, σ = 12]

8. A fund manager is considering investment in the equity shares of one of two companies. The criterion for selecting the company for investment is consistency of return on net worth. The following data have been collected:

Financial Year	Return on Net worth (%)	
	Modern Industries Ltd. (MIL)	Pioneer Industries Ltd. (PIL)
2001—2002	19	20
2000—2001	20	24
1999—2000	16	16
1998—1999	13	15
1997—1998	12	10

You are required to identify the company in which the fund manager should invest.
[Hint: See Example 58]

[Ans. For MIL : $\bar{X}$ = 16%, σ = 3.16%, C.V. = 19.76%
For PIL : $\bar{X}$ = 17%, σ = 4.73%, C.V. = 27.83%
MIL is more consistent and investment be made in MIL]

9. The coefficient of variation of wages of male workers and female workers are 55 per cent and 70 per cent respectively, while the standard deviations are 22.0 and 15.4 respectively. Calculate the overall average wages of all workers given that 80 per cent of the workers are male. [**Ans.** $\bar{X}_{12} = 36.4$]

10. The number of employees, wages per employee and variance of the wage per employee for two factories are given below:

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

(*i*) In which factory is there greater variation in the distribution of wages per employee?

(*ii*) Suppose in factory B, the wages of an employee were wrongly noted as Rs. 120 instead of Rs. 100. What would be the correct variance for factory B?

[**Ans.** (*i*) $C.V._A = 2.5$; $C.V._B = 4.71$ B is more variable (*ii*) $\sigma^2 = 5.96$]

(6) LORENZ CURVE

It is a graphical method of studying dispersion. Lorenz curve was given by famous statistician **Max O Lorenz**. Lorenz curve has great utility in the study of degree of inequality in the distribution of income and wealth between the countries. It is also useful for comparing the distribution of wages, profits, etc., over different business groups. Lorenz curve is a cumulative percentage curve in which the percentage of frequency (persons or workers) is combined with the percentage of other items such as income, profits, wages, etc.

Construction of a Lorenz Curve

Following steps are used while drawing a Lorenz Curve:

(*i*) The size of items (variable values) and frequencies are both cumulated. Taking grand total for each as 100, percentages are obtained for these various cumulative values.

(*ii*) Cumulative frequencies are plotted on X-axis while cumulative items are plotted on the Y-axis.

(*iii*) On both the axis, we start from 0 to 100 and both **X** and **Y** axis take the values from 0 to 100.

(*iv*) Draw a diagonal line **Y = X** joining the origin 0 (0.0) with the point P(100, 100). The line OP is called the **line of equal distribution**. Any point on this diagonal line shows the same per cent of X and Y.

(*v*) Plot the percentages of the cumulated values on the graph and a curve is obtained by joining different points. It is called **Lorenz curve**.

(*vi*) Closeness of the Lorenz curve to the line of equal distribution shows lesser variation in the distribution. Larger the gap between the line of equal distribution and the Lorenz curve, greater is the variation.

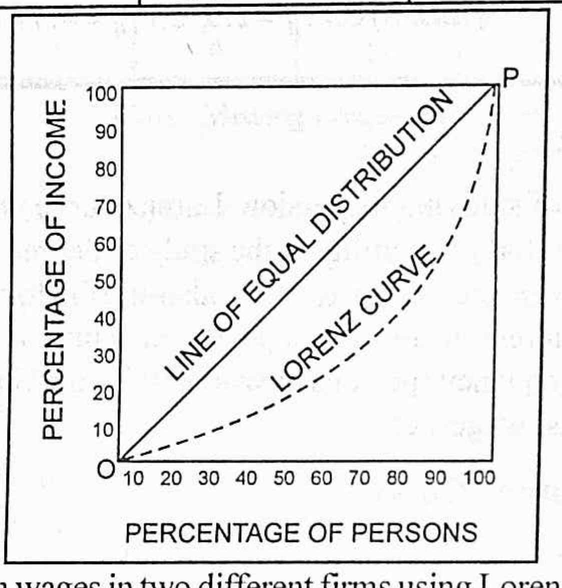
The following examples illustrate the procedure of drawing a Lorenz curve:

Example 40. Draw a Lorenz curve of the data given below:

Income:	100	200	400	500	800
No. of persons:	80	70	50	30	20

Solution:

Income	Cumulative Income	Cumulative percentage	No. of persons	Cumulative total	Cumulative percentage
100	100	$\frac{100}{2000}\times 100 = 5$	80	80	$\frac{80}{250}\times 100 = 3$
200	300	$\frac{300}{2000}\times 100 = 15$	70	150	$\frac{150}{250}\times 100 = 6$
400	700	$\frac{700}{2000}\times 100 = 35$	50	200	$\frac{200}{250}\times 100 = 8$
500	1,200	$\frac{1200}{2000}\times 100 = 60$	30	230	$\frac{230}{250}\times 100 = 92$
800	2,000	$\frac{2000}{2000}\times 100 = 100$	20	250	$\frac{250}{250}\times 100 = 10$



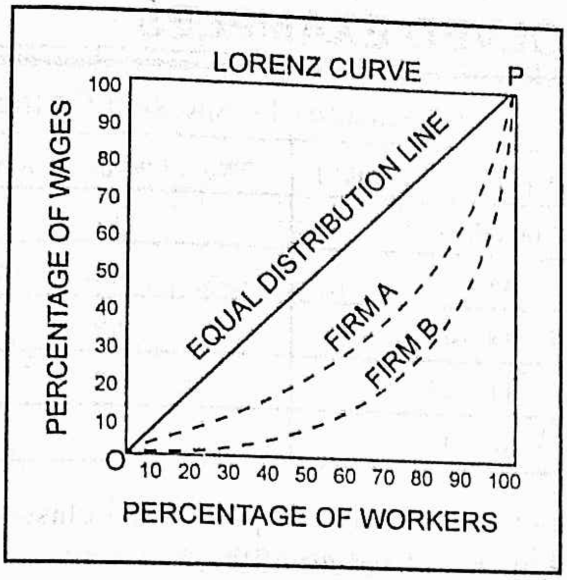
Example 41. Show inequality in wages in two different firms using Lorenz curve from following data:

Wages (Rs.):	50—70	70—90	90—110	110—130	130—150
No. of workers in firm A:	20	15	20	25	20
No. of workers in firm B:	150	100	90	110	50

Solution:

Wages (Rs.)	Mid-values (Rs.)	Cumulative Wages	Cumulative %	Firm A			Firm B		
				No. of workers	Cumulative total	Cumulative %	No. of workers	Cumulative total	Cumulative %
50—70	60	60	12	20	20	20	150	150	30
70—90	80	140	28	15	35	35	100	250	50
90—110	100	240	48	20	55	55	90	340	68
110—130	120	360	72	25	80	80	110	450	90
130—150	140	500	100	20	100	100	50	500	100

Note: The percentages are approximated to the nearest whole numbers.



It is obvious from the above figure that inequalities in the distribution of wages are more in Firm B than in Firm A.

EXERCISE 6.10

1. The following table shows number of firms in two different areas according to their annual profits. Present the data by way of Lorenz Curve.

Profit ('000 rupees):	6	25	60	84	105	150	170	400
Firms in Area A:	6	11	13	14	15	17	10	14
Firms in Area B:	2	38	52	28	38	26	12	4

2. The distribution of 9,400 Indian families according to income size is given below. Show inequality in the distribution of income by using Lorenz Curve.

Income:	0—1000	1000—5000	5000—10000	10000—20000	20000—40000
Families:	1,348	4,210	1,892	1,460	490

[**Hint:** Find out mid-value of class intervals]

3. Use Lorenz curve to compare the extant of inequalities in income distribution in two groups:

Monthly Income (Rs.):	1200—1400	1400—1600	1600—1800	1800—2000	2000—2200	2200—2400
No. of Persons in Gr.A:	800	960	1040	600	480	120
No. of Persons in Gr.B:	4800	6400	9600	3600	8000	4000

[**Ans.** Inequalities are more in Gr.A than Gr.B]

MISCELLANEOUS SOLVED EXAMPLES

Example 42. Calculate an appropriate measure of dispersion for the following data:

Wages per weak (in Rs.)	No. of wage earners
less than 35	14
35—37	62
38—40	99
41—43	18
over 43	7

Solution: The given distribution consists of open ended classes. One is less than 35 and the other is 'over 43'. The mid-values of these classes cannot be determined. Therefore, the appropriate measure of dispersion is Q.D. and coefficient of Q.D.

Calculation of Quartile Deviation

Wages (Rs.)	f	$c.f.$
Less 35	14	14
35—37	62	76
38—40	99	175
41—43	18	193
Over 43	7	200
	N = 200	

Location of Q_1

Size of Q_1 item $= \frac{N}{4} = \frac{200}{4}$, *i.e.*, 50th item which lies in 35—37 group which after adjustments becomes 34.5 or 37.5.

Now $$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i$$

$$= 34.5 + \frac{50-14}{62} \times 2$$

$$= 34.5 + \frac{36}{62} \times 2 = 34.5 + \frac{72}{62} = 35.66$$

$\therefore \quad Q_1 = 35.66$

Location of Q_3

Size of Q_3 item $= \frac{3N}{4} = \frac{3 \times 200}{4} = 150$th item which lies in 38—40 group which after adjustment becomes $37.5 - 40.5$.

Now, $$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i$$

$$= 37.5 + \frac{150-76}{99} \times 3 = 37.5 + \frac{74}{33} = 37.5 + 2.24 = 39.74$$

Calculation of Q.D. and its coefficient

$$\text{Q.D.} = \frac{Q_3 - Q_1}{2} = \frac{39.74 - 35.66}{2} = \frac{4.08}{2} = 2.04$$

$$\text{Coefficient of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{39.74 - 35.66}{39.74 + 35.66} = \frac{4.08}{75.4} = 0.54 \text{ approx.}$$

Example 43. The following is the record number of bricks laid each day for 10 days by two brick-layers A and B. Calculate the coefficient of variation in each case and discuss the relative consistency of the two brick-layers.

A:	700	675	725	625	650	700	650	700	600	650
B:	550	600	575	550	650	600	550	525	625	600

If the figures for A were in every case 10 more and those of B in every case 20 more than figure given above, how would the answer be affected?

Solution: (*i*) **Calculation for Mean and Standard Deviation**

Brick-layer A			Brick-layer B		
X	$dx' = \frac{X-700}{25}$	dx'^2	Y	$dy' = \frac{Y-625}{25}$	dy'^2
700	0	0	550	– 3	9
675	– 1	1	600	– 1	1
725	1	1	575	– 2	4
625	– 3	9	550	– 3	9
650	– 2	4	650	1	1
700 = A	0	0	600	– 1	1
650	– 2	4	550	– 3	9
700	0	0	525	– 4	16
600	– 4	16	625 = A	0	0
650	– 2	4	600	– 1	1
Total	**– 13**	**39**		**– 17**	**51**

Brick-layer A:

$$\overline{X} = A + \frac{\sum dx'}{N} \times i$$

$$= 700 - \frac{13}{10} \times 25 = 700 - 32.5 = 667.5 \text{ bricks per day}$$

$$\sigma_X = \sqrt{\frac{\sum dx'^2}{N} - \left(\frac{\sum dx'}{N}\right)^2} \times i$$

$$= \sqrt{\frac{39}{10} - \left(\frac{-13}{10}\right)^2} \times 25 = \sqrt{3.9 - 1.69} \times 25$$

$$= \sqrt{2.21} \times 25 = 1.486 \times 25 = 37.15$$

$$\text{C.V. (A)} = \frac{\sigma_X}{\overline{X}} \times 100 = \frac{37.15}{667.5} \times 100 = 5.56\%$$

Brick-layer B:

$$\overline{Y} = 625 - \frac{17}{10} \times 25 = 582.5 \text{ bricks per day}$$

$$\sigma_Y = \left[\sqrt{\frac{51}{10} - \left(\frac{-17}{10}\right)^2}\right] \times 25$$

$$= \sqrt{2.21} \times 25 = 37.15$$

$$\text{C.V. (B)} = \frac{37.15}{582.5} \times 100 = 6.38\%$$

As the coefficient of variation for brick-layer A is less than that of brick-layer B, brick-layer A is more consistent.

(*ii*) If the figures for A in every case were 10 more and that of B were 20 more, the arithmetic mean of A will increase by 10 and that of B by 20 but the standard deviations of both of them will remain unchanged.

[∵ S.D. is independent of the change of origin]

∴ A.M. of A will be 667.5 + 10 = 677.5 bricks per day

and A.M. of B will be 582.5 + 20 = 602.5 bricks per day

∴ Coefficient of variation of A will be $= \frac{37.15}{677.5} \times 100 = 5.48\%$

and coefficient of variation of B will be $= \frac{37.15}{602.5} \times 100 = 6.16\%$

After the change also brick-layer A remains more consistent than brick-layer B.

Example 44. Calculate arithmetic mean, median, mode and standard deviation for the following series:

Daily wages (Rs):	0—34.5	0—44.5	0—54.5	0—64.5	0—74.5	0—84.5
No. of workers:	4	24	62	86	96	100

Solution: The given data is in cumulative frequency form. It should first be converted into simple frequency data.

Calculation of $\overline{X}$, M, Z and σ

Daily wages	f	M.V. (m)	d	$d' = \frac{d}{10}$	fd'	fd'^2	$c.f.$
24.5—34.5	4	29.5	−20	−2	−8	16	4
34.5—44.5	20	39.5	−10	−1	−20	20	24
44.5—54.5	38	49.5 = A	0	0	0	0	62
54.5—64.5	24	59.5	+10	+1	+24	24	86
64.5—74.5	10	69.5	+20	+2	+20	40	96
74.5—84.5	4	79.5	+30	+3	+12	36	100
	$N = 100$				$\Sigma fd' = 28$	$\Sigma fd'^2 = 136$	

Arithmetic Mean

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 49.5 + \frac{28}{100} \times 10 = 52.3$$

Median: Size of median item $= \frac{N}{2} = \frac{100}{2} = 50$th item which lies in the class 44.5—54.5.

Now,
$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$M = 44.5 + \frac{50 - 24}{38} \times 10 = 44.5 + \frac{26}{38} \times 10 = 44.5 + 6.84 = 51.34$$

Mode: By inspection, mode lies in the class 44.5 – 54.5

$$Z = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i$$

$l_1 = 44.5,\ \Delta_1 = 38 - 20 = 18,\ \Delta_2 = 38 - 24 = 14,\ i = 10$

$$\therefore \quad Z = 44.5 + \frac{18}{18 + 14} \times 10 = 44.5 + \frac{18}{32} \times 10$$

$$= 44.5 + \frac{180}{32} = 44.5 + 5.625 = 50.125$$

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

$$= \sqrt{\frac{136}{100} - \left(\frac{28}{100}\right)^2} \times 10$$

$$= \sqrt{1.36 - 0.0784} \times 10 = \sqrt{1.2816} \times 10 = 11.32$$

Example 45. For two groups of observations the following results were available:

Group I	Group II
$\Sigma(X-5)=8$	$\Sigma(X-8)=-10$
$\Sigma(X-5)^2=40$	$\Sigma(X-8)^2=70$
$N_1=20$	$N_2=25$

Find the mean and standard deviation of both the groups taken together.

Solution: **Group I:**

$$\text{Let } \Sigma d_1=\Sigma(X-5)=8$$

$$\Sigma d_1^2=\Sigma(X-5)^2=40$$

$$\overline{X}_1=A+\frac{\Sigma d_1}{N}=5+\frac{8}{20}=5.40$$

$$\sigma_1=\sqrt{\frac{\Sigma d_1^2}{N}-\left(\frac{\Sigma d_1}{N}\right)^2}=\sqrt{\frac{40}{20}-\left(\frac{8}{20}\right)^2}=1.36$$

Group II:

$$\text{Let } \Sigma d_2=\Sigma(X-8)=-10$$

$$\Sigma d_2^2=\Sigma(X-8)^2=70$$

$$\overline{X}_2=A+\frac{\Sigma d_2}{N}=8+\frac{(-10)}{25}=7.6$$

$$\sigma_2=\sqrt{\frac{\Sigma d_2^2}{N}-\left(\frac{\Sigma d_2}{N}\right)^2}=\sqrt{\frac{70}{25}-\left(\frac{-10}{25}\right)^2}=1.62$$

$$\text{Combined Mean}(\overline{X}_{12})=\frac{N_1\overline{X}_1+N_2\overline{X}_2}{N_1+N_2}=\frac{20\times5.40+25\times7.6}{20+25}=6.62$$

$$\text{Combined S.D. }(\sigma_{12})=\sqrt{\frac{N_1\sigma_1^2+N_2\sigma_2^2+N_1d_1^2+N_2d_2^2}{N_1+N_2}}$$

$$d_1=\overline{X}_1-\overline{X}_{12}=5.40-6.62=-1.22$$

$$d_2=\overline{X}_2-\overline{X}_{12}=7.6-6.62=+0.98$$

$$\sigma_{12}=\sqrt{\frac{20(1.36)^2+25(1.62)^2+20(-1.22)^2+25(0.98)^2}{20+25}}$$

$$=\sqrt{3.476}=1.864$$

Example 46. "After settlement the average weekly wage in a factory had increased from Rs. to Rs. 12,000 and the standard deviation had increased from Rs. 100 to Rs. 150 settlement the wage has become higher and more uniform." Do you agree?

Solution: C.V. before settlement $= \frac{100}{8000} \times 100 = 1.25\%$

C.V. after settlement $= \frac{150}{12000} \times 100 = 1.25\%$

Since, there is no change in C.V., there is no improvement in uniformity.

Example 47. Construct a continuous frequency distribution with class interval of 20 for the following table showing weight (in grams) of 50 applies:

110	103	89	75	98	121	110	108	93	128
185	123	113	92	86	70	126	78	139	120
129	119	105	120	100	116	85	99	114	185
205	111	141	136	123	90	115	128	160	78
90	107	81	137	125	184	104	100	87	115

Also calculate the coefficient of variation of this distribution.

Solution: The lowest value is 70 and highest is 205. We have to take a class interval of 20. The various classes will be 70—90, 90—110, and so on up to 190—210.

Weight (in grams)	Tally Bars	Frequency
70—90	𝍸 𝍸 |	11
90—110	𝍸 𝍸 |	11
110—130	𝍸 𝍸 𝍸 ||||	19
130—170	||||	4
150—170	|	1
170—190	|||	3
190—210	|	1
		$N = 50$

Calculation of Coefficient of Variation

Weight	f	M.V.	d	d'	fd'	fd'^2
70—90	11	80	−60	−3	−33	99
90—110	11	100	−40	−2	−22	44
110—130	19	120	−20	−1	−19	19
130—170	4	140 = A	0	0	0	0
150—170	1	160	+20	+1	1	1
170—190	3	180	+40	+2	6	12
190—210	1	200	+60	+3	3	9
	$N = 50$				$\Sigma fd' = -64$	$\Sigma fd'^2 = 184$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 140 - \frac{64}{50} \times 20 = 114.4$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{184}{50} - \left(\frac{-64}{50}\right)^2} \times 20$$

$$= \sqrt{368 - 1.6384} \times 20 = 28.5769$$

$$\text{Coefficient of variation} = \frac{\sigma}{\overline{X}} \times 100 = \frac{28.5769 \times 100}{114.4} = 24.97\%$$

Example 48. The coefficient of variation of a series is 58%. The standard deviation is 21.2. What is the arithmetic mean?

Solution:

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100$$

$$\Rightarrow \quad \overline{X} = \frac{\sigma}{C.V.} \times 100$$

$$\text{Mean } (\overline{X}) = \frac{21.2 \times 100}{58} = 36.6$$

Example 49. During nine days in a festival the highest sale of a shop was on Sunday and was Rs. 90 more than the average sale for other days. If the standard deviation of the sale during the festival is 33.33, find the standard deviation leaving that the highest sale.

Solution:

$$\sigma = \sqrt{\frac{\Sigma(X - \overline{X})^2}{N}} \quad \text{(Formula of S.D.)}$$

$$33.33 = \sqrt{\frac{\Sigma(X - \overline{X})^2}{9}}$$

$$\Rightarrow \quad \Sigma x^2 = \frac{100}{3} \times \frac{100}{3} \times 9 = 10{,}000 \quad [\because x = X - \overline{X}]$$

$$\text{Sunday value} = \overline{X} + 90$$

$$x = X - \overline{X}$$

$$= (\overline{X} + 90) - \overline{X} = 90$$

Sum of the square of deviations excluding Sunday.

$$= 10000 - (90)^2 = 1900$$

$$\sigma \text{ for 8 days} = \sqrt{\frac{\Sigma(X - \overline{X})^2}{8}}$$

$$= \sqrt{\frac{1900}{8}} = \sqrt{237.5} = 15.4$$

Example 50: If the mean and standard deviation of 75 observations is 40 and 8 respectively, the new mean and standard deviation if

(*i*) Each observation is multiplied by 5.

(*ii*) 7 is added to each observation.

Solution: (*i*) New $\overline{X} = 200$, New $\sigma = 40$

The reason is that the change of scale affects the value of both $\overline{X}$ and σ.

(*ii*) New $\overline{X} = 47$, New $\sigma = 8$

The reason is that the mean is affected by change of origin but S.D. is not affected by change of origin.

Example 51. 5 observations of a series are 4, 6, 8, 12 and 15. Their mean and standard deviation are 9 and 4 respectively. Make such alternations in the terms of the series that new S.D. is 20 and new mean is 50.

Solution: Given $\overline{X} = 9, \sigma = 4$

The series with new S.D. is obtained when each observations is multiplied by 5. This operation will also increase the value of mean 5 times. In the given example, if each observation is multiplied by 5, the mean becomes 45 and S.D. becomes 20. The reason is that change of scale affects both mean and standard deviation.

Now, if 5 is added to each observation the mean becomes 50 while S.D. remains 20. The reason is that the change of origin affects only $\overline{X}$ and not σ. The transformation series corresponds to this can be written as $U = 5X + 5$. Thus, the changed observations will be: 25, 35, 45, 65 and 80.

Example 52. A collar manufacturer is considering the production of a new style of collar to attract youngmen. The following statistics of neck circumference are available based on measurements of a typical group of college students. Compute the SD and use the criterion $(\overline{X} \pm 3\sigma)$, to determine the largest and smallest sizes of collars, he should make in order to meet the needs of practically all his customers, bearing in mind, that collars are worn, on average $\frac{3}{4}$ inch larger than neck size.

Mid-points:	12.5	13.0	13.5	14.0	14.5	15.0	15.5	16.0	16.5
f:	4	19	30	63	66	29	18	1	1

Solution: Here, $i = 0.5$

Let, $A = 14.5 \quad \therefore d' = \left(\dfrac{X - A}{i}\right)$

Mid-points (X)	f	d'	d'^2	fd'	fd'^2
12.5	4	–4	16	–16	64
13.0	19	–3	9	–57	171
13.5	30	–2	4	–60	120
14.0	63	–1	1	–63	63
14.5	66	0	0	0	0
15.0	29	+1	1	29	29
15.5	18	+2	4	36	72
16.0	1	+3	9	3	9
16.5	1	+4	16	4	16
	$N = 231$			$\Sigma fd' = -124$	$\Sigma fd'^2 = 544$

Now, $\overline{X} = A + \frac{\Sigma fd'}{N} \times i$

$= 14.5 + \frac{-124}{231} \times 0.5 = 14.5 - 0.268 = 14.232$ inches

$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{544}{231} - \left(\frac{-124}{231}\right)^2} \times 0.5$

$= \sqrt{2.355 - 0.288} \times 0.5 = \sqrt{2.067} \times 0.5$

$= 1.437 \times 0.5 = 0.719$ inches.

Using the criterion $\overline{X} \pm 3\sigma$

Largest and smallest neck size

$= \overline{X} \pm 3\sigma = 14.232 \pm 3(0.719)$

$= 14.232 \pm 2.157 = 12.075$ and 16.389

Since, the collars are worn on an average $\frac{3}{4}$ inch longer than the neck size, we should add 0.75 to these limits. Thus, the smallest and largest sizes of the collar should be $(12.075 + 0.75)$ and $(16.389 + 0.75) = 12.825$ and 17.139

Thus, the smallest size of the collar should be 12.825 inches long and largest 17.139 inches long.

Example 53. The mean age and standard deviation of a group of 200 persons (grouped in intervals 0—5, 5—10, ..., etc.) were found to be 40 and 15. Later on it was discovered that the age 43 was misread as 53. Find the correct mean and standard deviation.

Solution: $N = 200$, $\overline{X} = 40$ and $\sigma = 15$

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

Incorrected $\Sigma fm = N\overline{X} = 200 \times 40 = 8000$

Corrected age is 43 which falls in the group 40—45, the mid-value of which is 42.5.

Incorrected age is 53 which falls in the group 50—55, the mid-value of which is 52.5.

$$\text{Corrected } \overline{X} = \frac{8000 + 42.5 - 52.5}{200} = \frac{7990}{200} = 39.95$$

$$\text{Incorrected } \Sigma fm^2 = N(\sigma^2 + \overline{X}^2) = 200\,(15^2 + 40^2) = 3{,}65{,}000$$

$$\text{Corrected } \Sigma fm^2 = 3{,}65{,}000 - (52.5)^2 + (42.5)^2$$

$$= 3{,}65{,}000 - 2756.25 + 1806.25 = 3{,}64{,}050$$

$$\text{Corrected } \sigma = \sqrt{\frac{3{,}64{,}050}{200} - (39.95)^2}$$

$$= \sqrt{1820.25 - 1596.0025} = 14.97$$

Example 54. The monthly wages (in Rs.) of 100 workers are distributed as follows:

Wages (Rs.):	0—100	100—200	200—300	300—400	400—500	500—600
No. of workers:	12	x	27	y	17	6

If model wage is Rs. 256.25, find the missing frequencies and hence find % variation in the distribution.

Solution: Let the missing frequencies be x and y

Wages (Rs.)	f	$c.f.$
0—100	12	12
100—200	x f_0	$12 + x$
200—300	27 f_1	$39 + x$
300—400	y f_2	$39 + x + y$
400—500	17	$56 + x + y$
500—600	6	$62 + x + y$
	N = 100	

$$62 + x + y = 100$$

$$\Rightarrow \quad x + y = 100 - 62 = 38$$

As $Z = 256.25$, Mode lies in 200–300

Applying the formula,

$$Z = l_1 + \frac{\Delta_1}{\Delta_1 + \Delta_2} \times i \quad \text{where, } \Delta_1 = |f_1 - f_0|;\ \Delta_2 = |f_2 - f_2|$$

$$256.25 = 200 + \frac{27 - x}{(27 - x) + (27 - y)} \times 100$$

$$256.25 - 200 = \frac{2700 - 100x}{27 - x + 27 - (38 - x)} \qquad [\because\ y = 38 - x]$$

$$\Rightarrow \quad 56.25 = \frac{2700 - 100x}{27 - x + 27 - 38 + x}$$

$$\Rightarrow \quad 56.25 = \frac{2700 - 100x}{16}$$

$$\Rightarrow \quad 56.25 \times 16 = 2700 - 100x$$

$$\Rightarrow \quad 900 - 2700 = -100x$$

$$\Rightarrow \quad -1800 = -100x$$

$$\therefore \quad x = \frac{1800}{100} = 18$$

Now, $\quad x + y = 38 \Rightarrow y = 38 - x = 38 - 18 = 20$

$$\therefore \quad x = 18 \text{ and } y = 20$$

Now, in order to find out % variation in the distribution, we have to find out the co-efficient of variation.

Wages (Rs.)	f	M.V. (m)	d	d'	fd'	fd'^2
0—100	12	50	–200	–2	–24	48
100—200	18	150	–100	–1	–18	18
200—300	27	250 = A	0	0	0	0
300—400	20	350	100	1	20	20
400—500	17	450	200	2	34	68
500—600	6	550	300	3	18	54
	$N = 100$				$\Sigma fd' = 30$	$\Sigma fd'^2 = 208$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 250 + \frac{30}{100} \times 100 = 280$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{208}{100} - \left(\frac{30}{100}\right)^2} \times 100$$

$$= \sqrt{2.08 - 0.09} \times 100 = \sqrt{1.99} \times 100$$

$$= 1.4107 \times 100 = 141.07$$

$$C.V. = \frac{\sigma}{\overline{X}} \times 100 = \frac{141.07}{280} \times 100 = 50.38\%$$

Example 55. The following table shows the marks obtained by three students A, B and C in an examination:

Student ↓ Subject →	Maximum marks in each subject				
	800 S_1	700 S_2	900 S_3	600 S_4	1000 S_5
A	560	553	549	540	500
B	480	420	540	360	600
C	424	427	423	426	420

Determine which student has shown (i) most consistent performance and (ii) most inconsistent performance.

Solution: The given data has been shown below:

Student ↓ Subject →	Percentage marks					Average % marks	Standard deviation	C.V.
	S_1	S_2	S_3	S_4	S_5			
A	70	79	61	90	50	70	13.87	19.81
B	60	60	60	60	60	60	0	0
C	53	61	47	71	42	54.8	10.32	18.83

$$\text{A's average marks} = \frac{70+79+61+90+50}{5} = \frac{350}{5} = 70$$

$$\text{B's average marks} = \frac{60+60+60+60+60}{5} = \frac{300}{5} = 60$$

$$\text{C's average marks} = \frac{53+61+47+71+42}{5} = \frac{274}{5} = 54.8$$

A's standard deviation (σ_A)

$$= \sqrt{\frac{0+(9)^2+(-9)^2+(20)^2+(-20)^2}{5}} = \sqrt{\frac{962}{5}} = \sqrt{192.4} = 13.87$$

$$\text{B's standard deviation } (\sigma_B) = \sqrt{\frac{0}{5}} = 0$$

$$\text{C's standard deviation } (\sigma_C) = \sqrt{\frac{(1.8)^2+(6.2)^2+(-7.8)^2+(16.3)^2+(-12.8)^2}{5}}$$

$$= \sqrt{\frac{530.05}{5}} = \sqrt{106.41} = 10.32$$

$$\text{A's coefficient of variation (C.V.)} = \frac{\sigma}{\bar{X}} \times 100 = \frac{13.87}{70} \times 100 = 19.81\%$$

$$\text{B's coefficient of variation (C.V.)} = \frac{0}{60} \times 100 = 0\%$$

$$\text{C's coefficient of variation (C.V.)} = \frac{10.32}{54.8} \times 100 = 18.83\%$$

Hence, it is clear from the above data that

(*i*) B is most consistent as in his case, the coefficient of variation is the least.

(*ii*) A is most inconsistent as in his case, the coefficient of variation is the greatest.

Example 56. The salaries paid to the managers of a company had a mean of Rs. 20,000 with a standard deviation of Rs. 3,000. What will be the mean and standard deviation if all the salaries are increased by (*i*) 10%, (*ii*) 10% of the existing mean?

Which policy would you recommend if the management does not want to have increased disparities in wages?

Solution: Increasing all the salaries by 10% ⇒ multiplying each salary each by 1.1. Hence, the mean is also multiplied by 1.1. Since, S.D. defends on change of scale, S.D. is also multiplied by 1.1.

∴ When all salaries are increased by 10%, the S.D. also increases by 10%.

If each salary is increased by 10%, the mean is also increased by 10%.

Increasing all the salaries by 10% of the existing mean ⇒ Adding a constant amount to each salary.

Since S.D. is independent of the change of origin, it will remain unchanged if each salary is increased by 10% of the mean. But mean will increase by 10% of the original mean.

If the management do not want to have increased disparities, it showed increase in the salary of the workers by 10% of the existing mean.

Example 57. If 10 is subtracted from every item in a data set then the coefficient of variation of the resulting set of data is 20%. If 20 is added to every item of the same data set then the coefficient of variation of the resulting set of data is 10%.

You are required to find out the coefficient of variation of the original set of data.

Solution: Let the average and standard deviation of the original data set be $\overline{X}$ and s.

Average of all items '$X - 10$' $= \dfrac{\Sigma(X-10)}{N} = \dfrac{\Sigma X}{N} = \dfrac{N10}{N} = \overline{X} - 10$

Standard deviation of all items '$N - 10$' $= s$

(This is because the value of standard deviation remains the same if each observation in a series is increased or decreased by the same quantity)

Given : $\dfrac{s}{\overline{X}-10} \times 100 = 20$

$$\Rightarrow \quad 100s = 20\overline{X} - 200$$

$$\Rightarrow \quad 20\overline{X} - 100s = 200$$

Average of all items '$X + 20$' $= \dfrac{\Sigma(X+20)}{N} = \dfrac{\Sigma X}{N} = \dfrac{N20}{N} = \overline{X} + 20$

Standard deviation of all items '$X + 20$' $= s$

(This is because the value of standard deviation remains the same if each observation in a series is increased or decreased by the same quantity)

Given : $\dfrac{s}{\overline{X}+20} \times 100 = 10$

$$\Rightarrow \quad 100s = 10\overline{X} + 200$$

$$\Rightarrow \quad 10\overline{X} - 100s = 200$$

Subtracting equation (*ii*) from equation (*i*), we get

$$(20\overline{X} - 100s) - (10\overline{X} - 100s) = 200 - (-200)$$

$$\Rightarrow \quad 10\overline{X} = 400$$

$$\Rightarrow \quad \overline{X} = \frac{400}{10} = 40$$

Putting the value of $\overline{X}$ in equation (*ii*), we get

$$10(40) - 100s = -200$$

$$\Rightarrow \quad 400 - 100s = -200$$

$$\Rightarrow \quad 100s = 600$$

$$\Rightarrow \quad s = \frac{600}{100} = 6$$

$\therefore$ Coefficient of variation of the original data set:

$$= \frac{s}{\overline{X}} \times 100 = \frac{6}{40} \times 100 = 15\%$$

Example 58. A fund manager is considering investment in the equity shares of one of two companies. The criterion for selecting the company for investment is consistency of return on net worth. The following data have been collected:

Financial Year	Return on Net worth (%)	
	Modern Industries Ltd. (MIL)	Pioneer Industries Ltd. (PIL)
2001—2002	19	20
2000—2001	20	24
1999—2000	16	16
1998—1999	13	15
1997—1998	12	10

You are required to identify the company in which the fund manager should invest.

Solution: **Modern Industries Ltd. (MIL):**

$$\text{Mean return on equity } (\overline{X}) = \frac{\Sigma X}{N} = \frac{19+20+16+13+12}{5} = \frac{80}{5} = 16\%$$

$$\text{Standard deviation (S.D.)} = \left[\frac{\Sigma(X-\overline{X})^2}{N}\right]^{\frac{1}{2}}$$

$$= \left[\frac{(19-16)^2+(20-16)^2+(16-16)^2+(13-16)^2+(12-16)^2}{5}\right]^{\frac{1}{2}}$$

$= 3.16\%$ (approx.)

$$\text{Coefficient of variation} = \frac{\text{S.D.}}{\overline{\text{X}}}\times 100 = \frac{3.16}{16}\times 100 = 19.76\% \text{ (approx.)}$$

Pioneer Industries Ltd. (PIL):

$$\text{Mean return on equity } (\overline{X}) = \frac{\Sigma X}{N} = \frac{20+24+16+15+10}{5} = \frac{85}{5} = 17\%$$

$$\text{Standard deviation (S.D.)} = \left[\frac{\Sigma(X-\overline{X})^2}{N}\right]^{\frac{1}{2}}$$

$$= \left[\frac{(20-17)^2+(24-17)^2+(16-17)^2+(15-17)^2+(10-17)^2}{5}\right]^{\frac{1}{2}}$$

$= 4.73\%$ (approx.)

$$\text{Coefficient of variation} = \frac{\text{S.D.}}{\overline{X}}\times 100 = \frac{4.73}{17}\times 100 = 27.83\% \text{ (approx.)}$$

Since, the coefficient of variation for MIL is less than coefficient of variation for PIL, it can be inferred that profitability of MIL is more consistent than PIL. Hence, investment should be made in MIL.

IMPORTANT FORMULAE

1. Range

$$\text{Range} = L - S$$

$$\text{Coeff. of Range} = \frac{L - S}{L + S}$$

2. Quartile Deviation

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

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

3. Mean Deviation

For Individual Series:

$$\text{M.D.} = \frac{\Sigma\,|D|}{N}$$

For Discrete and Continuous Series:

$$\text{M.D.} = \frac{\Sigma f\,|D|}{N}$$

$$\text{Coeff. of M.D.} = \frac{\text{M.D.}}{\text{Average}}$$

4. Standard Deviation

For Individual Series:

(*i*) $\sigma = \sqrt{\dfrac{\Sigma (X - \bar{X})^2}{N}}$: Actual Mean Method

(*ii*) $\sigma = \sqrt{\dfrac{\Sigma d^2}{N} - \left(\dfrac{\Sigma d}{N}\right)^2}$: Assumed Mean Method

(*iii*) $\sigma = \sqrt{\dfrac{\Sigma d'^2}{N} - \left(\dfrac{\Sigma d'}{N}\right)^2} \times i$: Step Deviation Method

For Discrete / Continuous Series:

(*i*) $\sigma = \sqrt{\dfrac{\Sigma f\,(X - \bar{X})^2}{N}}$: Actual Mean Method

(*ii*) $\sigma = \sqrt{\dfrac{\Sigma fd^2}{N} - \left(\dfrac{\Sigma fd}{N}\right)^2}$: Assumed Mean Method

(*iii*) $\sigma = \sqrt{\dfrac{\Sigma fd'^2}{N} - \left(\dfrac{\Sigma fd'}{N}\right)^2} \times i$: Step Deviation Method

5. Combined Standard Deviation

$$\sigma_{12} = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1 d_1^2 + N_2 d_2^2}{N_1 + N_2}}$$

Where, $d_1 = \overline{X}_1 - \overline{X}_{12}$ and $d_2 = \overline{X}_2 - \overline{X}_{12}$

6. Coefficient of Variation

$$\text{C.V.} = \frac{\sigma}{\overline{X}} \times 100$$

7. Variance

$$\text{Variance} = (S.D.)^2$$

or $\sigma = \sqrt{\text{variance}}$

QUESTIONS

1. What is meant by dispersion? What purpose does a measure of dispersion serve?
2. What are the various measures of dispersion. Explain the relative merits and demerits of each.
3. (*i*) What are the properties of a good measure of variation?
   (*ii*) Why is standard deviation considered a better measure of dispersion.
4. What is coefficient of variation? What purpose does it serve? Also distinguish between variance and coefficient of variation.
5. Define range, interquartile range, quartile deviation, mean deviation and standard deviation. Describe their merits and demerits.
6. Define dispersion. Discuss the merits and demerits of different measures of dispersion.
7. What do you understand by standard deviation? Explain its important properties.
8. Explain the method of measuring inequalities of income by using Lorenz curve.
9. What do you understand by Lornez curve? Discuss the usefulness of Lornez curve.
10. Why is S.D. (σ) the most widely used measure of dispersion? Explain.
11. If a constant is subtracted from each score in a series, what will be its effect on $\overline{X}$ and σ?

7 Measures of Skewness

◘ INTRODUCTION

In the preceeding two chapters, we have discussed the measures of central tendency and dispersion of frequency distributions for their summarisation and comparison with each other. These measures, however, do not adequately describe a frequency distribution in the sense that there could be two or more distributions with the same mean and standard deviation but still different from each other with regard to shape or pattern of distribution. This implies that there is need to develop some measures to further describe the distribution. These measures are known as measures of skewness.

◘ MEANING OF SKEWNESS

The term skewness means lack of symmetry in a frequency distribution. Skewness denotes the degree of departure of a distribution from symmetry and reveals the direction of scatterness of the items. It gives us an idea about the shape of the frequency curve. When a distribution is not symmetrical, it is called a skewed distribution. Skewness tells us about the asymmetry of the frequency distribution.

◘ DEFINITION OF SKEWNESS

Some important definitions of skewness are given below:

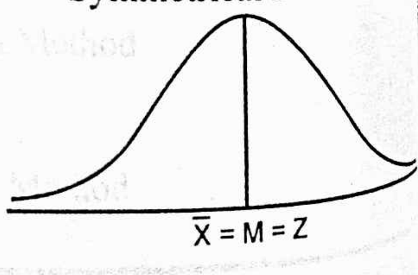
1. Skewness is the degree of asymmetry or departure from symmetry of a distribution.
   —M.R. Speigal
2. When a series is not symmetrical, it is said to be asymmetrical or skewed.
   —Croxten and Cowden
3. By skewness of a frequency distribution, we mean degree of its departure from symmetry.
   —Simpson and Kafka

◘ SKEWNESS AND FREQUENCY DISTRIBUTION

The concept of skewness will be made more clear from the following diagrams showing a symmetrical distribution, a positively skewed distribution and a negatively skewed distribution.

(1) Symmetrical Distribution: In a symmetrical distribution or symmetrical curve, skewness is not present. The values of mean, median and mode coincide, *i.e.*, $\overline{X} = M = Z$. The spread of the frequencies is the same on both sides of the central point of the curve.

Symmetrical Distribution



(2) Skewed Distribution: A distribution which is not symmetrical is called skewed distribution or asymmetrical distribution. A skewed distribution may be either positively skewed or negatively skewed.

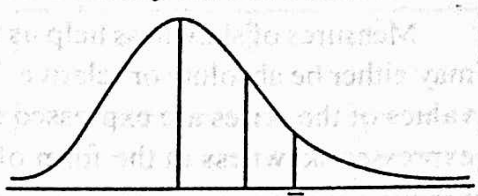
(a) Postively Skewed Distribution: If the longer tail of the frequency curve of distribution lies to the right of the central point, it is called a positively skewed distribution.

In the positively skewed distribution, the value of the mean will be greater than median and median be greater than mode, *i.e.*, $\bar{X} > M > Z$.

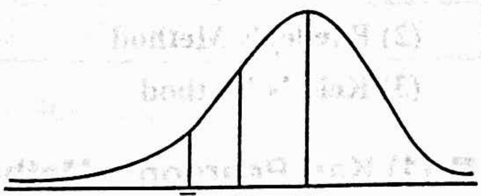
(b) Negatively Skewed Distribution: If the longer tail of the frequency curve of the distribution lies to the left of the central point, it is called a negatively skewed distribution.

In the negatively skewed distribution, the value of the mean will be less then median and median be less than mode, *i.e.*, $\bar{X} < M < Z$.

Postively Skewed Distribution



Negatively Skewed Distribution



DIFFERENCE BETWEEN DISPERSION AND SKEWNESS

The main points of difference between dispersion and skewness are given as under:

(1) Dispersion is concerned with measuring the amount of variation in a series rather than with its direction. Skewness is concerned with direction of variation or the departure from symmetry.

(2) Dispersion tells us about the composition of the series whereas skewness tells us about the shape of the series.

(3) Measures of dispersion are based on averages of the first order such as $\bar{X}, M, Z$, etc., whereas measures of skewness are based on averages of first and second order such as $\bar{X}, M, Z, \sigma$, etc.

TESTS OF SKEWNESS

In order to find out whether a distribution is skewed or not, the following tests may be applied:

(1) Relationship between Averages: If in a distribution, the values of mean, median and mode are equal, *i.e.*, $\bar{X} = M = Z$, then skewness is absent in it. On the other hand, if the values of mean, median and mode are not identical, *i.e.*, $\bar{X} \neq M \neq Z$, then skewness is found present in the distribution.

(2) Distance of Quartiles from the Median: If in a distribution, the quartiles (Q_1 and Q_3) are equidistant from the median, *i.e.*, $Q_3 - M = M - Q_1$, then skewness is absent and if $Q_3 - M \neq M - Q_1$, then skewness is present in the distribution.

(3) Graph of the Data: When the data plotted on the graph paper gives us a bell shaped curve, skewness is absent. On the other hand, when the data plotted on the graph do not give the normal bell shaped curve, *i.e.*, the two values of the curve are not equal, then skewness is present in the distribution.

MEASURES OF SKEWNESS

Measures of skewness help us to find out the direction and extent of asymmetry in a series. They may either be absolute or relative. The measures which expresses skewness in the units in which the values of the series are expressed are called **absolute measures of skewness.** The measures which expresses skewness in the form of ratios or percentage are called **relative measures of skewness.** Relative measures of skewness, also called **coefficient of skewness** are useful to compare the skewness of two or more series.

There are three important methods of measuring skewness, namely

(1) Karl Pearson's Method

(2) Bowley's Method

(3) Kelly's Method

(1) Karl Pearson's Method

Karl Pearson's method is based on arithmetic mean ($\bar{X}$), mode (Z), median (M) and standard deviation (σ). Karl Pearson has given the following formulae for measuring skewness:

Absolute Measure of Skewness	Coefficient of Skewness
$S_K = \bar{X} - Z$	Coefficient of $S_K = \frac{\bar{X} - Z}{\sigma}$
When mode (Z) is ill defined, then $S_K = 3(\bar{X} - M)$	When mode (Z) is ill defined, then Coefficient of $S_K = \frac{3(\bar{X} - M)}{\sigma}$

The value of Karl Pearson's coefficient of skewness usually lies between ± 1. In case mode is ill defined, the value of coefficient of skewness lies between ± 3.

Steps for Calculation

(1) Calculate mean ($\bar{X}$) of the distribution.

(2) Calculate mode (Z) of the distribution.

(3) Calculate median (M) of the distribution.

(4) Calculate standard deviation (σ).

(5) Put these values in the formulae.

The following examples illustrate the procedure of calculating Pearson's coefficient of skewness.

Calculation of Coefficient of Skewness—Discrete Series

Example 1. From the following data, find out Karl Pearson's Coefficient of Skewness:

Height (in inches)	No. of persons
58	10
59	18
60	30
61	42
62	35
63	28

Solution:

Calculation of Karl Pearson's Coefficient of Skewness

Height (X)	f	A = 60 (X – 60) d	fd	fd^2
58	10	–2	–20	40
59	18	–1	–18	18
60 A	30	0	0	0
61	42	+1	+42	42
62	35	+2	+70	140
63	28	+3	+84	252
	N = 163		$\Sigma fd = 158$	$\Sigma fd^2 = 492$

$$\overline{X} = A + \frac{\Sigma fd}{N} = 60 + \frac{158}{163} = 60 + 0.969 = 60.969$$

$$\sigma = \sqrt{\frac{fd^2}{N} - \left(\frac{\Sigma fd}{N}\right)^2} = \sqrt{\frac{492}{163} - \left(\frac{158}{163}\right)^2}$$

$$= \sqrt{3.0184 - 0.9395} = \sqrt{2.0789} = 1.4418$$

By inspection, mode is 61 (as its frequency is maximum)

Thus, $\overline{X} = 60.969$, $\sigma = 1.4418$, $Z = 61$

$$\therefore \text{ Coefficient of Skewness} = \frac{\overline{X} - Z}{\sigma} = \frac{60.969 - 61}{1.4418}$$

$$= \frac{-0.031}{1.4418} = -0.0215$$

Calculation of Pearson's Coefficient of Skewness—Continuous Series

Example 2. Calculate Karl Pearson's Coefficient of Skewness from the following data:

Variable:	0 – 5	5 – 10	10 – 15	15 – 20	20 – 25	25 – 30	30 – 35	35 – 40
Frequency:	2	5	7	13	21	16	8	3

Solution:

Calculation of Karl Pearson's Coefficient of Skewness

Variable	f	M.V. (m)	d	$d' = \frac{d}{5}$	fd'	fd'^2
0 – 5	2	2.5	–20	–4	–8	32
5 – 10	5	7.5	–15	–3	–15	45
10 – 15	7	12.5	–10	–2	–14	28
15 – 20	13	17.5	–5	–1	–13	13
20 – 25	21	22.5 = A	0	0	0	0
25 – 30	16	27.5	+5	+1	16	16
30 – 35	8	32.5	+10	+2	16	32
35 – 40	3	37.5	+15	+3	9	27
	N = 75				$\Sigma fd' = -9$	$\Sigma fd'^2 = 193$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 22.5 + \frac{(-9)}{75} \times 5 = 22.5 - 0.6 = 21.9$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{fd'}{N}\right)^2} \times i = \sqrt{\frac{193}{75} - \left(\frac{-9}{75}\right)^2} \times 5$$

$$= \sqrt{2.5733 - 0.0144} \times 5 = \sqrt{2.5589} \times 5 = 1.599 \times 5 = 7.995$$

By inspection, modal class is 20—25

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 20 + \frac{21 - 13}{42 - 13 - 16} \times 5$$

$$= 20 + \frac{40}{13} = 20 + 3.08 = 23.08$$

Thus, $\overline{X} = 21.9$, $\sigma = 7.995$, $Z = 23.08$

$$\therefore \text{ Coefficient of Skewness} = \frac{\overline{X} - Z}{\sigma} = \frac{21.9 - 23.08}{7.995} = -0.15$$

Example 3. Calculate Karl Pearson's Coefficient of Skewness from the following data:

Wages:	300—400	400—500	500—600	600—700	700—800
No. of workers:	5	10	10	3	2

Solution: Since the given series is a bimodel series, the following formula for calculating skewness is used:

$$\text{Coeff. of } S_K = \frac{3(\overline{X} - M)}{\sigma}$$

Calculation of Coefficient of Skewness

Wages	f	M.V. (m)	d	d'	fd'	fd'^2	c.f.
300—400	5	350	−200	−2	−10	20	5
400—500	10	450	−100	−1	−10	10	15
500—600	10	550	0	0	0	0	25
600—700	3	650	+100	+1	3	3	28
700—800	2	750	+200	+2	4	8	30
	$N = 30$				$\Sigma fd' = -13$	$\Sigma fd'^2 = 41$	

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 550 - \frac{13}{30} \times 10$$

$$= 550 - 43.33 = 506.67$$

$$\text{Median} = \frac{N}{2} = \frac{30}{2} = 15\text{th item.}$$

Median lies in the class interval 400—500.

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 400 + \frac{15 - 5}{10} \times 100 = 400 + 100 = 500$$

$$\sigma=\sqrt{\frac{\Sigma fd'^2}{N}-\left(\frac{\Sigma fd'}{N}\right)^2}\times i=\sqrt{\frac{41}{30}-\left(\frac{-13}{30}\right)^2}\times 100$$

$$=\sqrt{1.367-0.188}\times 100=\sqrt{1.179}\times 100=1.086\times 100=108.6$$

$\therefore\ \overline{X}=506.67$, Median $=500$, $\sigma=108.6$

$$\text{Coefficient of } S_K=\frac{3(\overline{X}-M)}{\sigma}=\frac{3(506.67-500)}{108.6}=\frac{20.01}{108.6}=+0.184$$

Example 4. Calculate the coefficient of skewness from the following data:

Mid-point:	15	20	25	30	35	40
Frequency:	12	18	25	24	20	21

Solution: As the mid-points of the different class intervals are given, we first find actual class intervals by using the formula $m\pm i/2$, where, m = mid-point and i = difference between two mid-points.

Calculation of Coefficient of Skewness

Classes	f	Mid-point m	d $(m-25)$	$d'=\frac{d}{5}$	fd'	fd'^2
12.5—17.5	12	15	−10	−10	−24	48
17.5—22.5	18	20	−5	−5	−18	18
22.5—27.5	25	25 = A	0	0	0	0
27.5—32.5	24	30	5	5	+24	24
32.5—37.5	20	35	10	+10	+40	80
37.5—42.5	21	40	15	+15	+63	189
	$N=120$				$\Sigma fd'=85$	$\Sigma fd'^2=359$

$$\overline{X}=A+\frac{\Sigma fd'}{N}\times i=25+\frac{85}{120}\times 5=25+3.542=28.542$$

$$\sigma=\sqrt{\frac{\Sigma fd'^2}{N}-\left(\frac{\Sigma fd'}{N}\right)^2}\times i=\sqrt{\frac{359}{120}-\left(\frac{85}{120}\right)^2}\times 5$$

$$=\sqrt{2.992-0.502}\times 5=\sqrt{2.4902}\times 5=1.578\times 5=7.89$$

Mode: The highest frequency is 25, mode lies corresponding to mid-point 25, *i.e.*, in the class 22.5—27.5.

$$Z=l_1+\frac{\Delta_1}{\Delta_1+\Delta_2}\times i$$

Here, $l_1=22\cdot5, \Delta_1=25-18=7, \Delta_2=25-24=1, i=5$

$$Z=22.5+\frac{7}{7+1}\times 5=22.5+4.375=26.875$$

$$S_{K_P}=\frac{\overline{X}-Z}{\sigma}=\frac{28.542-26.875}{7.89}=\frac{1.667}{7.89}=0.211$$

There is a low degree of positive skewness.

Example 5. Find the mean, mode, standard deviation and pearson's coefficient of skewness the following data:

Year under:	10	20	30	40	50	60
No. of Persons:	15	32	51	78	97	109

Solution: Since it is a cumulative frequency series, it should first be converted into sim frequency series:

Years	f	M.V. (m)	d	d'	fd'	fd'^2
0—10	15	5	–30	–3	–45	135
10—20	17	15	–20	–2	–34	68
20—30	19	25	–10	–1	–19	19
30—40	27	35A	0	0	0	0
40—50	19	45	+10	+1	+19	19
50—60	12	55	+20	+2	+24	48
Total	$N = 109$				$\Sigma fd' = -55$	$\Sigma fd'^2 = 28$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 35 + \frac{(-55)}{109} \times 10 = 35 - 5.045 = 29.95$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{289}{109} - \left(\frac{-55}{109}\right)^2} \times 10$$

$$= \sqrt{2.6513 - 0.2546} \times 10 = \sqrt{2.3967} \times 10$$

$$= 1.548 \times 10 = 15.48$$

By inspection, modal class is 30 – 40

$$\therefore\ Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 30 + \frac{27 - 19}{54 - 19 - 19} \times 10 = 30 + \frac{8}{16} \times 10 = 35$$

$$\therefore \text{ Coefficient of Skewness} = \frac{\overline{X} - Z}{\sigma} = \frac{29.95 - 35}{15.48} = -0.32$$

Example 6. Calculate Pearson's coefficient of skewness from the following data:

Marks above:	10	20	30	40	50	60	70	80	90
No. of students:	100	97	90	70	40	25	15	8	3

Solution: Since it is a cumulative frequency series, it should first be converted into simp frequency series.

Marks	f	M.V. (m)	d	d'	fd'	fd'^2
10—20	3	15	–40	–4	–12	48
20—30	7	25	–30	–3	–21	63
30—40	20	35	–20	–2	–40	80
40—50	30	45	–10	–1	–30	30

50—60	15	55 = A	0	0	10	0
60—70	10	65	+10	+1	10	10
70—80	7	75	+20	+2	14	28
80—90	5	85	+30	+3	15	45
90—100	3	95	+40	+4	12	48
	$N = 100$				$\Sigma fd' = -52$	$\Sigma fd'^2 = 352$

$$\bar{X} = A + \frac{\Sigma fd'}{N} \times i = 55 - \frac{52}{100} \times 10 = 55 - 5.2 = 49.8$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{352}{100} - \left(\frac{-52}{100}\right)^2} \times 10$$

$$= \sqrt{3.52 - 0.2704} \times 10 = \sqrt{3.2496} \times 10$$

$$= 1.8026 \times 10 = 18.02$$

By inspection, modal class is 40—50

$$\therefore \quad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 40 + \frac{30-20}{60-20-15} \times 10$$

$$= 40 + \frac{10 \times 10}{25} = 40 + 4 = 44$$

$$\therefore \text{ Coefficient of Skewness} = \frac{\bar{X} - Z}{\sigma} = \frac{49.8 - 44}{18.02} = \frac{5.8}{18.02} = 0.32$$

Example 7. For a group of 20 items, $\Sigma X = 1452, \Sigma X^2 = 144280$ and mode = 63.7. Find Karl Pearson's co-efficient of Skewness.

Solution: Coefficient of skewness $= \dfrac{\bar{X} - Z}{\sigma}$

$$\bar{X} = \frac{\Sigma X}{N} = \frac{1452}{20} = 72.6$$

$$\sigma = \sqrt{\frac{\Sigma X^2}{N} - (\bar{X})^2} = \sqrt{\frac{144280}{20} - (72.6)^2}$$

$$= \sqrt{7214 - 5270.76} = 44.082$$

$$\text{Coefficient of Skewness} = \frac{\bar{X} - Z}{\sigma} = \frac{72.6 - 63.7}{44.082} = 0.202$$

Example 8. In a certain distribution the following results were obtained:
C.V. = 40%, $\bar{X} = 25$, Z = 20
Find out coefficient of skewness.

Solution:

$$\text{C.V.} = \frac{\sigma}{\bar{X}} \times 100 \Rightarrow 40 = \frac{\sigma}{25} \times 100 \Rightarrow \sigma = 10$$

$$\text{Coefficient of Skewness} = \frac{\bar{X} - Z}{\sigma} = \frac{25 - 20}{10} = \frac{5}{10} = \frac{1}{2} = +0.5$$

EXERCISE 7.1

1. Calculate skewness and its coefficient from the following data: (Use Pearson's Formula).

Wages (Rs.) :	10	11	12	13	14	15	16
No . of workers :	4	7	9	15	8	5	2

[**Ans.** $S_K = -2.22$, coeff. of $S_K = -1.457$]

2. Calculate Pearson's Coefficient of Skewness from the following data:

Profits (Rs. lakhs):	70—80	60—70	50—60	40—50	30—40	20—30	10—20	0—10
No. of company:	11	22	30	35	21	11	6	5

[**Ans.** $\overline{X} = 46.84$, $Z = 47.86$, $\sigma = 17.08$, Coeff. of $S_K = -0.0304$]

3. Calculate Pearson's Coefficient of Skewness from the following:

Marks above:	0	10	20	30	40	50	60	70	80
No. of students:	150	140	100	80	80	70	30	14	0

[**Hint:** See Example 19] [**Ans.** $\overline{X} = 39.27, \sigma = 22.8, M = 45$; Coeff. of $S_K = -0.75$]

4. Calculate Pearson coefficient of skewness based on mean, median and standard deviation from the following data:

Age Groups:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70 and above
No. of workers:	18	16	15	12	10	5	2	1

[**Ans.** $\overline{X} = 26$ app., $M = 23.67$, $\sigma = 17.46$, Coeff. of $S_K = 0.4$]

5. The daily expenditure of 100 families is given below:

Daily expenditure:	0—20	20—40	40-60	60-80	80-100
No. of families:	13	?	27	?	16

If the mode of the distribution is 44, calculate Karl Pearson's coefficient of skewness.
[**Hint:** See Example 18] [**Ans.** Coeff. of $S_K = 0.237$]

6. Find Pearson's Coefficient of skewness from the following data:

Height (inches):	60—62	63—65	66—68	69—71	72—74
Frequency:	5	18	42	27	8

[**Ans.** Coeff. of $S_K = 0.034$]

7. From the marks secured by 120 students in Section A and 120 in Section B, the following measures are obtained :

Section A : $\overline{X} = 35.0$, $\sigma = 7$, $Z = 32$

Section B : $\overline{X} = 40.0$, $\sigma = 10$, $Z = 30$

Determine which distribution of marks is more skewed. [**Ans.** B is more skewed]

(2) Bowley's Method

Prof. Bowley has given another method of measuring skewness. It is based upon median (M), first quartile (Q_1) and third quartile (Q_3). It is also called **quartile method of measuring skewness.** Bowley has given the following formulae for measuring skewness:

Absolute Measure of Skewness	Bowley's Coefficient of Skewness
$Sk = Q_3 + Q_1 - 2M$	Coeff. of Sk $= \dfrac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$

Steps for Calculation

(1) Calculate Q_1, *i.e.*, first quartile

(2) Calculate Q_3, *i.e.*, third quartile

(3) Calculate M, *i.e.*, median

(4) Substitute these values in the formulae.

The following examples illustrate the procedure of calculating Bowley's measure of skewness:

Calculation of Bowley's Coefficient of Skewness in Discrete Series

Example 9. Find Bowley's coefficient of skewness for the following frequency distribution:

No. of children per family:	0	1	2	3	4	5	6
No. of families:	7	10	16	25	18	11	8

Solution:

Calculation of Bowley's Coefficient of Skewness

No. of children (X)	No. of families (f)	c.f.
0	7	7
1	10	17
2	16	33
3	25	58
4	18	76
5	11	87
6	8	95
	$N = 95$	

Bowley's coefficient of skewness is given by

$$\text{Coefficient of } S_K = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$Q_1 = \text{Size of} \left(\frac{N+1}{4}\right)^{th} \text{item} = \frac{95+1}{4} = 24^{th} \text{item}$$

Size of 24th item is 2. Hence, $Q_1 = 2$

$$Q_3 = \text{Size of} \frac{3(N+1)}{4} \text{th item} = \frac{3(95+1)}{4} = 72 \text{ th item}$$

Size of 72th item is 4. Hence, $Q_3 = 4$

$$M = \text{Size of } \frac{1}{2}(N+1)^{th} \text{ item} = \frac{95+1}{2} = 48^{th} \text{ item.}$$

Size of 48^{th} item is 3. Hence, Median = 3

$$\text{Coeff. of skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$= \frac{4+2-2\times 3}{4-2} = \frac{0}{2} = 0$$

Calculation of Bowley's Coefficient of Skewness in Continuous Series

Example 10. Calculate coefficient of skewness based on quartiles and median from the following data:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60	60—70	70—80
No. of students:	10	25	20	15	10	35	25	10

Solution: **Calculation of Bowley's Coefficient of Skewness**

Marks	f	$c.f.$
0—10	10	10
10—20	25	35
20—30	20	55
30—40	15	70
40—50	10	80
50—60	35	115
60—70	25	140
70—80	10	150
	$N = 150$	

$Q_1 = \frac{N}{4} = \frac{150}{4} = 37.5$th item. Q_1 lies in the class interval 20—30

$$Q_1 == l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 20 + \frac{37.5-35}{20} \times 10 = 20 + 1.25 = 21.25$$

$$Q_3 = \frac{3N}{4} = \frac{3}{4} \times 150 = 112.5\text{th item.}$$

Q_3 lies in the class interval 50—60.

$$Q_3 = l_1 + \frac{\frac{3}{4}N - c.f.}{f} \times i = 50 + \frac{112.5-80}{35} \times 10 = 50 + 9.29 = 59.29$$

$$\text{Median item} = \frac{N}{2} = \frac{150}{2} = 75 \text{ th item.}$$

Median lies in the class interval 40—50.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 40 + \frac{75 - 70}{10} \times 10 = 45$$

$\therefore \quad Q_1 = 21.25, Q_3 = 59.29, M = 45$

$$\text{Coeff. of } S_K = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$= \frac{59.29 + 21.25 - 2 \times 45}{59.29 - 21.25}$$

$$= \frac{80.54 - 90}{38.04} = \frac{-9.46}{38.04} = -0.249$$

Example 11. Calculate Coefficient of Q.D. and Bowley's Coefficient of Skewness from the data given below:

Profits in lakhs (less than):	10	20	30	40	50	60	70
No. of Companies:	8	20	40	50	56	59	60

Solution: Since it is a cummulative frequency series, first we convert it into simple frequency series:

Size	f	$c.f.$
0—10	8	8
10—20	12	20
20—30	20	40
30—40	10	50
40—50	6	56
50—60	3	59
60—70	1	60
	$N = 60$	

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

$Q_1 = \text{Size of} \left(\frac{N}{4}\right) \text{th item} = \frac{60}{4} = 15\text{th item}$. Q_1 lies in the class 10—20.

$$Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 10 + \frac{15 - 8}{12} \times 10 = 10 + 5.833 = 15.833$$

$Q_3 = \text{Size of} \left(\frac{3}{4} N\right) \text{th item} = \frac{3}{4} \times 60 = 45\text{th item}$. Q_3 lies in the class 30—40.

$$Q_3 = l_1 + \frac{\frac{3}{4} N - c.f.}{f} \times i = 30 + \frac{45 - 40}{10} \times 10 = 35$$

$$\therefore \quad \text{Coeff. of Q.D.} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{35 - 15.833}{35 + 15.833} = \frac{19.167}{50.833} = 0.377$$

Median item = Size of $\frac{N}{2}$th item $= \frac{60}{2} = 30$th item. Median lies in the class 20—30.

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 20 + \frac{30 - 20}{20} \times 10 = 20 + 5 = 25$$

$$\text{Coeff. of } S_K = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{35 + 15.833 - 2(25)}{35 - 15.833} = \frac{0.833}{19.167} = 0.043$$

Example 12. Calculate Bowley's Coefficient of Skewness from the following data:

Mid-values:	1	2	3	4	5	6	7	8	9	10
Frequency:	2	9	11	14	20	24	20	16	5	2

Solution: As the mid-values of the different class intervals are given, we first find actual class intervals by using the formula $m \pm \frac{i}{2}$ where m = mid-value, i = difference between two mid-values.

Classes	Mid-values (m)	f	$c.f.$
0.5—1.5	1	2	2
1.5—2.5	2	9	11
2.5—3.5	3	11	22
3.5—4.5	4	14	36
4.5—5.5	5	20	56
5.5—6.5	6	24	80
6.5—7.5	7	20	100
7.5—8.5	8	16	116
8.5—9.5	9	5	121
9.5—10.5	10	2	123
		$N = 123$	

$$Q_1 = \text{size of } \frac{N}{4} \text{ th item} = \frac{123}{4} = 30.75$$

Q_1 lies in the class 3.5 – 4.5.

$$\therefore \quad Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 3.5 + \frac{30.75 - 22}{14} \times 1$$

$$= 3.5 + 0.625 = 4.125$$

$$Q_3 = \text{size of } \frac{3N}{4} \text{ th item} = \frac{3 \times 123}{4} = 92.25\text{th item}$$

Q_3 lies in the class 6.5 – 7.5

$$Q_3 = l_1 + \frac{\frac{3N}{4} - c.f.}{f} \times i = 6.5 + \frac{92.25 - 80}{20} \times 1$$

$$= 6.5 + 0.6125 = 7.1125$$

Median item = Size of $\left(\frac{N}{2}\right)$th item $= \frac{123}{2} = 61.5$th item.

Median lies in the class 5.5 – 6.5.

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 5.5 + \frac{61.5 - 56}{24} \times 1$$

$$= 5.5 + 0.2292 = 5.7292$$

$$\text{Coeff. of } S_K = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{7.1125 + 4.125 - 2(5.7292)}{7.1125 - 4.125}$$

$$= \frac{11.2375 - 11.4582}{2.9875} = \frac{-0.2207}{2.9875} = -0.073$$

Example 13. For a distribution, Bowley's coefficient of skewness is –0.56, $Q_1 = 16.4$ and Median = 24.2. Find Q_3 and coefficient of quartile deviation.

Solution: **Given:** Coeff. of $S_K = -0.56$, $Q_1 = 16.4$, $M = 24.2$

$$\text{Coeff. of } S_K = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

$$-0.56 = \frac{Q_3 + 16.4 - 2(24.2)}{Q_3 - 16.4}$$

$$\Rightarrow -0.56\,(Q_3 - 16.4) = Q_3 + 16.4 - 48.4$$

$$-0.56Q_3 + 9.184 = Q_3 + 16.4 - 48.4$$

$$-0.56Q_3 - Q_3 = 16.4 - 48.4 - 9.184$$

$$-1.56Q_3 = -41.184$$

$$Q_3 = 26.4$$

$$\therefore \quad Q.D. = \frac{Q_3 - Q_1}{2} = \frac{26.4 - 16.4}{2} = 5.$$

ample 14. Find coefficient of skewness from the following information:

Difference of two quartiles = 8

Mode = 11

Sum of two quartiles = 22

Mean = 8

Solution: We know, $Z = 3M - 2\overline{X}$

$\Rightarrow$ $3M = Z + 2\overline{X} = 11 + 2 \times 8 = 27$

$\Rightarrow$ $M = \dfrac{27}{3} = 9$

$$\text{Bowley's Coeff. of Skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1} = \frac{22 - 2\times 9}{8} = \frac{22-18}{8} = \frac{4}{8} = \frac{1}{2} = 0.5$$

Example 15. For a distribution the distance of the median from the first quartile is five times of the third quartile from the median. Calculate Bowley's Coefficient of Skewness for the distribution.

Solution: We are given : $M - Q_1 = 5(Q_3 - M)$

Bowley's coefficient of skewness is given by:

$$S_K(\text{Bowley}) = \frac{(Q_3 - M) - (M - Q_1)}{(Q_3 - M) + (M - Q_1)} = \frac{(Q_3 - M) - 5(Q_3 - M)}{(Q_3 - M) + 5(Q_3 - M)}$$

$$= \frac{-4(Q_3 - M)}{6(Q_3 - M)} = \frac{-2}{3} = -0.67$$

EXERCISE 7.2

1. Calculate coefficient of quartile deviation and Bowley's coefficient of skewness from the following data:

Size :	Below 10	10—20	20—30	30—40	40—50	Above 50
f:	5	12	20	16	5	2

[**Ans.** Q_1 = 18.33, Q_3 = 35, Coeff. of Q.D. = 0.313, Coeff. of S_K = 0.019]

2. Find out Quartiles and Coefficient of Skewness from the following data:

X:	3	4	5	6	7	8	9	10
f:	2	5	7	11	10	8	5	3

[**Ans.** Q_1 = 5, M = 7, Q_3 = 8, Coeff.of S_K = – 0.332]

3. Calculate Bowley's Coefficient of skewness from the following data:

Mid-values:	75	100	125	150	175	200	225	250
Frequency:	35	40	48	100	125	80	50	22

[**Ans.** Coeff. of S_K = – 0.032]

4. The mean, mode and Q.D. of a distribution are 42, 36 and 15 respectively. If its Bowley's coefficient of skewness is 1/3, find the values of two quartiles.

   [**Hint:** Find $M = \overline{X} - \dfrac{1}{3}Z$] [**Ans.** $Q_1 = 20, Q_3 = 50$]

5. Calculate the quartile co-efficient of skewness for the following distribution:

Class:	1—5	6—10	11—15	16—20	21—25	26—30	31—35
f:	3	4	68	30	10	6	2

[**Ans.** Coeff. of S_K = 0.265]

(3) Kelly's Method

The third method of measuring skewness is given by Prof. Kelly. It is based on percentiles and deciles. Kelly has given the following formulae for measuring skewness:

Absolute Measures of S_K	**Coefficient of S_K**
1. $S_K = P_{90} + P_{10} - 2M$	1. Coeff. of $S_K = \dfrac{P_{90} + P_{10} - 2M}{P_{90} - P_{10}}$
or	or
2. $S_K = D_9 + D_1 - 2M$	2. Coeff. of $S_K = \dfrac{D_9 + D_1 - 2M}{D_9 - D_1}$

This method is not very popular in practice. It is suitable when the skewness is based on percentiles or deciles.

Steps for Calculation

(1) Calculate P_{90}, *i.e.*, Nineteenth Percentile

(2) Calculate P_{10}, *i.e.*, Tenth Percentile

(3) Calculate M, *i.e.*, Median.

(4) Substitute these in the formulae.

Example 16. From the data given below, find out Kelly's Coefficient of Skewness based on percentiles:

Marks:	0—10	10—20	20—30	30—40	40—50	50—60
No. of students:	4	6	20	10	7	3

Solution:

Marks	***f***	***c.f.***
0—10	4	4
10—20	6	10
20—30	20	30
30—40	10	40
40—50	7	47
50—60	3	50
	$N = 50$	

$$P_{90} = \text{Size of } \frac{90N}{100} \text{ th item} = \frac{90 \times 50}{100} = 45\text{th item.}$$

P_{90} lies in the class interval 40—50.

$$\therefore \quad P_{90} = l_1 + \frac{\frac{90N}{100} - c.f.}{f} \times i = 40 + \frac{45 - 40}{7} \times 10 = 47.14$$

$$P_{10} = \text{Size of } \frac{10N}{100} \text{ th item} = \frac{10 \times 50}{100} = 5\text{th item.}$$

P_{10} lies in the class interval 10 — 20.

$$\therefore \quad P_{10} = l_1 + \frac{\frac{10N}{100} - c.f.}{f} \times i$$

$$= 10 + \frac{5-4}{6} \times 10 = 11.67$$

$$\text{Median item} = \text{Size of } \frac{N}{2}\text{th item} = \frac{50}{2} = 25\text{th item.}$$

M lies in the class interval 20–30.

$$\therefore \quad M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 20 + \frac{25-10}{20} \times 10$$

$$= 20 + \frac{15 \times 10}{20} = 27.5$$

$$\therefore \quad P_{90} = 47.17, \ P_{10} = 11.67, \ M = 27.5$$

$$\text{Kelly's Coeff. of Skew.} = \frac{P_{90} + P_{10} - 2M}{P_{90} - P_{10}} = \frac{47.14 + 11.67 - 2 \times 27.5}{47.14 - 11.67} = 0.11$$

EXERCISE 7.3

1. Calculate Kelly's coefficient of skewness from the data given below:

X:	110—115	115—120	120—125	125—130	130—135
f:	4	10	26	49	72
X:	135—140	140—145	145—150	150—155	155—160
f:	90	52	33	17	7

[**Ans.** Coeff. of $S_K = 0.0$

2. Compute Kelly's coefficient of skewness based on percentiles from the following:

Marks:	15—20	20—25	25—30	30—35	35—40	40—45
No. of students:	1	2	15	22	7	3

[**Ans.** Coeff. of $S_K = 0.0$

MISCELLANEOUS SOLVED EXAMPLES

Example 17: Calculate arithmetic mean, mode, standard deviation and coefficient of skewness f[illegible] the following:

Marks (less than):	10	20	30	40	50	60
No. of students:	4	10	30	40	47	50

Solution: The above data are in cumulative form. Firstly these data will be converted into simple form:

Marks (X)	f	M.V. (m)	d	d'	fd'	fd'^2
0—10	4	5	–20	–2	–8	16
10—20	6	15	–10	–1	–6	6
20—30	20	25 = A	0	0	0	0
30—40	10	35	+10	+1	10	10
40—50	7	45	+20	+2	14	28
50—60	3	55	+30	+3	9	27
	$N = 50$				$\Sigma fd' = 19$	$\Sigma fd'^2 = 87$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 25 + \frac{19}{50} \times 10 = 28.8$$

By inspection, mode lies in the class 20—30

$$\therefore \qquad Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i = 20 + \frac{20-6}{40-6-10} \times 10$$

$$= 20 + \frac{14}{24} \times 10 = 20 + 5.83 = 25.83$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{87}{50} - \left(\frac{19}{50}\right)^2} \times 10$$

$$= \sqrt{1.74 - 0.1444} \times 10 = 1.263 \times 10 = 12.63$$

$$\text{Coefficient of skewness} = \frac{\overline{X} - Z}{\sigma} = \frac{28.8 - 25.83}{12.63} = \frac{2.97}{12.63} = +0.235$$

Example 18. The daily expenditure of 100 families is given below:

Daily expenditure:	0—20	20—40	40—60	60—80	80—100
No. of families:	13	?	27	?	16

If the mode of the distribution is 44, calculate Karl Pearson Coefficient of skewness.

Solution: Let the missing frequency for the class 20—40 be X. The frequency for the class 60—80 shall be $100 - (56 + x) = 44 - x$

Expenditure	f	$c.f.$
0—20	13	13
20—40	x	$13+x$
40—60	27	$40+x$
60—80	$44-x$	84
80—100	16	100
	$N = 100$	

The formula of mode is:

$$Z = l_1 + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times i$$

Since the given modal value is 44, it lies in the class 40—60.

$$44 = 40 + \frac{27 - x}{54 - x - (44 - x)} \times 20$$

$$44 = 40 + \frac{27 - x}{10} \times 20$$

or $$44 - 40 = \frac{27 - x}{10} \times 20$$

or $$4 = \frac{27 - x}{10} \times 20$$

or $$40 = (27 - x)20$$

$$27 - x = 2$$

or $$x = 25$$

Thus, the frequency for the class 20—40 is 25 and the frequency of the class 60—80 is 44—25 = 19. Thus, the completed frequency distribution is:

0—20	20—40	40—60	60—80	80—100
13	25	27	19	16

Calculation of Coefficient of Skewness

Daily Expenditure	f	M.V. (m)	d	d'	fd'	fd'^2
0—20	13	10	− 40	− 2	− 26	52
20—40	25	30	− 20	− 1	− 25	25
40—60	27	50 = A	0	0	0	0
60—80	19	70	+ 20	+ 1	19	19
80—100	16	90	+ 40	+ 2	32	64
	$N = 100$				$\Sigma fd' = 0$	$\Sigma fd'^2 = 160$

Karl Pearson coefficient of skewness $= \frac{\overline{X} - Z}{\sigma}$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 50 + \frac{0}{100} \times 10 = 50$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{160}{100} - \left(\frac{0}{100}\right)^2} \times 20$$

$$= \sqrt{1.6 - 0} \times 20$$

$$= 1.265 \times 20 = 25.3$$

$Z = 44$ (given)

$$\therefore \quad \text{Coeff. of } S_K = \frac{\overline{X} - Z}{\sigma} = \frac{50 - 44}{25.3} = \frac{6}{25.3} = 0.237$$

Example 19. Calculate Karl Pearson's Coefficient of skewness from the following data:

Marks above:	0	10	20	30	40	50	60	70	80
No. of students:	150	140	100	80	80	70	30	14	0

Solution: The above data are in cummulative frequency. Firstly we convert these data into simple form:

Marks	f	M.V. (m)	d	d'	fd'	fd'^2	$c.f.$
0—10	10	5	–40	–4	–40	160	10
10—20	40	15	–30	–3	–120	360	50
20—30	20	25	–20	–2	–40	80	70
30—40	0	35	–10	–1	0	0	70
40—50	10	45 = A	0	0	0	0	80
50—60	40	55	+10	+1	40	40	120
60—70	16	65	+20	+2	32	64	136
70—80	14	75	+30	+3	42	126	150
80—90	0	85	+40	+4	0	0	150
	$N = 150$				$\Sigma fd' = -86$	$\Sigma fd'^2 = 830$	

As this is a bimodal series (*i.e.*, there are two maximum frequencies), we will find coefficient of skewness by using the formula

$$\text{Coeffi. of } S_K = \frac{3(\overline{X} - M)}{\sigma}$$

$$\overline{X} = A + \frac{\Sigma fd'}{N} \times i = 45 + \frac{(-86)}{150} \times 10 = 45 - \frac{86}{15} = 39.27$$

$$\text{Median item} = \text{Size of } \frac{N}{2} = \frac{150}{2} = 75\text{th item. Median lies in class 40—50.}$$

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i$$

$$= 40 + \frac{75 - 70}{10} \times 10 = 40 + \frac{5}{10} \times 10 = 40 + 5 = 45$$

$$\sigma = \sqrt{\frac{\Sigma fd'^2}{N} - \left(\frac{\Sigma fd'}{N}\right)^2} \times i = \sqrt{\frac{830}{150} - \left(\frac{-86}{150}\right)^2} \times 10$$

$$= \sqrt{5.53 - 0.33} \times 10 = \sqrt{5.20} \times 10 = 2.28 \times 10 = 22.8$$

$$\therefore \text{ Coeff. of skewness} = \frac{3(\overline{X} - M)}{\sigma} = \frac{3(39.27 - 45)}{22.8} = \frac{3(-5.73)}{22.8} = \frac{-17.19}{22.8} = -0.75$$

Example 20. You are given the position in a factory before and after the settlement of an industrial dispute. Comment on the gains or losses from the point of view of workers and that of management.

	Before	After
No. of workers	2,400	2,350
Mean wages (Rs.)	45.5	47.5
Median wages (Rs.)	48.0	45.0
Standard Deviation (Rs.)	12.0	10.0

Solution: The following comments can be made on the basis of the information given:

(*i*) By comparing the total wage bill we can comment on the increase or decrease in the level of wages.

Total wage bill *before* the settlement of dispute = 2,400 × 45.5 = Rs. 1,09,200

Total wage bill *after* the settlement of dispute = 2,350 × 47.5 = Rs. 1,11,625.

Hence the total wage bill has gone up after the settlement of dispute even though the number of workers has decreased from 2,400 to 2,350. This means that the average wage is now higher. This is definitely a gain to the workers.

Conversely, we cannot say that increased wage bill is necessarily a loss to management because if it results in greater efficiency of workers and, therefore, higher productivity, it would be a positive gain to management also.

(*ii*) Median before settlement of the dispute was 48 and after settlement it is 45. This means that formerly 50% of workers used to get above Rs. 48 and now they get only above Rs. 45.

(*iii*) By comparing the coefficient of variation before and after the settlement of dispute we can comment on the distribution of wages.

Coefficient of variation *before* the settlement of dispute

$$C.V. = \frac{\sigma}{\overline{X}} \times 100, \text{ where, } \sigma = 12, \overline{X} = 45.5$$

$$\therefore \quad C.V. = \frac{12}{45.5} \times 100 = 26.37$$

Coefficient of variation *after* the settlement of dispute $\sigma = 10, \overline{X} = 47.5$

$$\therefore \quad C.V. = \frac{10}{47.5} \times 100 = 21.05$$

Since the value of the coefficient of variation has decreased from 26.4 to 21.05 there is sufficient evidence to conclude that wages are more uniformly distributed after the settlement of dispute or, in other words, there is lesser inequality in the distribution of wages after the dispute is settled.

(*iv*) By comparing skewness, we can comment upon the nature of the distribution.

Coefficient of skewness *before* the settlement of dispute

$$S_{Kp} = \frac{3(\overline{X} - M)}{\sigma} = \frac{3(45.5 - 48)}{12} = \frac{-7.5}{12} = -0.625$$

Coefficient of skewness *after* the settlement of dispute

$$S_{Kp} = = \frac{3(47.5 - 45)}{10} = \frac{7.5}{10} = +0.75$$

Thus, the distribution is positively skewed after the settlement of dispute whereas it was negatively skewed before the settlement of dispute. This suggests that the number of workers getting low wages has increased considerably and that of workers getting high wages fallen, though the actual wage of workers has increased.

Example 21. Calculate Bowley's coefficient of skewness for the following data:

Size :	5—7	8—10	11—13	14—16	17—19
f:	14	24	38	20	4

Solution: Class intervals are in inclusive form. For finding median and quartiles, we convert the given distribution into exclusive form:

X	f	$c.f.$
4.5—7.5	14	14
7.5—10.5	24	38
10.5—13.5	38	76
13.5—16.5	20	96
16.5—19.5	4	100
	$N = 100$	

$$Q_1 = \text{Size of } \frac{N}{4} = \frac{100}{4} = 25\text{th item.}$$

Q_1 lies in the class interval 7.5—10.5.

$$\therefore \quad Q_1 = l_1 + \frac{\frac{N}{4} - c.f.}{f} \times i = 7.5 + \frac{25 - 14}{24} \times 3 = 8.87$$

$$Q_3 = \text{Size of } \frac{3N}{4} = \frac{3(100)}{4} = 75\text{th item.}$$

Q_3 lies in class interval 10.5—13.5.

$$Q_3 = l_1 + \frac{\frac{3}{4}N - c.f.}{f} \times i = 10.5 + \frac{75 - 38}{38} \times 3 = 13.42$$

$$\text{Median} = \text{Size of } \frac{N}{2} = \frac{100}{2} = 50\text{th item.}$$

M lies in the class interval 10.5—13.5.

$$M = l_1 + \frac{\frac{N}{2} - c.f.}{f} \times i = 10.5 + \frac{50-38}{38} \times 3 = 11.447$$

$\therefore$ Bowley's Coefficient of Skewness $= \dfrac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$

$$= \frac{13.42 + 8.87 - 2 \times 11.447}{13.42 - 8.87} = -0.13$$

Example 22. In a frequency distribution, the coefficient of skewness based on quartiles is 0.6, the sum of upper and lower quartiles is 100 and Median is 38, find the value of lower and upper quartiles.

Solution: Given: $Q_1 + Q_3 = 100, M = 38$ and $S_K = 0.6$

Bowley coeff. of skewness $= \dfrac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$

Substituting the values in Bowley's formula, we get

$$0.6 = \frac{100 - 2(38)}{Q_3 - Q_1}$$

or $\quad 0.6(Q_3 - Q_1) = 100 - 76 = 24$

or $\quad Q_3 - Q_1 = \dfrac{24}{0.6} = 40$

Now $Q_3 + Q_1 = 100$ ...(*i*) (Given)

$Q_3 - Q_1 = 40$ ...(*ii*)

By adding (*i*) and (*ii*), we get

$2Q_3 = 140$

$\Rightarrow \quad Q_3 = 70$

Also $\quad Q_1 = 100 - 70 = 30$

$\therefore \; Q_1 = 30, \; Q_3 = 70$

Example 23. Pearson's coefficient of skewness for a distribution is 0.4 and coefficient of variance is 30%. Its mode is 88. Find the mean and median.

Solution: We are given $S_{Kp} = 0.4$, mode = 88 and coeff. of variance 30 %. We have to calculate mean and median.

$$S_K = \frac{\text{Mean} - \text{Mode}}{\sigma}$$

Coeff. of variance is 30% or $\dfrac{\sigma}{\bar{X}} = 0.3$

$$= \frac{1 - \frac{\text{Mode}}{\text{Mean}}}{\frac{\text{S.D.}}{\text{Mean}}} = \frac{1 - \frac{88}{\text{Mean}}}{0.3} = 0.4 = 1 - \frac{88}{\text{Mean}} = 0.12 = 1 - \frac{88}{\text{Mean}}$$

$$\Rightarrow \qquad \frac{88}{\text{Mean}} = 1 - 0.12 = 0.88$$

$$0.88 \text{ Mean} = 88 \text{ or Mean} = \frac{88}{0.88} = 100$$

$$\text{Mode} = 3 \text{ Median} - 2\overline{X}$$

$$88 = 3 \text{ Median} - 2(100)$$

or $$3 \text{ Median} = 288 \text{ or Median} = 96$$

Hence, mean and median are 100 and 96 respectively.

Example 24. Consider the following distributions:

Items	Distribution A	Distribution B
Mean	100	90
Mode	90	80
Standard Deviation	10	10

(*i*) Distribution A has the same degree of the variation as distribution B.

(*ii*) Both distributions have the same degree of skewness. True/False? Comment, giving reasons.

Solution: (*i*) C.V. for distribution A $= 100 \times \dfrac{\sigma_A}{\overline{X}_A} = 100 \times \dfrac{10}{100} = 10$

C.V. for distribution A $= 100 \times \dfrac{\sigma_B}{\overline{X}_B} = 100 \times \dfrac{10}{90} = 11.11$

Since C.V.(B) > C.V.(A), the distribution B is more variable than the distribution A. Hence, the given statement that the distribution A has the same degree of variation as distribution B is wrong.

(*ii*) Karl Pearson's coefficient of skewness for the distributions A and B is given by:

$$S_K(\text{A}) = \frac{3(\overline{X} - M)}{\sigma} = \frac{3(100-90)}{10} = 3$$

and $$S_K(\text{B}) = \frac{3(90-80)}{10} = 3$$

Since $S_K(\text{A}) = S_K(\text{B}) = 3$, the statement that both the distributions have the same degree of skewness is true.

IMPORTANT FORMULAE

MEASURES OF SKEWNESS

1. Pearson's Measures

Absolute skewness:

$$\text{Skewness} = \overline{X} - Z$$

when mode (Z) is ill defined, then

$$\text{Skewness} = 3(\overline{X} - M)$$

Relative measure of skewness:

(*i*) $\text{Coefficient of Skewness} = \dfrac{\overline{X} - Z}{\sigma}$

(*ii*) $\text{Coefficient of Skewness} = \dfrac{3(\overline{X} - M)}{\sigma}$ (when Z is ill defined)

2. Bowley's Measures

Absolute skewness:

$$\text{Skewness} = Q_3 + Q_1 - 2M$$

Relative measure of skewness:

$$\text{Coefficient of skewness} = \frac{Q_3 + Q_1 - 2M}{Q_3 - Q_1}$$

3. Kelly's Measures

Absolute skewness:

$$\text{Skewness} = P_{90} + P_{10} - 2M \quad \text{or} \quad D_9 + D_1 - 2M$$

Relative measure of skewness:

$$\text{Coefficient of skewness} = \frac{P_{90} + P_{10} - 2M}{P_{90} - P_{10}} \quad \text{or} \quad \frac{D_9 + D_1 - 2M}{D_9 - D_1}$$

QUESTIONS

1. What is skewness? How does it differ from dispersion? Describe the various measures of skewness.
2. Distinguish between dispersion and skewness and point out the various methods of measuring skewness.
3. What is skewness? What are tests of skewness? Draw rough sketches to indicate different types of skewness and locate rough the relative position of mean, median and mode in each case.
4. (*i*) Define skewness. How does it differ from dispersion?

   (*ii*) Explain different measures of skewness.

Moments and Measures of Kurtosis

8

INTRODUCTION

Like average, dispersion and skewness, kurtosis is the fourth characteristic of a frequency istribution which gives us an idea about the shape of a frequency distribution. Kurtosis indicates, hether a frequency distribution is more flat-topped or more peaked than the normal distribution. efore taking up a detailed study of Kurtosis, it is necessary to introduce the concept of moments hich is essential for its study.

MOMENTS

Moments are the general statistical measures used to describe and analyse the characteristics of frequency distribution. There are three basis for defining moments:

(1) Moments about the Mean

(2) Moments about Assumed Mean

(3) Moments about zero.

(1) Moments about the Mean

The moment about the mean are called **central moments.** They are denoted by Greek symbol μ (read as mu). If $X_1, X_2, X_3 ... X_n$ be the n values of a variable X and $\overline{X}$ be its actual mean, then the rth moment about the mean is defined and given by:

$$\mu_r = \frac{\Sigma(X-\overline{X})^r}{N} \quad \text{where } \mu_r = r\text{th moment about the mean, } r = 1, 2, 3, 4...$$

For a frequency distribution (or grouped data), the rth moment about mean is defined as:

$$\mu_r = \frac{\Sigma f(X-\overline{X})^r}{N} \quad \text{where } N = \Sigma f, r = 1, 2, 3, 4...$$

Putting $r = 1, 2, 3$ and 4, the various central moments are as follows:

Individual Series	**Discrete/Continuous Series**
$\mu_1 =$ First Central Moment $= \frac{\Sigma(X-\overline{X})^1}{N} = 0$	$\mu_1 = \frac{\Sigma f(X-\overline{X})^1}{N} = 0$
$\mu_2 =$ Second Central Moment $= \frac{\Sigma(X-\overline{X})^2}{N}$	$\mu_2 = \frac{\Sigma f(X-\overline{X})^2}{N}$
$\mu_3 =$ Third Central Moment $= \frac{\Sigma(X-\overline{X})^3}{N}$	$\mu_3 = \frac{\Sigma f(X-\overline{X})^3}{N}$
$\mu_4 =$ Fourth Central Moment $= \frac{\Sigma(X-\overline{X})^4}{N}$	$\mu_4 = \frac{\Sigma f(X-\overline{X})^4}{N}$

Moments are extended to higher powers but in practice the first four moments are obtained because of the difficulty of computation.

Note 1: The first central moment μ_1 is always zero, *i.e.*, $\mu_1 = 0$ because the sum of the deviations taken from the mean is zero $(\Sigma(X-\overline{X})=0)$.

Note 2: The second central moment μ_2 is the square of the standard deviation, *i.e.*, $\mu_2 = (S.D.)^2$ is equal to the variance of the distribution, *i.e.*, $\mu_2 = \text{Variance} = \sigma^2$.

Example 1. Find the first four central moments of the following numbers: 1, 3, 7, 9, 10

Solution:

Calculation of Moments

X	$\overline{X}=6$ $(X-\overline{X})$	$(X-\overline{X})^2$	$(X-\overline{X})^3$	$(X-\overline{X})^4$
1	$1-6=-5$	25	−125	625
3	$3-6=-3$	9	−27	81
7	$7-6=1$	1	1	1
9	$9-6=3$	9	27	81
10	$10-6=4$	16	64	256
$\Sigma X=30$ $N=5$	$\Sigma(X-\overline{X})=0$	$\Sigma(X-\overline{X})^2=60$	$\Sigma(X-\overline{X})^3=-60$	$\Sigma(X-\overline{X})^4=104$[illegible]

$$\overline{X} = \frac{\Sigma X}{N} = \frac{30}{5} = 6$$

$$\mu_1 = \frac{\Sigma(X-\overline{X})}{N} = \frac{0}{5} = 0;\ \mu_2 = \frac{\Sigma(X-\overline{X})^2}{N} = \frac{60}{5} = 12$$

$$\mu_3 = \frac{\Sigma(X-\overline{X})^3}{N} = \frac{-60}{5} = -12;\ \mu_4 = \frac{\Sigma(X-\overline{X})^4}{N} = \frac{1044}{5} = 208.8.$$

Example 2. Calculate $\mu_1, \mu_2, \mu_3, \mu_4$ for the following frequency distribution:

Marks:	5—15	15—25	25—35	35—45	45—55	55—65
No. of students:	10	20	25	20	15	10

Solution:

Calculation of Moments

Marks	No. of students (f)	Mid-values X	fX	$\overline{X}=34$ $(X-\overline{X})$	$f(X-\overline{X})$	$f(X-\overline{X})^2$	$f(X-\overline{X})^3$	$f(X-\overline{X})^4$
5—15	10	10	100	−24	−240	5760	−138240	3317760
15—25	20	20	400	−14	−280	3920	−54880	768320
25—35	25	30	750	−4	−100	400	−1600	6400
35—45	20	40	800	6	120	720	4320	25920
45—55	15	50	750	16	240	3840	61440	983040
55—65	10	60	600	26	260	6760	175760	4569760
	$N=100$		$\Sigma fX=3400$		$\Sigma f(X-\overline{X})=0$	$\Sigma f(X-\overline{X})^2=21400$	$\Sigma f(X-\overline{X})^3=46800$	$\Sigma f(X-\overline{X})^4=29671200$

Now, $\bar{X} = \frac{\Sigma fX}{N} = \frac{3400}{100} = 34;$ $\mu_1 = \frac{\Sigma f(X-\bar{X})}{N} = \frac{0}{100} = 0$

$\mu_2 = \frac{\Sigma f(X-\bar{X})^2}{N} = \frac{21400}{100} = 214;$ $\mu_3 = \frac{\Sigma f(X-\bar{X})^3}{N} = \frac{46800}{100} = 468;$

$\mu_4 = \frac{\Sigma f(X-\bar{X})^4}{N} = \frac{9671200}{100} = 96712.$

(2) Moments about Assumed Mean

When the arithmetic mean is not in whole numbers but in fractions, the calculation of deviations from the mean would involve too many calculations and would take a lot of time. In such a case, the moments from the assumed mean are first calculated and then converted into central moments. The moments about assumed mean are called **non-central moments**. They are denoted by the Greek symbol μ'(pronounced as mu dash). If $X_1, X_2, X_3, ..., X_n$ be the n values of a variable X and A is its assumed mean, then the rth moment about assumed mean is defined and given by:

$$\mu'_r = \frac{\Sigma(X-A)^r}{N}$$ Where μ'_r = rth moment about assumed mean A

$r = 1, 2, 3, 4.$

For a frequency distribution, the rth moment about assumed mean (A) is defined as:

$$\mu'_r = \frac{\Sigma f(X-A)^r}{N}$$

Putting r = 1, 2, 3 and 4, the various non-central moments are as follows:

Individual Series	Discrete/Continuous Series
μ'_1 = First Moment about A = $\frac{\Sigma(X-A)^1}{N}$	$\mu'_1 = \frac{\Sigma f(X-A)^1}{N} = \frac{\Sigma fd^1}{N}$
μ'_2 = Second " " " = $\frac{\Sigma(X-A)^2}{N}$	$\mu'_2 = \frac{\Sigma f(X-A)^2}{N} = \frac{\Sigma fd^2}{N}$
μ'_3 = Third " " " = $\frac{\Sigma(X-A)^3}{N}$	$\mu'_3 = \frac{\Sigma f(X-A)^3}{N} = \frac{\Sigma fd^3}{N}$
μ'_4 = Fourth " " " = $\frac{\Sigma(X-A)^4}{N}$	$\mu'_4 = \frac{\Sigma f(X-A)^4}{N} = \frac{\Sigma fd^4}{N}$

Moments can be extended to higher powers in a similar fashion, but in practice, only the first four moments are computed because of the difficulty of computation.

Note: If there is some common factor in the X-column/Mid-value column (m) of a frequency distribution, the computation process of moments can further be simplified by dividing the deviations (d) taken from assumed mean by a common factor (i), and multiply the various moments by i, i^2, i^3 and i^4. Thus the four non-central moments μ'_1, μ'_2, μ'_3 and μ'_4 are calculated as follows:

$$\mu_1' = \frac{\Sigma fd'}{N} \times i; \qquad \mu_2' = \frac{\Sigma fd'^2}{N} \times i^2;$$

$$\mu_3' = \frac{\Sigma fd'^3}{N} \times i^3; \qquad \mu_4' = \frac{\Sigma fd'^4}{N} \times i^4.$$

Where, $d' = \dfrac{X - A}{i}$, i = common factor, X = values or mid-values of the X-column.

(3) Moments about Zero or Origin

The moments about zero or origin are denoted by Greek symbol ν(read as nu). $X_1, X_2, \ldots X_n$ be the values of a variable X, then the rth moment about zero is defined and given by

$$\nu_r = \frac{\Sigma(X-0)^r}{N} = \frac{\Sigma X^r}{N}$$

For a frequency distribution, rth moment about zero is defined by

$$\nu_r = \Sigma f(X-0)^r = \frac{\Sigma fX^r}{N}$$

Putting r = 1, 2, 3 and 4, the various moment about zero are as follows:

Individual Series	Discrete/Continuous Series
$\nu_1 = \dfrac{\Sigma(X-0)^1}{N} = \dfrac{\Sigma X^1}{N} = \overline{X}$	$\nu_1 = \dfrac{\Sigma fX}{N} = \overline{X}$
$\nu_2 = \dfrac{\Sigma(X-0)^2}{N} = \dfrac{\Sigma X^2}{N}$	$\nu_2 = \dfrac{\Sigma fX^2}{N}$
$\nu_3 = \dfrac{\Sigma(X-0)^3}{N} = \dfrac{\Sigma X^3}{N}$	$\nu_3 = \dfrac{\Sigma fX^3}{N}$
$\nu_4 = \dfrac{\Sigma(X-0)^4}{N} = \dfrac{\Sigma X^4}{N}$	$\nu_4 = \dfrac{\Sigma fX^4}{N}$

Moments about zero can be extended to higher powers but in practice the first four moments are computed because of the difficulty of computation.

Conversion of Non-Central Moments including Zero into Central Moments

The central moments can be easily computed from the moments about the assumed mean including zero by using the following relations:

Using Moments about Assumed Mean	Using Moments about Origin
$\mu_1 = \mu_1' - \mu_1' = 0$ (Always)	$\mu_1 = 0$
$\mu_2 = \mu_2' - (\mu_1')^2$	$\mu_2 = \nu_2 - \nu_1^2$
$\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2(\mu_1')^3$	$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3$
$\mu_4 = \mu_4' - 4\mu_3' \cdot \mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$	$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2 \cdot \nu_1^2 - 3\nu_1^4$
$\overline{X} = A + \mu_1'$;	$\sigma^2 = \mu_2' - \mu_1'^2$

Conversion of Central Moments into Non-Central Moments including Zero

The moments about any value 'A' (including zero) can be easily computed from the central moments by using the following relations:

Moments about any Value 'A' from Central Moments	Moments about Zero from Central Moments
$\mu'_1 = \bar{X} - A$	$v_1 = A + \mu'_1$ or $\bar{X}$
$\mu'_2 = \mu_2 + (\mu'_1)^2$	$v_2 = \mu_2 + v_1^2$
$\mu'_3 = \mu_3 + 3\mu'_2.\mu'_1 - 2(\mu'_1)^3$	$v_3 = \mu_3 + 3v_2.v_1 - 2v_1^3$
$\mu'_4 = \mu_4 + 4\mu'_3.\mu'_1 - 6\mu'_2(\mu'_1)^2 + 3(\mu'_1)^4$	$v_4 = \mu_4 + 4v_3.v_1 - 6v_2.v_1^2 + 3v_1^4$

Note: 1. The signs are reverse of what we had while converting moments about assumed mean into central moments.

2. It is necessary to find $\bar{X}$ for converting central moments into non-central moments.

UTILITY OF MOMENTS

Moments are useful in analysing the different aspects of frequency distribution. With the help of moments we can measure the central tendency of a set of observations, their variability, their asymmetry and the height of the peak their curves would make. The following is the summary of how moments help in analysing a frequency distribution:

	Moments	What it measures
1.	First moment about origin or zero (v_1)	Mean
2.	Second moment about the mean (μ_2)	Variance
3.	Second and third moments about the mean (μ_2 and μ_3)	Skewness
4.	Second and fourth moments about the mean (μ_2 and μ_4)	Kurtosis

Example 3. Calculate the first four moments about mean from the following distribution:

X :	1	2	3	4	5	6	7
f :	2	9	25	35	20	8	1

Solution: We shall first determine moments about assumed mean, then calculate the central moments using the appropriate formula.

Calculations of Moments

X	f	$d = X - A$ $A = 4$	fd	fd^2	fd^3	fd^4
1	2	–3	–6	18	–54	162
2	9	–2	–18	36	–72	144
3	25	–1	–25	25	–25	25
4 A	35	0	0	0	0	0
5	20	+1	20	20	20	20
6	8	+2	16	32	64	128
7	1	+3	3	9	27	81
	$N = 100$		$\Sigma fd = -10$	$\Sigma fd^2 = 140$	$\Sigma fd^3 = -40$	$\Sigma fd^4 = 560$

$$\mu_1' = \frac{\Sigma fd'^1}{N} = \frac{-10}{320} = -0.1 \qquad \mu_2' = \frac{\Sigma fd'^2}{N} = \frac{140}{100} = 1.4$$

$$\mu_3' = \frac{\Sigma fd'^3}{N} = \frac{-40}{100} = -0.4 \qquad \mu_4' = \frac{\Sigma fd'^4}{N} = \frac{560}{100} = 5.6$$

Now, we convert moments about assumed mean into central moments by using the formula

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 1.4 - (-0.1)^2 = 1.4 - 0.01 = 1.39$$

$$\mu_3 = \mu_3' - 3\mu_2'.\mu_1' + 2\mu_1'^3 = (-0.4) - 3 \times 1.4 \times (-0.1) + 2(-0.1)^3$$

$$= -0.4 + 0.42 - 0.002 = 0.018$$

$$\mu_4 = \mu_4' - 4\mu_3'.\mu_1' + 6\mu_2'.\mu_1'^2 - 3\mu_1'^4$$

$$= 5.6 - 4 \times (-0.4) \times (-0.1) + 6 \times 1.4 \times (-0.1)^2 - 3(-0.1)^4$$

$$= 5.6 - 0.16 + 0.084 - 0.0003 = 5.5237$$

Example 4. Calculate the first four moments about mean for the following distribution:

X:	2.0	2.5	3.0	3.5	4.0	4.5	5.0
f:	5	38	65	92	70	40	10

Solution: We shall first determine moments about assumed mean and then calculate the central moments using the appropriate formula.

Calculation of Moments

X	f	$d = X - A$ $A = 3.5$	$d' = d/.5$	fd'	fd'^2	fd'^3	fd'^4
2.0	5	−1.5	−3	−15	45	−135	405
2.5	38	−1.0	−2	−76	152	−304	608
3.0	65	−0.5	−1	−65	65	−65	65
3.5A	92	0	0	0	0	0	0
4.0	70	+0.5	+1	70	70	70	70
4.5	40	+1	+2	80	160	320	640
5.0	10	+1.5	+3	30	90	270	810
	N = 320			$\Sigma fd' = 24$	$\Sigma fd'^2 = 582$	$\Sigma fd'^3 = 156$	$\Sigma fd'^4 =$ 2[illegible]

$$\mu_1' = \frac{\Sigma fd'}{N} \times i = \frac{24}{320} \times (0.5) = 0.0375$$

$$\mu_2' = \frac{\Sigma fd'^2}{N} \times i^2 = \frac{582}{320} \times (0.5)^2 = 0.4547$$

$$\mu_3' = \frac{\Sigma fd'^3}{N} \times i^3 = \frac{156}{320} \times (0.5)^3 = 0.0609$$

$$\mu_4' = \frac{\Sigma fd'^4}{N} \times i^4 = \frac{2598}{320} \times (0.5)^4 = 0.5074$$

Now we convert moments about assumed mean into central moments by using the formula

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 0.4547 - (0.0375)^2 = 0.4534$$

$$\mu_3 = \mu_3' - 3\mu_2'.\mu_1' + 2\mu_1'^3 = 0.0609 - 3(0.4547)(0.0375) + 2(0.0375)^3 = 0.0099$$

$$\mu_4 = \mu_4' - 4\mu_3'.\mu_1' + 6\mu_2'.\mu_1'^2 - 3\mu_1'^4$$

$$= 0.5074 - 4(0.0609)(0.0375) + 6(0.454)(0.0375)^2 - 3(.0375)^4 = 0.5021$$

Example 5. Calculate first four central mements from the following distribution:

Height (in inches):	60—62	63—65	66—68	69—71	72—74
Frequency:	5	18	42	27	8

Solution: We shall first determine moments about assumed mean and then calculate the central moments using the appropriate formula.

Calculation of Moments

Height (X)	f	M.V. (m)	d	$d' = d/3$	fd'	fd'^2	fd'^3	fd'^4
60—62	5	61	–6	–2	–10	20	–40	80
63—65	18	64	–3	–1	–18	18	–18	18
66—68	42	67 = A	0	0	0	0	0	0
69—71	27	70	+3	+1	+27	27	27	27
72—74	8	73	+6	+2	+16	32	64	128
	N=100				$\Sigma fd' = 15$	$\Sigma fd'^2 = 97$	$\Sigma fd'^3 = 33$	$\Sigma fd'^4 = 253$

$$\mu_1' = \frac{\Sigma fd'}{N} \times i = \frac{15}{100} \times 3 = 0.45$$

$$\mu_2' = \frac{\Sigma fd'^2}{N} \times i^2 = \frac{97}{100} \times 9 = 8.73$$

$$\mu_3' = \frac{\Sigma fd'^3}{N} \times i^3 = \frac{33}{100} \times 27 = 8.91$$

$$\mu_4' = \frac{\Sigma fd'^4}{N} \times i^4 = \frac{253}{100} \times 81 = 204.93$$

Now, we convert moments about assumed mean into moments about mean by using the formula.

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 8.73 - (0.45)^2 = 8.73 - 0.20 = 8.53$$

$$\mu_3 = \mu_3' - 3\mu_2'.\mu_1' + 2(\mu_1')^3 = 8.91 - 3(8.73)(0.45) + 2(0.45)^3$$

$$= 8.91 - 11.79 + 0.18 = -2.70$$

$$\mu_4 = \mu_4' - 4\mu_3' \cdot \mu_1' + 6\mu_2' \cdot (\mu_1')^2 - 3(\mu_1')^4$$
$$= 204.93 - 4\ (8.91)\ (0.45)\ + 6\ (8.73)\ (0.45)^2 - 3\ (0.45)^4$$
$$= 204.93 - 16.04 + 6\ (8.73)\ (0.25) - 3\ (0.04)$$
$$= 204.93 - 16.04 + 10.61 - 0.12 = 199.38$$

Example 6. The first four moments of a distribution about $x = 2$ are: 1, 2.5, 5.5 and 16. Calculate the four moments about $\overline{X}$ and about zero.

Solution: We are given $A = 2, \mu_1' = 1, \mu_2' = 2.5, \mu_3' = 5.5$ and $\mu_4' = 16$. From these moments about assumed mean (2), we can find out moments about mean with the help of the following formulae:

Moments about mean:

$$\mu_1 = 0$$
$$\mu_2 = \mu_2' - (\mu_1')^2 = 2.5 - (1)^2 = 1.5$$
$$\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2\mu_1'^3 = 5.5 - 3(2.5)(1) + 2(1)^3 = 5.5 - 7.5 + 2 = 0$$
$$\mu_4 = \mu_4' - 4\mu_3' \cdot \mu_1' + 6\mu_2' \cdot (\mu_1')^2 - 3(\mu_1')^4$$
$$= 16 - 4(5.5)(1) + 6(2.5)(1)^2 - 3(1)^4 = 16 - 22 + 15 - 3 = 6$$

Thus, moments about mean are $\mu_1 = 0, \mu_2 = 1.5, \mu_3 = 0, \mu_4 = 6$

Moments about zero:

$$\overline{X} = A + \mu_1' = 2 + 1 = 3$$
$$v_1 = \overline{X} = 3$$
$$v_2 = \mu_2 + v_1^2 = 1.5 + (3)^2 = 10.5$$
$$v_3 = \mu_3 + 3v_2 . v_1 - 2v_1^3 = 0 + 3(10.5)(3) - 2(3)^3 = 40.5$$
$$v_4 = \mu_4 + 4v_3 . v_1 - 6v_2 . v_1^2 + 3v_1^4$$
$$= 6 + 4(40.5)\ (3) - 6(10.5)(3)^2 + 3(3)^4 = 6 + 486 - 567 + 243 = 168$$

Thus, moments about zero are $v_1 = 3, v_2 = 10.5, v_3 = 40.5, v_4 = 168$

Example 7. The arithmetic mean of a series is 22 and first four central moments are 0, 81, – and 14817. Find the first four moments (*i*) about the assumed mean '25' and (*ii*) a origin or zero.

Solution: Given $\overline{X} = 22$, $\mu_1 = 0, \mu_2 = 81, \mu_3 = -144$ and $\mu_4 = 14817$

(*i*) About Assumed mean '25' (A = 25)

$$\mu_1' = \overline{X} - A = 22 - 25 = -3$$
$$\mu_2' = \mu_2 + (\mu_1')^2 = 81 + (-3)^2 = 81 + 9 = 90$$
$$\mu_3' = \mu_3 + 3\mu_2' . \mu_1' - 2(\mu_1')^3$$
$$= -144 + 3(90)\ (-3) - 2(-3)^3$$
$$= -144 - 810 + 54 = -900$$

$$\mu_4' = \mu_4 + 4\mu_3'.\mu_1' - 6\mu_2'.\mu_1'^2 + 3(\mu_1')^4$$
$$= 14817 + 4(-900)(-3) - 6(90)(-3)^2 + 3(-3)^4$$
$$= 14817 + 10800 - 4860 + 243 = 21,000$$

(ii) About origin zero ($A = 0$)

$$v_1 = \overline{X} = 22$$
$$v_2 = \mu_2 + v_1^2 = 81 + (22)^2 = 81 + 484 = 565$$
$$v_3 = \mu_3 + 3v_2.v_1 - 2v_1^3 = -144 + 3(565)(22) - 2(22)^3$$
$$= -144 + 37290 - 21296 = 15850$$
$$v_4 = \mu_4 + 4v_3.v_1 - 6v_2.v_1^2 + 3v_1^4$$
$$= 14817 + 4(15850)(22) - 6(565)(22)^2 + 3(22)^4$$
$$= 14817 + 1394800 - 1640760 + 702768 = 4,71,625$$

SHEPPARD CORRECTIONS FOR GROUPING ERRORS IN MOMENTS

In computing various moments in case of grouped data, it is assumed that the values of all items lying in a class are concentrated at the mid-point of the class. This assumption leads to grouping error in finding the values of the moments. This error is corrected by famous mathematician Sheppard, and therefore, called Sheppard's Corrections. According to Sheppard, first (μ_1) and third (μ_3) moments need no corrections. He has suggested the following formulae for correcting the second (μ_2) and the fourth (μ_4) moments which he regards as crude moments liable to be affected by the grouping error of a continuous series.

$$\mu_2 \text{ (corrected)} = \mu_2(\textit{uncorrected}) - \frac{i^2}{12}$$

$$\mu_4 \text{ (corrected)} = \mu_4(\textit{uncorrected}) - \frac{1}{2}i^2\ \mu_2\ (\textit{uncorrected}) + \frac{7}{240}i^4$$

Where, i = width of class interval.

The first and third moments need no correction.

The following conditions should be satisfied for the application of Sheppard's corrections:

(i) The correction should not be made unless the frequency is at least 1000 otherwise the moments will be more affected by sampling errors than by grouping errors.

(ii) The correction is not applicable to J-or U-shaped distributions or even to the skew form.

(iii) The observations should relate to a continuous variable.

(iv) The curve should approach the base line gradually and slowly at each end of the distribution.

Example 8. The first four central moments of a continuous series with class intervals of 3 are arrived at 0, 43.353, –9.774 and 5508.567. Find their corrected values using Sheppard's corrections.

Solution: According to Sheppard, the first and third moments about mean need no correction. Hence, the 2nd and 4th moments only are corrected as follows:

We are given, $\mu_2 = 43.353$ and $\mu_4 = 5508.567$ and $i = 3$

We have,

$$\mu_2(\text{corrected}) = \mu_2 - \frac{i^2}{12} = 43.353 - \frac{(3)^2}{12} = 43.353 - 0.75 = 42.603$$

$$\mu_4\,(\text{corrected}) = \mu_4 - \frac{1}{2}i^2, \mu_2 + \frac{7}{240}i^4 = 5508.567 - \frac{1}{2}(3)^2(43.353) + \frac{7}{240}(3)$$

$$= 5508.567 - 195.0885 + 2.3625 = 5315.841$$

BETA AND GAMMA COEFFICIENTS (OR BETA AND GAMMA MEASURES) BASED ON MOMENTS

Karl Pearson has developed Beta and Gamma Coefficients (or Beta and Gamma Measures) based on the central moments which are given below:

Beta Coefficients or Beta Measures	Gamma Coefficients or Gamma Measures
$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$	$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}}$
$\sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}}$	$\gamma_2 = \beta_2 - 3$ or $\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$
$\beta_2 = \frac{\mu_4}{\mu_2^2}$	

Note: These coefficients are used in the calculation of skewness and kurtosis. It needs to be mentioned here that the above coefficients are pure numbers and independent of the units of measurements.

Example 9. The first four central moments are: 0, 4, 8 and 144. Find β and γ coefficients.

Solution: We are given: $\mu_1 = 0, \mu_2 = 4, \mu_3 = 8, \mu_4 = 144$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(8)^2}{(4)^3} = \frac{64}{64} = 1 \qquad \therefore \qquad \beta_1 = 1$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{144}{(4)^2} = \frac{144}{16} = 9 \qquad \therefore \qquad \beta_2 = 9$$

$$\gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{8}{\sqrt{(4)^3}} = \frac{8}{\sqrt{64}} = \frac{8}{8} = 1$$

$$\gamma_2 = \beta_2 - 3 = 9 - 3 = 6$$

MEASURE OF SKEWNESS BASED ON CENTRAL MOMENTS

A mesure of skewness may be obtained by making use of the second and third central moments. Skewness is measured by β_1 coefficient (read as β_1 coefficient) which is defined and given by:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(\text{Third Central Moment})^2}{(\text{Second Central Moment})^3}$$

In a symmetrical distribution, β_1 shall be zero. The greater the values of β_1, the more skewed the distribution. But β_1 as a measure of skewness cannot tell us about the direction of skewness, *i.e.*, whether it is positive or negative. Therefore, instead of β_1, sometimes $\sqrt{\beta_1}$ is used as a measure of skewness. It is obtained as:

$$\sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}} \qquad \text{where } \sqrt{\beta_1} = \text{Moment Coefficient of Skewness}$$

Interpretation

The value of $\sqrt{\beta_1}$ is interpreted as follows:

(a) If $\sqrt{\beta_1} = 0$, there is no skewness, *i.e.*, the distribution is symmetric.

(b) If $\sqrt{\beta_1} > 0$, there is positive skewness, *i.e.*, the distribution is positively skewed.

(c) If $\sqrt{\beta_1} < 0$, there is negative skewness, *i.e.*, the distribution is negatively skewed.

Example 10. The first three central moments of a distribution are: 0, 2.5, 0.7. Find the moment coefficient of skewness.

Solution: We are given: $\mu_1 = 0, \mu_2 = 2.5, \mu_3 = 0.7$

$$\text{Moment coefficient of skewness} = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{0.7}{\sqrt{(2.5)^3}} = \frac{0.7}{\sqrt{15.625}} = \frac{0.7}{3.953} = 0.177$$

Example 11. The first four moments of a distribution about the value 5 of the variable are 2, 20, 40 and 50. Calculate the moment coefficient of skewness.

Solution: We are given: $A = 5, \mu_1' = 2, \mu_2' = 20, \mu_3' = 40, \mu_4' = 50$

$$\mu_1 = 0,\ \mu_2 = \mu_2' - (\mu_1')^2 = 20 - (2)^2 = 16$$

$$\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2\mu_1'^3 = 40 - 3(2)(20) + 2(2)^3 = -64$$

$$\text{Moment coefficient of skewness} = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}} = \frac{-64}{\sqrt{(16)^3}} = -1$$

EXERCISE 8.1

1. Calculate the first four moments about mean for the following data:

X:	3	6	8	10	18

[**Ans.** $\mu_1 = 0, \mu_2 = 25.6, \mu_3 = 97.2, \mu_4 = 1588$]

2. Calculate the first four moments about mean for the following data:

X :	2	3	4	5	6
f:	1	3	7	3	1

[**Ans.** $\mu_1 = 0, \mu_2 = 0.933, \mu_3 = 0, \mu_4 = 2.530$]

3. Calculate the first four central moments from the following data and also make Sheppards' corrections:

Variable :	0—10	10—20	20—30	30—40
Frequency :	1	3	4	2

[Ans. $\mu_1 = 0, \mu_2 = 81, \mu_3 = -144, \mu_4 = 14{,}817$, μ_2 (corrected) = 71, μ_4 (corrected) = 11058.67]

4. Calculate the first four moments about the mean from the following data and also find the value of β_1 and β_2:

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

[Ans. $\mu_1 = 0$, $\mu_2 = 177.39$, $\mu_3 = 47.982$, $\mu_4 = 95009.364$; $\beta_1 = 0.0004$, $\beta_2 = 3.02$]

5. The first four moments of a distribution about the value A=5 are –2, 10, –25 and 50. Find the first four about $\overline{X}$ and about zero.

[Ans. $\mu_1 = 0$, $\mu_2 = 6$, $\mu_3 = 19$, $\mu_4 = 42$; $\nu_1 = 1$, $\nu_2 = 7$, $\nu_3 = 38$, $\nu_4 = 155$]

6. The arithmetic mean of a series is 5 and the first four central moments are 0, 3, 0 and 26. Find the four moments (*i*) based on assumed mean '4' and (*ii*) based on zero.

[Ans. (*i*) 1, 4, 10 and 45 (*ii*) 5, 28, 170, 1101]

7. Examine whether the following results for obtaining 2nd order central moments are consistent or not: $N = 50$, $\Sigma X = 100$, $\Sigma X^2 = 160$.
[**Hint:** See Example 27]

[Ans. Inconsistent]

8. If the first three moments about origin for distribution are 10, 225 and 0 respectively, calculate the first three moments about value '5' for the distribution.

[Ans. $\mu_1' = 5, \mu_2' = 150, \mu_3' = -2750$]

9. The first four central moments of a distribution are 0, 2.5, 0.7 and 18.75. Test the skewness of the distribution.

[Ans. $\beta_1 = 0.31$, the distribution is slightly skewed]

10. The first four moments of a distribution about value 2 are 1, 2.5, 5.5 and 16 respectively. Calculate the four moments about mean and comment on the nature of distribution.

[Ans. $\mu_1 = 0$, $\mu_2 = 1.5$, $\mu_3 = 0$, $\mu_4 = 6$, $\beta_1 = 0$, symmetrical, $\beta_2 = 2.67$, platy-kurtic]

11. The first four central moments of a continous series with class intervals of 6 are arrived at 0, –60, 900 and –9500. Find their corrected values accending to Sheppard's corrections.

[Ans. – 63, –8382.2]

KURTOSIS

Kurtosis is a Greek word meaning bulkiness. In statistics, it refers to the degree of flatness or peakedness of a frequency curve. The degree of kurtosis (or peakedness) of a distribution is measured relative to the peakedness of the normal curve. To quote **M.R. Speigal,** *"Kurtosis is the degree of peakedness of a distribution, usually taken relative to a normal distribution"*. According to **Croxten and Cowden** *"A measure of kurtosis indicates the degree to which a frequency distribution is peaked or flat-topped". Thus, a measure of kurtosis tells us the extent to which a distribution is more peaked or flat-topped than the normal curve.*

Types of Kurtosis

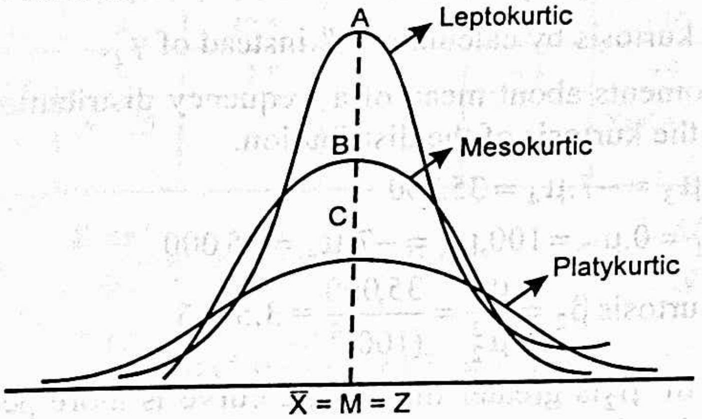
There are three types of kurtosis in a distribution:

(1) Lepto-kurtic: A curve having a high peak than the normal curve is called lepto-kurtic. In such a curve, there is too much concentration of the items near the centre.

(2) Platy-kurtic: A curve having a low peak (or flat topped) than the normal curve is called platy-kurtic. In such a curve, there is less concentration of items near the centre.

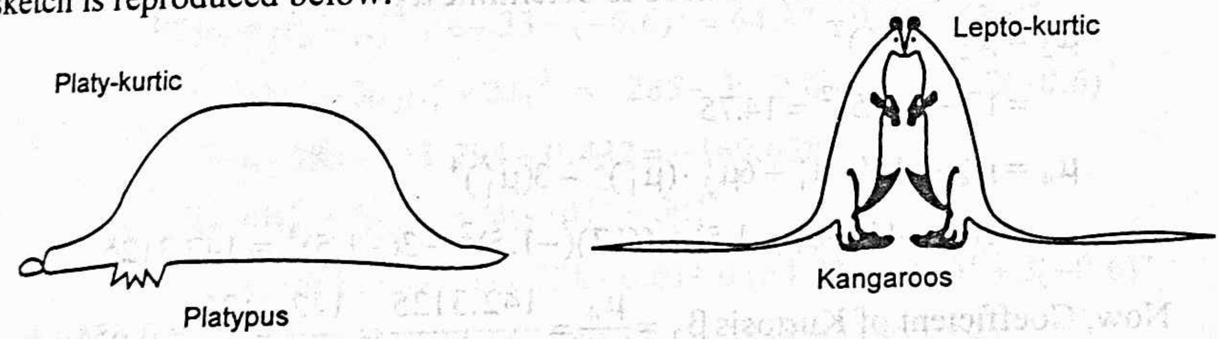
(3) Meso-kurtic: A curve having normal peak or the normal curve itself is called meso-kurtic. In such a curve, there is equal distribution of items around the central value.

The figure below illustrates three different curves explained above:



(A) Lepto-kurtic, (B) Meso-kurtic, (C) Platy-kurtic

A famous British statistician **William Gosset** (known as "Student") has very humorously described the nature of the curves in these words **"platy-kurtic curves are squat with short tails, like the platyplus, lepto-kurtic curves are high with long tails like the Kangaroos"**. Gosset's little sketch is reproduced below:



Measures of Kurtosis

Kurtosis is measured by β_2(read as beta two) which is defined and given by:

$$\text{Measure of kurtosis } \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{\text{Fourth Central Moment}}{(\text{Second Central Moment})^2}$$

Interpretation

The value of β_2 is interpreted as follows:

(*i*) If $\beta_2 > 3$, the curve is more peaked than the normal curve, *i.e.*, lepto-kurtic.

(*ii*) If $\beta_2 < 3$, the curve is less peaked than the normal curve, *i.e.*, platy-kurtic.

(*iii*) If $\beta_2 = 3$, the curve is having moderate peak, *i.e.*, meso-kurtic.

▶ Alternative Measure

Sometimes, the Kurtosis is measured by γ_2(read as Gamma two) which is defined and given by

$$\gamma_2 = \beta_2 - 3 = \frac{\mu_4}{\mu_2^2} - 3$$

▶ Interpretation

The value of γ_2 is interpreted as follows:

(*i*) If γ_2 or $\beta_2 - 3 = 0$, the curve is meso-kurtic

(*ii*) If γ_2 or $\beta_2 - 3 > 0$ the curve is lepto-kurtic

(*iii*) If γ_2 or $\beta_2 - 3 < 0$ the curve is platy-kurtic.

Note: It is easier to interpret kurtosis by calculating β_2 instead of γ_2.

Example 12. The first four moments about mean of a frequency distribution are 0, 100, −7 and 35,000. Discuss the kurtosis of the distribution.

$\mu_1 = 0, \mu_2 = 100, \mu_3 = -7, \mu_4 = 35{,}000$

Solution: We are given: $\mu_1 = 0, \mu_2 = 100, \mu_3 = -7, \mu_4 = 35{,}000$

$$\text{Coefficient of Kurtosis } \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{35{,}000}{(100)^2} = 3.5 > 3.$$

Since the value of β_2 is greater than 3, the curve is more peaked than the normal curve, *i.e.*, lepto-kurtic.

Example 13. The first four moments of a distribution about the value '4' of the variable are −1.5, 17, −30 and 108. Discuss the kurtosis of the distribution.

Solution: We are given: $A = 4, \mu_1' = -1.5, \mu_2' = 17, \mu_3' = -30, \mu_4' = 108$

For determining kurtosis, we need to determine μ_2 and μ_4.

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

$$= 17 - (-1.5)^2 = 14.75$$

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

$$= 108 - 4(-30)(-1.5) + 6(17)(-1.5)^2 - 3(-1.5)^4 = 142.3125$$

$$\text{Now, Coefficient of Kurtosis } \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{142.3125}{(14.75)^2} = \frac{142.3125}{217.5625} = 0.654 < 3$$

Since, $\beta_2 < 3$, the distribution is platy-kurtic.

Example 14. Calculate first four central moments and coefficient of kurtosis for the following distribution and comment on the result.:

Variable:	0—5	5—10	10—15	15—20	20—25	25—30	30—35	35—40
Frequency:	2	5	7	13	21	16	8	3

Solution: We shall first determine moments about assumed mean, and then calculate the central moments using the appropriate formulae:

Calculation of Moments

Variable	f	M.V. (m)	$d = X - A$	$d' = d/5$	fd'	fd'^2	fd'^3	fd'^4
0—5	2	2.5	–20	–4	–8	32	–128	512
5—10	5	7.5	–15	–3	–15	45	–135	405
10—15	7	12.5	–10	–2	–14	28	–56	112
15—20	13	17.5	–5	–1	–13	13	–13	13
20—25	21	22.5 = A	0	0	0	0	0	0
25—30	16	27.5	+5	+1	+16	16	+16	16
30—35	8	32.5	+10	+2	+16	32	+64	128
35—40	3	37.5	+15	+3	+9	27	+81	243
	N=75				$\Sigma fd' = -9$	$\Sigma fd'^2$ =193	$\Sigma fd'^3$ =–171	$\Sigma fd'^4$ =1429

$$\mu_1' = \frac{\Sigma fd'}{N} \times i = \frac{-9}{75} \times 5 = -0.6 \qquad \mu_2' = \frac{\Sigma fd'^2}{N} \times i^2 = \frac{193}{75} \times 25 = 64.33$$

$$\mu_3' = \frac{\Sigma fd'^3}{N} \times i^3 = \frac{-171}{75} \times 125 = -285$$

$$\mu_4' = \frac{\Sigma fd'^4}{N} \times i^4 = \frac{1429}{75} \times 625 = 11908.33$$

Using moments about assumed mean, central moments are calculated as:

$$\mu_1 = 0 \text{ (Always)}$$

$$\mu_2 = \mu_2' - \mu_1'^2 = 64.33 - (-0.6)^2 = 64.33 - 0.36 = 63.97$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2\mu_1'^3 = -285 - 3(-0.6)(64.33) + 2(-0.6)^3$$

$$= -285 + 115.794 - 0.432 = -169.638$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2' \cdot \mu_1'^2 - 3\mu_1'^4$$

$$= 11908.33 - 4(-285)(-0.6) + 6(64.33)(-0.6)^2 + 3(-0.6)^4$$

$$= 11908.33 - 684 + 138.953 - 0.3888 = 11362.895$$

$$\text{Coefficient of kurtosis } \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{11362.895}{(63.97)^2} = \frac{11362.895}{4092.1609} = 2.776$$

Since $\beta_2 < 3$, the distribution is platy-kurtic.

Example 15. The standard deviation of a symmetrical distribution is 3. What must be the value of the fourth moment about the mean in order that the distribution be meso-kurtic?

Solution: For a meso-kurtic distribution $\beta_2 = 3$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \qquad \text{...}(i)$$

We are given: $\sigma = 3$

$$\mu_2 = \sigma^2 = (3)^2 = 9$$

Thus, $\beta_2 = 3$, $\mu_2 = 9$

Putting the value in (*i*)

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \Rightarrow 3 = \frac{\mu_4}{9^2} \text{ or } \mu_4 = 243$$

Thus, the fourth moment about the mean must be 243 in order that distribution be meso-kurtic.

Example 16. The following data are given to an economist for the purpose of economic analysis. The data refers to the length of a certain type of batteries:

$N = 100,\ \Sigma fd = 50,\ \Sigma fd^2 = 1970, \Sigma fd^3 = 2948$

and $\Sigma fd^4 = 86{,}752$ in which $d = X - 48$

Do you think that the distribution is platy-kurtic?

Solution: Given, $N = 100, \Sigma fd = 50, \Sigma fd^2 = 1970, \Sigma fd^3 = 2948$ and $\Sigma fd^4 = 86{,}752$.

We shall first determine moments about assumed mean, then calculate the central moments using the appropriate formula:

$$\mu_1' = \frac{\Sigma fd}{N} = \frac{50}{100} = 0.50; \qquad \mu_2' = \frac{\Sigma fd^2}{N} = \frac{1970}{100} = 19.70$$

$$\mu_3' = \frac{\Sigma fd^3}{N} = \frac{2948}{100} = 29.48; \qquad \mu_4' = \frac{\Sigma fd^4}{N} = \frac{86752}{100} = 867.52$$

Using moments about assumed mean, the central moments are:

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 19.70 - (0.50)^2 = 19.45$$

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

$$= 867.52 - 4(29.48)(0.50) + 6(19.70)(0.5)^2 - 3(0.5)^4$$

$$= 867.52 - 58.96 + 29.55 - 0.1875 = 837.9225$$

$$\text{Coefficient of Kurtosis } \beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{837.9225}{(19.45)^2} = \frac{837.9225}{378.3025} = 2.215 < 3$$

Since $\beta_2 < 3$, the distribution is platy-kurtic.

COMBINED EXAMPLES ON SKEWNESS AND KURTOSIS

Example 17. Compute the coefficient of skewness (β_1) and coefficient of kurtosis (β_2) based on moments for the following data:

Age:	25—30	30—35	35—40	40—45	45—50	50—55	55—60	60—65
Frequency:	2	8	18	27	25	16	7	2

Solution:

Calculation of Moments

Age	f	M.V.	d	$d' = \frac{d}{5}$	fd'	fd'^2	fd'^3	fd'^4
25—30	2	27.5	–15	–3	–6	18	–54	162
30—35	8	32.5	–10	–2	–16	32	–64	128
35—40	18	37.5	–5	–1	–18	18	–18	18
40—45	27	42.5 = A	0	0	0	0	0	0
45—50	25	47.5	+5	+1	+25	25	+25	25
50—55	16	52.9	+10	+2	+32	64	+128	256
55—60	7	57.5	+15	+3	+21	63	+189	567
60—65	2	62.5	+20	+4	+8	32	+128	512
	N=105				$\Sigma fd'$ = 46	$\Sigma fd'^2$ =252	$\Sigma fd'^3$ =334	$\Sigma fd'^4$ =1668

$$\mu_1' = \frac{\Sigma fd'}{N} \times i = \frac{46}{105} \times 5 = 2.19, \qquad \mu_2' = \frac{\Sigma fd'^2}{N} \times i^2 = \frac{252}{105} \times 5^2 = 60$$

$$\mu_3' = \frac{\Sigma fd'^3}{N} \times i^3 = \frac{334}{105} \times 5^3 = 397.625 \qquad \mu_4' = \frac{\Sigma fd'^4}{N} \times i^4 = \frac{1668}{105} \times 5^4 = 9892.85$$

Moments about Mean

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 60 - (2.19)^2 = 55.20$$

$$\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2\mu_1'^3 = 397.625 - 3(60)(2.19) + 2(2.19)^3 = 24.43$$

$$\mu_4' = \mu_4' - 4\mu_3' \cdot \mu_1' + 6\mu_2' \cdot \mu_1'^2 - 3\mu_1'^4$$

$$= 9892.85 - 4\,(397.625)(2.19) + 6(60)(2.19)^2 - 3(2.19)^4$$

$$= 9892.85 - 3483.195 + 1726.596 - 69.007 = 8067.25$$

Now, Moment Coefficient of Skewness $(\beta_1) = \frac{\mu_3^2}{\mu_2^3} = \frac{(24.43)^2}{(55.20)^3} = 0.0035$

Moment Coefficient of Kurtosis $(\beta_2) = \frac{\mu_4}{\mu_2^2} = \frac{8067.25}{(55.20)^2} = 2.65 < 3$

Example 18. The first four central moments of a distribution are 0, 2.5, 0.7 and 18.75. Test the skewness and kurtosis of the distribution.

Solution: Given $\mu_1 = 0, \mu_2 = 2.5, \mu_3 = 0.7$ and $\mu_4 = 18.75$.

Testing Skewness

Skewness is measured by the coefficient β_1 which is defined as:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(0.7)^2}{(2.5)^3} = 0.031$$

Since, $\beta_1 = 0.031$, the distribution is slightly skewed, *i.e.*, it is not perfectly symmetrical.

Testing Kurtosis

For testing kurtosis, we compute the value of β_2 which is defined as

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{18.75}{(2.5)^2} = \frac{18.75}{6.25} = 3$$

Since β_2 is exactly 3, the distribution is meso-kurtic.

Example 19. The first four moments of a distribution about the value 4 are 1, 4, 10 and 45. Obtain a measure of skewness and kurtosis.

Solution: We are given $A = 4, \mu_1' = 1, \mu_2' = 4, \mu_3' = 10$ and $\mu_4' = 45$

Moments about Mean

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 4 - (1)^2 = 3$$

$$\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2\mu_1'^3 = 10 - 3(4)(1) + 2(1)^3 = 10 - 12 + 2 = 0$$

$$\mu_4 = \mu_4' - 4\mu_3' \cdot \mu_1' + 6\mu_2' \cdot \mu_1'^2 - 3\mu_1'^4$$

$$= 45 - 4(10)(1) + 6(4)(1)^2 - 3(1)^4 = 45 - 40 + 24 - 3 = 26$$

Measure of Skewness

Skewness is measured by the coefficient β_1 which is defined as

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0}{(3)^3} = 0$$

Since $\beta_1 = 0$, the distribution is symmetrical.

Measures of Kurtosis

Kurtosis is measured by the coefficient β_2 which is defined as

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{26}{(3)^2} = 2.89$$

Since $\beta_2 < 3$, the distribution is platy-kurtic.

EXERCISE 8.2

1. The first four moments of a distribution about $x = 4$ are –1.5, 17, –30, 108. Discuss the kurtosis of the distribution. **[Ans.** $\beta_1 = 0.65$, platy-kurtic]
2. The first four central moments of a distribution are 0, 19.67, 29.26 and 866. Test the skewness and kurtosis of the distribution. **[Ans.** $\beta_1 = 1125$, $\beta_2 = 2.238$, platy-kurtic]
3. Find a measure of kurtosis for the following distribution:

Marks:	30—35	35—40	40—45	45—50	50—55	55—60	60—65	65—70
No. of students:	5	14	16	25	14	12	8	6

[Ans. $\beta_2 = 2.34$, platy-kurtic]

4. For a meso-kurtic distribution, the first moment about 7 is 23 and second moment about origin is 1000. Find coefficient of variation and fourth moment about mean.
[Hint: See Example 29] [Ans. C.V. = 33.33, μ_4 = 30,000]

5. Analyse the frequency distribution by the method of moments.

X:	2	3	4	5	6
f:	1	3	7	3	1

[Hint: See Example 24] [Ans. $\overline{X} = 4, \sigma = 0.966, \sqrt{\beta_1} = 0,\ \beta_2 = 2.91$]

6. For the following distribution, calculate the first four central moments and two beta coefficients:

Class-interval:	20—30	30—40	40—50	50—60	60—70	70—80	80—90
Frequency:	5	14	20	25	17	11	8

[Ans. $\mu_1 = 0,\ \mu_2 = 254,\ \mu_3 = 540,\ \mu_4 = 1{,}49{,}000,\ \beta_1 = 0.177945, \beta_2 = 2.31$]

7. For a distribution it has been found that the first four moments about 27 are 0, 256, – 2871 and 1,88,462 respectively. Obtain the first four moments about zero. Also calculate the values of β_1 and β_2 and comment.
[Hint: See Example 28] [Ans. $v_1 = 27, v_2 = 985, v_3 = 37548, v_4 = 15{,}29{,}579, \beta_1 = 0.49, \beta_2 = 2.875$]

8. For a distribution mean is 10, standard deviation is 4, $\sqrt{\beta_1} = 1$ and $\beta_2 = 4$. Obtain the first four moments about origin, *i.e.*, zero.
[Hint: See Example 21] [Ans. $v_1 = 10,\ v_2 = 116,\ v_3 = 1544,\ v_4 = 23{,}184$]

MISCELLANEOUS EXAMPLES

Example 20. Calculate first four central moments from the following and also find the value of β_1 and β_2:

Sales (Rs. crores):	40—50	50—60	60—70	70—80	80—90
No. of companies:	10	25	30	23	12

Solution:

Calculations for Moments

Sales (Rs. crores)	f	Mid values m	$d = m - A$	$d' = \frac{d}{10}$	fd'	fd'^2	fd'^3	fd'^4
40—50	10	45	– 20	–2	–20	40	–80	160
50—60	25	55	– 10	–1	–25	25	–25	25
60—70	30	65 = A	0	0	0	0	0	0
70—80	23	75	10	+1	+23	23	+23	23
80—90	12	85	20	+2	+24	48	+96	192
	$N = 100$				$\Sigma fd' = 2$	$\Sigma fd'^2 = 136$	$\Sigma fd'^3 = 14$	$\Sigma fd'^4 = 400$

$$\mu_1' = \frac{\Sigma fd'}{N} \times i = \frac{2}{100} \times 10 = 0.2; \qquad \mu_2' = \frac{\Sigma fd'^2}{N} \times i^2 = \frac{136}{100} \times 100 = 136$$

$$\mu_3' = \frac{\Sigma fd'^3}{N} \times i^3 = \frac{14}{100} \times 1000 = 140;\ \mu_4' = \frac{\Sigma fd'^4}{N} \times i^4 = \frac{400}{100} \times 10{,}000 = 40{,}0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 136 - (0.2)^2 = 135.96$$

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

$$= 140 - 3(136)(0.2) + 2(0.2)^3 = 140 - 81.6 + 0.016 = 58.416$$

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

$$= 40{,}000 - 4(140)(0.2) + 6(136)(0.2)^2 - 3(0.2)^4$$

$$= 40{,}000 - 112 = 32.64 - 0.0048 = 39{,}920.64$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(58.416)^2}{(135.96)^3} = 0.0014$$

β_1 is a measure of skewness. Since the value of β_1 is very close to zero, the distributio is more or less symmetrical.

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{39{,}920.64}{(135.96)^2} = 2.16$$

β_2 is a measure of kurtosis. Since the value of β_2 is less than 3, the curve is platy-kurtic.

Example 21. For a distribution mean is 10, standard deviation is 4, $\sqrt{\beta_1} = 1$ and $\beta_2 = 4$. Obtain the first four moments about the origin.

Solution: Given: $\overline{X} = 10,\ \sigma = 4,\ \sqrt{\beta_1} = 1,\ \beta_2 = 4,$

$$\mu_2 = \sigma^2 = (4)^2 = 16$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \Rightarrow 1 = \frac{\mu_3^2}{(16)^3} \Rightarrow \mu_3^2 = 4096$$

$$\therefore \quad \mu_3 = \sqrt{4096} = 64$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \Rightarrow 4 = \frac{\mu_4}{(16)^2} \Rightarrow \mu_4 = 4 \times 256 = 1024$$

$$\therefore \quad \mu_1 = 0 \text{ (always)},\ \mu_2 = 16,\ \mu_3 = 64,\ \mu_4 = 1024$$

Moments about Zero

$$v_1 = \overline{X} \text{ or } A + \mu_1' = 10$$

$$v_2 = \mu_2 + v_1^2 = 16 + (10)^2 = 116$$

$$v_3 = \mu_3 + 3v_2 . v_1 - 2v_1^3$$

$$= 64 + 3(116)(10) - 2(10)^3 = 64 + 3480 - 2000 = 1544$$

$$v_4 = \mu_4 + 4v_3 . v_1 - 6v_2 . v_1^2 + 3v_1^4$$

$$= 1024 + 4(1544)(10) - 6(116)(10)^2 + 3(10)^4$$

$$= 1024 + 61760 - 69600 + 30000 = 23{,}184$$

Example 22. The first four moments of a distribution about $X = 4$ are 1, 4, 10 and 45. Obtain the various characteristics of the distribution on the basis of the information given. Comment upon the nature of distribution.

Solution: We are given: $A = 4, \mu_1' = 1, \mu_2' = 4, \mu_3' = 10$ and $\mu_4' = 45$

According to the formulae on moments, the different possible characteristics of the distribution will be brought as under:

(*i*) Mean of distribution $\overline{X} = A + \mu_1' = 4 + 1 = 5$

(*ii*) S.D. of the distribution or $\sigma = \sqrt{\mu_2} = \sqrt{\mu_2' - (\mu_1')^2}$

$$= \sqrt{4 - (1)^2} = \sqrt{4 - 1} = \sqrt{3} = 1.732$$

(*iii*) Variance $= \sigma^2 = \mu_2 = \mu_2' - (\mu_1')^2 = 4 - 1 = 3$

(*iv*) Coefficient of variance or C.V. $= \dfrac{\sigma}{\overline{X}} \times 100$

$$= \frac{1.732}{5} \times 100 = 34.64\%$$

(*v*) Coefficient of skewness or $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

Where, $\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2(\mu_1')^3 = 10 - 3(4)(1) + 2(1)^3 = 10 - 12 + 2 = 0$

and $\mu_2 = 3$

Thus, $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{0}{(3)^2} = \dfrac{0}{27} = 0$

Comment: As $\beta_1 = 0$, the distribution is symmetric.

(*vi*) Coefficient of kurtosis or $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

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

$$= 45 - 4(10)(1) + 6(4)(1)^2 - 3(1)^4 = 45 - 40 + 24 - 3 = 26$$

and $\mu_2 = 3$

$\therefore \beta_2 = \dfrac{\mu_4}{\mu_2^2} = \dfrac{26}{(3)^2} = 2.88$

Comment: As $\beta_2 < 3$, the distribution is platy-kurtic.

Example 23. Compute the coefficient of skewness (β_1) and kurtosis (β_2) based on moments from the following data:

X:	4.5	14.5	24.5	34.5	44.5	54.5	64.5	74.5	84.5	94.5
f:	1	5	12	22	17	9	4	3	1	1

Solution:

Calculation of Skewness and Kurtosis

X	f	$d' = \frac{X-44.5}{10}$	fd'	fd'^2	fd'^3	fd'^4
4.5	1	−4	−4	16	−64	256
14.5	5	−3	−15	45	−135	405
24.5	12	−2	−24	48	−96	192
34.5	22	−1	−22	22	−22	22
44.5	17	0	0	0	0	0
54.5	9	+1	+9	9	+9	9
64.5	4	+2	+8	16	+32	64
74.5	3	+3	+9	27	+81	243
84.5	1	+4	+4	16	+64	256
94.5	1	+5	+5	25	+125	625
	$N = 75$		$\Sigma fd' = -30$	$\Sigma fd'^2 = 224$	$\Sigma fd'^3 = -6$	$\Sigma fd'^4 = 2{,}072$

$$\mu_1' = \frac{\Sigma fd'}{N} \times i = \frac{-30}{75} \times 10 = -4; \qquad \mu_2' = \frac{\Sigma fd'^2}{N} \times i^2 = \frac{224}{75} \times 10^2 = 298.66$$

$$\mu_3' = \frac{\Sigma fd'^3}{N} \times i^3 = \frac{-6}{75} \times 10^3 = -80 \qquad \mu_4' = \frac{\Sigma fd'^4}{N} \times i^4 = \frac{2{,}072}{75} \times 10^4 = 27626.66$$

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 298.66 - (-4)^2 = 282.66$$

$$\mu_3 = \mu_3' - 3\mu_2'.\mu_1' + 2\mu_1'^3 = -80 - 3\,(298.66)(-4) + 2(-4)^3 = 3375.92$$

$$\mu_4 = \mu_4' - 4\,{}_1\mu_3'\,\mu' + 6\mu_2'\,\mu_1'^2 - 3(\mu_1')^4$$

$$= 27626.66 - 4\,(-80)(-4) + 6(298.66)(-4)^2 - 3(-4)^4$$

$$= 27626.66 - 1280 + 28671.36 - 768 = 302890.02$$

Skewness: $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3} = \dfrac{(3375.92)^2}{(282.66)^3} = 0.504$

For kurtosis we have to compute the value of β_2

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{302890.02}{(282.66)^2} = 3.79 > 3$$

Since the value of β_2 is greater than 3, the curve is more peaked than the normal curve, *i.e.*, lepto-kurtic.

Example 24. Given the frequency distribution:

X:	2	3	4	5	6
f:	1	3	7	3	1

Show that the distribution is symmetric and platy-kurtic.

Solution: For determining the symmetricity and kurtosis of the distribution we are to assess the value of β_1 and β_2 and for this, we compute the first four moments about the mean, *i.e.*, μ_1, μ_2, μ_3 and μ_4.

Calculation for Central Moments

X	f	fX	$\bar{X}=4$ $X-\bar{X}$ x	fx	fx^2	fx^3	fx^4
2	1	2	–2	–2	4	–8	16
3	3	9	–1	–3	3	–3	3
4	7	28	0	0	0	0	0
5	3	15	+1	+3	3	+3	3
6	1	6	+2	+2	4	+8	16
	$N=15$	$\Sigma fX=60$		$\Sigma fx=0$	$\Sigma fx^2=14$	$\Sigma fx^3=0$	$\Sigma fx^4=38$

We have $\bar{X}=\dfrac{\Sigma fX}{N}=\dfrac{60}{15}=4$

First four central moments are:

$$\mu_1=\frac{\Sigma fx}{N}=\frac{0}{15}=0; \qquad \mu_2=\frac{\Sigma fx^2}{N}=\frac{14}{15}=0.933$$

$$\mu_3=\frac{\Sigma fx^3}{N}=\frac{0}{15}=0; \qquad \mu_4=\frac{\Sigma fx^4}{N}=\frac{38}{15}=2.533$$

$$\beta_1=\frac{\mu_3^2}{\mu_2^3}=\frac{(0)^2}{(0.933)^2}=0$$

Since, the value of β_1 is 0, the distribution is symmetric

$$\beta_2=\frac{\mu_4}{\mu_2^2}=\frac{2.533}{(0.933)^2}=2.91$$

Since, the value of β_2 is less than 3, the distribution is platy-kurtic.

Example 25. The first four central moments of a continuous series with class intervals of 10 are arrived at 0, 20, 40 and 50 respectively. Find the corrected values of the moments according to Sheppard's corrections.

Solution: According to Sheppard, the first and third moments about the mean need no correction. Hence, the 2nd and 4th moments only are corrected as follows:

We are given, $\mu_1'=0$, $\mu_2'=20$, $\mu_3'=40$ and $\mu_4'=50$, $i=10$

We have

$$\mu_2(\text{corrected})=\mu_2-\frac{i^2}{12}=20-\frac{10^2}{12}=20-\frac{100}{12}=20-8.33=11.67$$

$$\text{and } \mu_4(\text{corrected})=\mu_4-\frac{1}{2}i^2\cdot\mu_2+\frac{7}{240}(i)^4=50-\frac{1}{2}\cdot(10)^2(20)+\frac{7}{240}(10)^4$$

$$=50-1000+\frac{7}{240}(10{,}000)=50-1000+291.67=-658.33$$

IMPORTANT TYPICAL EXAMPLES

Example 26. For a distribution, the mean is 10, standard deviation is 4, $\sqrt{\beta_1} = 1$, and $\beta_2 = 4$. Obtain the first four moments about '4'.

Solution: Given, $\overline{X} = 10, \sigma = 4, \sqrt{\beta_1} = 1$ and $\beta_2 = 4$

$$\mu_2 = \sigma^2 = (4)^2 = 16$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \quad \Rightarrow \quad 1 = \frac{\mu_3^2}{(16)^3} \quad \Rightarrow \quad \mu_3^2 = 4096$$

$$\therefore \quad \mu_3 = \sqrt{4096} = 64$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \quad \Rightarrow \quad 4 = \frac{\mu_4}{(16)^2} \quad \Rightarrow \quad \mu_4 = 4 \times 256 = 1024$$

$$\therefore \quad \mu_1 = 0,\ \mu_2 = 16,\ \mu_3 = 64,\ \mu_4 = 1024$$

Moments about 4

$$\mu_1' = \overline{X} - A = 10 - 4 = 6$$

$$\mu_2' = \mu_2 + (\mu_1')^2 = 16 + (6)^2 = 52$$

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

$$= 64 + 3(52)(6) - 2(6)^3 = 64 + 936 - 432 = 568$$

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

$$= 1024 + 4(568)(6) - 6(52)(6)^2 + 3(6)^4$$

$$= 1024 + 13632 - 11232 + 3888 = 7312$$

Example 27. Examine whether the following results of a piece of computation for obtaining the second central moments are consistent or not:

$N = 120,\ \Sigma fX = -125,\ \Sigma fX^2 = 128$

Solution: $\mu_1 = \dfrac{-125}{120} = -1.042 \qquad \mu_2' = \dfrac{128}{120} = 1.066$

$$\mu_2 = \mu_2' - \mu_1'^2 = 1.066 - (1.042)^2 = 1.066 - 1.085 = -0.019$$

As the variance $\mu_2 = \sigma^2$ can never be negative, the data for obtaining μ_2 are not consistent.

Aliter:

$$\sigma^2 = \mu_2 = \frac{\Sigma fX^2}{N} - \left(\frac{\Sigma fX}{N}\right)^2 = \frac{128}{120} - \left(\frac{-125}{120}\right)^2 = 1.066 - 1.085 = -0.019$$

Example 28. For a distribution it has been found that the first four moments about 27 are 0, 256, – 2871 and 1,88,462 respectively. Outline the first four moments about zero. Also calculate the values of β_1 and β_2 and comment.

Solution: Given, $A = 27, \mu_1' = 0, \mu_2' = 256, \mu_3' = -2871, \mu_4' = 1,88,462$

Moments about Mean:

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 256 - 0 = 256$$

$$\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2(\mu_1'^3) = -2871 - 3(256)(0) + 2(0) = -2871$$

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

$$= 188462 - 4(-2871)(0) + 6(256)(0)^2 - 3(0)^4 = 188462$$

Moments about Zero:

$$v_1 = \overline{X} = A + \mu_1' = 27 + 0 = 27$$

$$v_2 = \mu_2 + v_1^2 = 256 + (27)^2 = 256 + 729 = 985$$

$$v_3 = \mu_3 + 3v_2 . v_1 - 2v_1^3$$

$$= -2871 + 3(985)(27) - 2(27)^3$$

$$= -2871 + 79785 - 39366 = 37548$$

$$v_4 = \mu_4 + 4v_3 . v_1 - 6v_2 .(v_1^2) + 3(v_1^4)$$

$$= 188462 + 4(37548)(27) - 6(985)(27)^2 + 3(27)^4$$

$$= 188462 + 4055184 - 4308390 + 1594323 = 1529579$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-2871)^2}{(256)^3} = \frac{8242641}{16777216} = 0.49$$

Comment : Since, $\beta_1 = 0.49$, the distribution is positively of skewed.

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{188462}{(256)^2} = \frac{188462}{65536} = 2.875$$

Comment : As $\beta_2 < 3$, the distribution is platy-kurtic.

Example 29. For a meso-kurtic distribution, the first moment about 7 is 23 and the second moment about origin is 1000. Find the coefficient of variation and the fourth moment about the mean.

Solution: Since the distribution is given to be meso-kurtic, we have

$$\beta_2 = 3 \quad \Rightarrow \quad \frac{\mu_4}{\mu_2^2} = 3 \quad \Rightarrow \quad \mu_4 = 3\mu_2^2 \quad \text{...}(i)$$

First moment about '7' is 23

i.e., μ_1' (about 7) = 23 (Given)

$$\therefore \quad \text{Mean} = 7 + \mu_1' = 7 + 23 = 30 \quad \text{...}(ii)$$

But mean is the first moment about origin.

$\therefore$ μ_1' (about origin) = 30

Moments About Origin

$\mu_1' = \text{Mean} = 30; \quad \mu_2' = 1{,}000$ (Given)

$\therefore \quad \mu_2 = \mu_2' - \mu_1'^2 = 1000 - (30)^2 = 100 \quad \Rightarrow \quad \text{Variance } (\sigma^2) = 100 \Rightarrow \text{S.D.}(\sigma) = 10$

$$\text{Coefficient of Variation (C.V.)} = \frac{100 \times \text{S.D.}}{\text{Mean}} = \frac{100 \times 10}{30} = 33.33$$

Substituting the value of $\mu_2 = 100$, in (*i*), the fourth moment about mean is given by:

$$\mu_4 = 3 \times (100)^2 = 30{,}000.$$

Example 30. If $\beta_1 = +1, \beta_2 = 4$ and variance $= 9$, find the values of μ_3 and μ_4 and comment upon the nature of the distribution.

Solution: We are given, $\beta_1 = +1, \beta_2 = 4$ and variance $= \mu_2 = 9$

$$\beta_1 = +1 \quad \Rightarrow \quad \frac{\mu_3^2}{\mu_2^3} = 1$$

$$\Rightarrow \quad \mu_3^2 = \mu_2^3 = 9 \times 9 \times 9 = (3 \times 3 \times 3)^2 = (27)^2 \quad \Rightarrow \quad \mu_3 = \pm 27$$

$$\text{Also, } \beta_2 = 4 \quad \Rightarrow \quad \frac{\mu_4}{\mu_2^2} = 4 \quad \Rightarrow \quad \mu_4 = 4 \times 9 \times 9 = 324$$

$$\therefore \quad \mu_3 = \pm 27 \quad \text{and} \quad \mu_4 = 324.$$

Nature of the Distribution: Since $\beta_1 \neq 0$, but $\beta_1 = 1$, the distribution is moderately skewed. Further, since μ_3 $(= \pm 27)$ can be positive or negative, we cannot tell the direction of the skewness.

Also $\beta_2 = 4 > 3$. Hence, the given distribution is lepto-kurtic, *i.e.*, more peaked than the normal curve.

Example 31. The first three moments of the distribution about the value '2' of the variables are 1, 16 and – 40. Show that the mean is 3, variance is 15 and $\mu_3 = -86$.

Solution: We are given, $A = 2, \mu_1' = 1, \mu_2' = 16, \mu_3' = -40$

$$\text{Mean } (\bar{X}) = A + \mu_1' = 2 + 1 = 3$$

$$\text{Variance } (\sigma^2) = \mu_2 = \mu_2' - (\mu_1')^2 = 16 - (1)^2 = 16 - 1 = 15$$

$$\begin{aligned} \mu_3 &= \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 \\ &= -40 - 3(16)(1) + 2(1)^3 \\ &= -40 - 48 + 2 = -86 \end{aligned}$$

Example 32. The first four moments of a distribution about the value '3' of the variable are 1.2, 13, – 22 and 180. Find the value of β_2.

Solution: We are given, $\mu_1' = 1.2, \mu_2' = 13, \mu_3' = -22, \mu_4' = 180$

But
$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

$$\mu_2 = \mu_2' - (\mu_1')^2 = 13 - (-1.2)^2 = 13 - 1.44 = 11.56.$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$
$$= 180 - 4(-22)(-1.2) + 6(13)(-1.2)^2 - 3(-1.2)^4$$
$$= 180 - 105.6 + 112.32 - 6.2208$$
$$= 292.32 - 111.8208 = 180.4992$$

Now, $\beta_2 = \dfrac{\mu_4}{\mu_2^2}$

Here, $\mu_4 = 180.4992$, $\mu_2 = 11.56$

$$\beta_2 = \frac{180.4992}{(11.56)^2} = \frac{180.4992}{133.6336} = 1.35$$

Since β_2 is less than 3, so the curve is platy-kurtic.

IMPORTANT FORMULAE

Moments about Mean

$$\mu_1 = \frac{\Sigma(X-\bar{X})^1}{N} = 0, \qquad \mu_2 = \frac{\Sigma(X-\bar{X})^2}{N}$$

$$\mu_3 = \frac{\Sigma(X-\bar{X})^3}{N}, \qquad \mu_4 = \frac{\Sigma(X-\bar{X})^4}{N}$$

For a Frequency Distribution

$$\mu_1 = \frac{\Sigma f(X-\bar{X})^1}{N}, \qquad \mu_2 = \frac{\Sigma f(X-\bar{X})^2}{N} \text{ etc.}$$

Moments about Arbitrary Origin 'A'

$$\mu_1' = \frac{\Sigma(X-A)^1}{N}, \qquad \mu_2' = \frac{\Sigma(X-A)^2}{N}$$

$$\mu_3' = \frac{\Sigma(X-A)^3}{N}, \qquad \mu_4' = \frac{\Sigma(X-A)^4}{N}$$

For a Frequency Distribution

$$\mu_1' = \frac{\Sigma f(X-A)^1}{N} \times i \quad \text{or} \quad \mu_1' = \frac{\Sigma fd'}{N} \times i$$

$$\mu_2' = \frac{\Sigma f(X-A)^2}{N} \times i^2 \quad \text{or} \quad \mu_2' = \frac{\Sigma fd'^2}{N} \times i^2$$

$$\mu_3' = \frac{\Sigma f(X-A)^3}{N} \times i^3 \quad \text{or} \quad \mu_3' = \frac{\Sigma fd'^3}{N} \times i^3$$

▶ **Moments about Zero**

$$v_1 = \frac{\Sigma X^1}{N}, \quad v_2 = \frac{\Sigma X^2}{N}$$

$$v_3 = \frac{\Sigma X^3}{N}, \quad v_4 = \frac{\Sigma X^4}{N}$$

▶ **Relationship between Central and Non-central Moments**

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$\mu_3 = \mu_3' - 3\mu_2' \cdot \mu_1' + 2\mu_1'^3$$

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

▶ **Relationship between Central Moments and Moments about Origin**

$$v_1 = \overline{X}, \quad v_2 = \mu_2 + v_1^2$$

$$v_3 = \mu_3 + 3\mu_2 . v_1 + (v_1)^3, \quad v_4 = \mu_4 + 4\mu_3 . v_1 + 6\mu_2 .(v_1)^2 + (v_1)^4$$

▶ **Skewness and Kurtosis**

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}, \quad \gamma_1 = \sqrt{\beta_1} = \frac{\mu_3}{\sqrt{\mu_2^3}}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}, \quad \gamma_2 = \beta_2 - 3$$

QUESTIONS

1. What do you understand by skewness and kurtosis? Give formulae for measuring them.
2. Explain the term kurtosis. How does kurtosis differ from skewness?
3. What is kurtosis? How is it measured?
4. Define moments. How are skewness and kurtosis calculated from central moments?
5. Distinguish between skewness and kurtosis.
6. What is kurtosis? What purpose does it serve?
7. Define Moments. How are they useful in analysing the different aspects of a frequency distribution?
8. Discuss various measures of kurtosis?
9. How do you measure skewness and kurtosis by using moments?
10. Give measures of skewness and kurtosis.
11. What is kurtosis? Explain the methods to measure kurtosis.
12. What are Sheppard's corrections for grouping errors ? State the conditions under which Sheppard's corrections are applicable.