Unit-6

BASIC STATISTICS

In this unit, students will learn how to

- *construct grouped frequency table.*
- *construct histograms with equal and unequal class intervals.*
- *construct a frequency polygon.*
- *construct a cumulative frequency table.*
- *draw a cumulative frequency polygon.*
- *calculate (for ungrouped and grouped data)*
  - *arithmetic mean by definition and using deviations from assumed mean.*
- *calculate median, mode, geometric mean, harmonic mean.*
- *recognize properties of arithmetic mean.*
- *calculate weighted mean and moving averages.*
- *estimate median, quartiles and mode graphically.*
- *measure range, variance and standard deviation.*

6.1 Frequency Distribution

A frequency distribution is a tabular arrangement for classifying data into different groups and the number of observations falling in each group corresponds to the respective group. The data presented in the form of frequency distribution is called Grouped Data. Hence a frequency distribution is a method to summarize data.

6.1(i) Construction of Frequency Table

On the basis of types of variable or data, there are two types of frequency distribution. These are:

(a) Discrete Frequency Distribution.

(b) Continuous Frequency Distribution.

(a) Discrete Frequency Table

Following steps are involved in making of a discrete frequency distribution:

(i) Find the minimum and maximum observation in the data and write the values of the variable in the variable column from minimum to the maximum.

(ii) Record the observations by using tally marks. (Vertical bar '|')

(iii) Count the tally and write down the frequency in the frequency column.

Example 1: Five coins are tossed 20 times and the number of heads recorded at each toss are given below: 3, 4, 2, 3, 3, 5, 2, 2, 2, 1, 1, 2, 1, 4, 2, 2, 3, 3, 4, 2.

Make frequency distribution of the number of heads observed.

Solution: Let X = Number of Heads. The frequency distribution is given below:

Frequency distribution of number of heads

X	Tally Marks	Frequency
1	|||	3
2	~~||||~~ |||	8
3	~~||||~~	5
4	|||	3
5	|	1

(b) Continuous Frequency Table

The making of continuous frequency distribution involves the following steps:

(i) Find the Range, where Range = $X_{max} - X_{min}$ (the difference between *maximum* and *minimum* observations).

(ii) Decide about the number of groups (denote it by k) into which the data is to be classified (usually an integer between 5 and 20). Usually it depends upon the range. The larger the range the more the number of groups.

(iii) Determine the size of class (denote by h) by using the formula:

$$h = \frac{\text{Range}}{k} \qquad \text{(Use formula when "}b\text{" is not given)}$$

Note: The rule of approximation is relaxed in determining h. For example, $h = 7.1$ or $h = 7.9$ may be taken as 8.

(iv) Start writing the classes or groups of the frequency distribution usually starting from the minimum observation and keeping in view the size of a class.

(v) Record the observations from the data by using tally marks.

(vi) Count the number of tally marks and record them in the frequency column for each class.

Example 2: The following are the marks obtained by 40 students in mathematics of class X.

Make a frequency distribution with a class interval of size 10.

51, 55, 32, 41, 22, 30, 35, 53, 30, 60, 59, 15, 7, 18, 40, 49, 40, 25, 14, 18,19, 2, 43, 22, 39, 26, 34, 19, 10, 17, 47, 38, 13, 30, 34, 54, 10, 21, 51, 52.

Solution: Let X = marks of a student.

From the above data we have $X_{min} = 2$, $X_{max} = 60$. It is given that $h = 10$. We can either start from 2 or the nearest smallest integer 0 for our convenience. There are two ways to make frequency distribution.

(a) We may write the actual observations falling in the respective groups. This is given as follows:

Classes/Groups	Observations	Frequency
0 —– 9	2, 7	2
10 —– 19	10, 10, 13, 14, 15, 17, 18, 18, 19, 19	10
20 —– 29	21, 22, 22, 25, 26	5
30 —– 39	30, 30, 30, 32, 34, 34, 35, 38, 39	9
40 —– 49	40, 40, 41, 43, 47, 49	6
50 —– 59	51, 51, 52, 53, 54, 55, 59	7
60 —– 69	60	1

(b) Use tally marks for recording each observation in the respective group. This is given in the following table:

Classes/Groups	Tally Marks	Frequency
0 —– 9	𝍷𝍷	2
10 —– 19	𝍸 𝍸	10
20 —– 29	𝍸	5
30 —– 39	𝍸 𝍷𝍷𝍷𝍷	9
40 —– 49	𝍸 𝍷	6
50 —– 59	𝍸 𝍷𝍷	7
60 —– 69	𝍷	1
Total		40

Note: The solution (b) is usually adopted to construct a frequency distribution.

Concepts involved in a Continuous frequency table:

The following terms are frequently used in a continuous frequency distribution:

(a) Class Limits: The minimum and the maximum values defined for a class or group are called Class limits. The minimum value is called the **lower class limit** and the maximum value is called the **upper class limit** of that class. In example 2, the lower class limits are 0, 10, 20, 30 etc., while the upper class limits are 9, 19, 29, 39 etc.

(b) Class Boundaries: As a continuous frequency distribution is based on measurable characteristic variable which involves the rule of approximation to record any observation. From example 2, some class boundaries are given below:

Class limits	Class Boundaries
0 ––– 9	–0.5 ––– 9.5
10 ––– 19	9.5 ––– 19.5
20 ––– 29	19.5 ––– 29.5

Hence referring to example 2, we may say that the real lower class limit of 10 is 9.5, as all values between 9.5 and 10.49 are recorded as 10. While the upper class limit of 19 is 19.5 as all values between 18.5 and 19.5 are recorded as 19. The real class limits of a class are called **class boundaries**. A **class boundary** is obtained by adding two successive class limits and dividing the sum by 2. The value so obtained is taken as **upper class boundary** for the previous class and **lower class boundary** for the next class.

(c) Midpoint or Class Mark: For a given class the average of that class obtained by dividing the sum of upper and lower class limits by 2, is called the **midpoint** or **class mark** of that class.

(d) Cumulative Frequency: The total of frequency up to an upper class limit or boundary is called the **cumulative frequency**.

The above concepts have been explained with reference to Example 2 below:

Example 3: Compute class boundaries, class marks and cumulative frequency for data of example 2.

Solution: Computation follows,

Class limits	Class Boundaries	Midpoint/ Class mark	Frequency	Cumulative frequency
0 ––– 9	–0.5 ––– 9.5	4.5	2	2
10 ––– 19	9.5 ––– 19.5	14.5	10	2 + 10 = 12
20 ––– 29	19.5 ––– 29.5	24.5	5	12 + 5 = 17
30 ––– 39	29.5 ––– 39.5	34.5	9	17 + 9 = 26
40 ––– 49	39.5 ––– 49.5	44.5	6	26 + 6 = 32
50 ––– 59	49.5 ––– 59.5	54.5	7	32 + 7 = 39
60 ––– 69	59.5 ––– 69.5	64.5	1	39 + 1 = 40
Total			40	

6.1(ii) Construction of Histograms

Histogram

A **Histogram** is a graph of adjacent rectangles constructed on *XY*-plane. It is a graph of frequency distribution. In practice both discrete and continuous frequency distributions are represented by means of histogram. However there is a little difference in the construction procedure. We explain this with the help of examples.

Equal Intervals Histogram:

Example 1: Make a Histogram of the following distribution of the number of heads when 5 coins were tossed.

X (number of heads)	Frequency
0	1
1	3
2	8
3	5
4	3
5	1

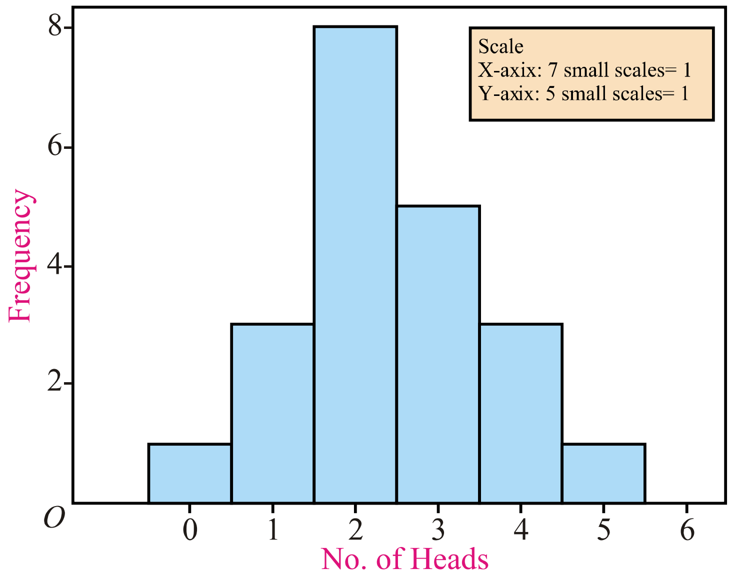
Solution: We proceed as follows,

Step 1: Mark the values of variable X along x-axis using a suitable interval.

Step 2: Mark the frequency along y-axis using a suitable scale.

Step 3: At each interval make a rectangle of height corresponding to the respective frequency of values of the variable X.

The resulting Histogram is given below:



Example 2: Make a Histogram for the following distribution of marks.

Class Boundaries	Frequency
–0.5 —— 9.5	2
9.5 —— 19.5	10
19.5 —— 29.5	5
29.5 —— 39.5	6
39.5 —— 49.5	7
49.5 —— 59.5	1

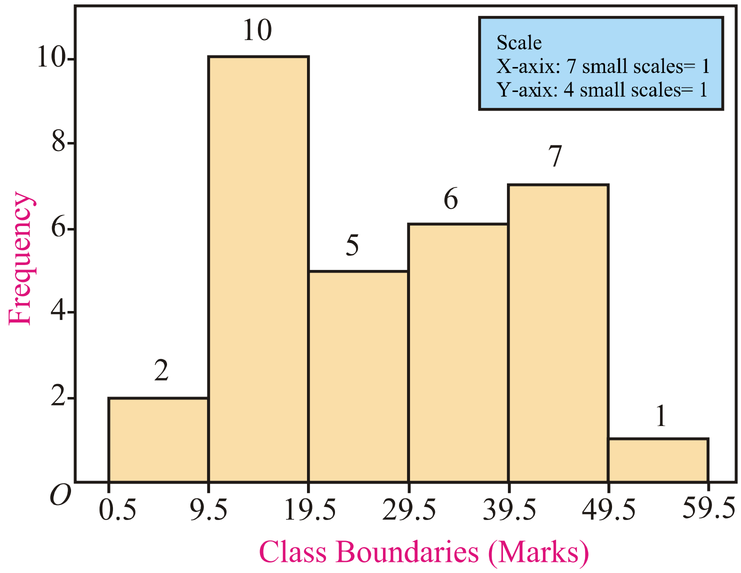
Solution: Since this is a continuous frequency distribution so we proceed as follows:

(a) Mark the class boundaries along x-axis using a suitable scale.

(b) Mark the frequency along y-axis using a suitable scale.

(c) At each class interval construct a rectangle of height corresponding to the frequency of that group.

Histogram



Note: On graphs above 0 and –0.5 are written on positive side of x-axis just to better understand the histogram.

Unequal Intervals Histogram

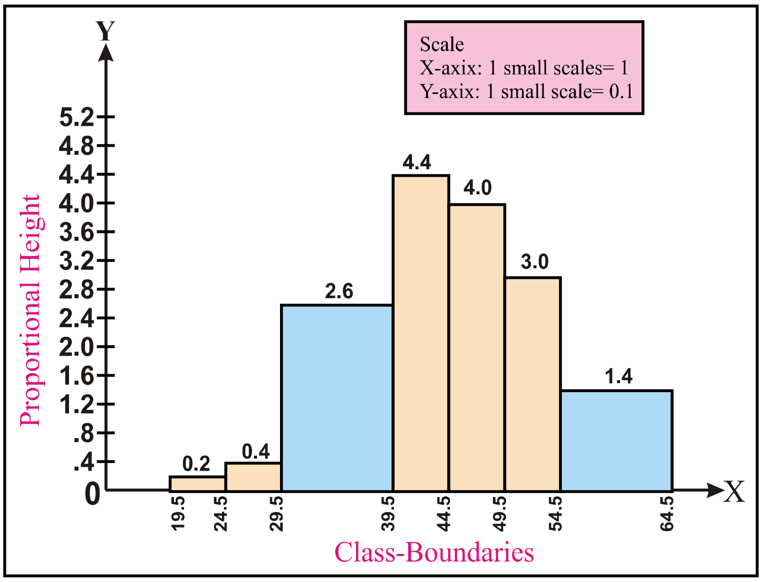
If the class intervals are un-equal, the frequency must be adjusted by dividing each class frequency on its class interval size. If the interval becomes double, then frequency is divided by 2, so that the area of the bar is in proportion to the areas of other bars etc.

Example 3: Draw a histogram illustrating the following data:

Age	Number of Men
20 —— 24	1
25 —— 29	2
30 —— 39	26
40 —— 44	22
45 —— 49	20
50 —— 54	15
55 —— 64	14

Solution: As the class intervals are unequal the height of each rectangle cannot be made equal to the frequency. Therefore, we obtain proportional heights by dividing each frequency with class interval size. This is shown in the following table:

Age	Class Boundaries	Class Interval (h)	Frequency	Proportional Heights
20 —— 24	19.5 —— 24.5	5	1	$1 \div 5 = 0.20$
25 —— 29	24.5 —— 29.5	5	2	$2 \div 5 = 0.4$
30 —— 39	29.5 —— 39.5	10	26	$26 \div 10 = 2.6$
40 —— 44	39.5 —— 44.5	5	22	$22 \div 5 = 4.4$
45 —— 49	44.5 —— 49.5	5	20	$20 \div 5 = 4.0$
50 —— 54	49.5 —— 54.5	5	15	$15 \div 5 = 3.0$
55 —— 64	54.5 —— 64.5	10	14	$14 \div 10 = 1.4$



6.1(iii) Construction of Frequency Polygon

A **Frequency Polygon** is a many sided closed figure. Its construction is explained by the following example:

Example 1: For the following data make a Frequency Polygon.

Class limits	Class Boundaries	Frequency
10—19	9.5—19.5	10
20—29	19.5—29.5	5
30—39	29.5—39.5	9
40—49	39.5—49.5	6
50—59	49.5—59.5	7
60—69	59.5—69.5	1

Solution:

Step 1. Take two additional groups with the same class interval size. One before the very first group and the second after the very last group. Also calculate midpoints for these two groups. These groups will have frequency '0'.

Class limits	Class Boundaries	Frequency
0 — 9	**–0.5 — 9.5**	0
10 — 19	9.5 — 19.5	10
20 — 29	19.5 — 29.5	5
30 — 39	29.5 — 39.5	9
40 — 49	39.5 — 49.5	6
50 — 59	49.5 — 59.5	7
60 — 69	59.5 — 69.5	1
70 — 79	**69.5 — 79.5**	0

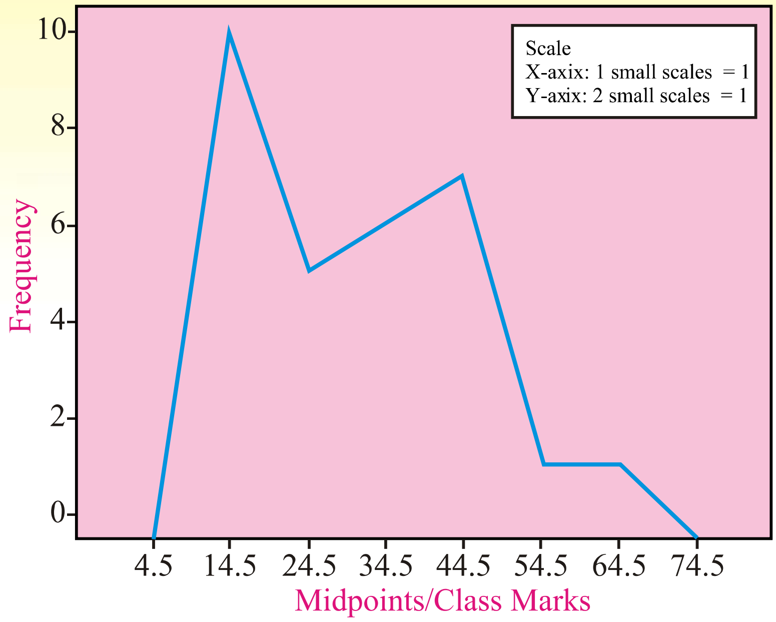
Step 2. Calculate class marks or midpoints for the given distribution.

Midpoints / Class marks	Frequency
4.5	0
14.5	10
24.5	5
34.5	9
44.5	6
54.5	7
64.5	1
74.5	0

Step 3. Mark midpoints at x-axis and frequency along y-axis using appropriate scale.

Step 4. Plot a point against the frequency for each of the corresponding midpoint/class mark.

Step 5. Join all the points by means of line segments.



6.2 Cumulative Frequency Distribution

6.2(i) Construction of Cumulative Frequency Table

A table showing cumulative frequencies against upper class boundaries is called a **cumulative frequency distribution**. It is also called ***a less than cumulative frequency distribution***.

Example 1: Construct a cumulative frequency distribution for the following data.

Classes	20 – 24	25 – 29	30 – 34	35 – 39	40 – 44	45 – 49	50 – 54
Frequency	1	2	26	22	20	15	14

Solution: The **cumulative frequency distribution** is constructed below:

Class Boundaries	Frequency (f)	Cumulative Frequency	Class Boundaries	Cumulative Frequency
10.5 — 19.5	0	0	Less than 19.5	0
19.5 — 24.5	1	0 + 1 = 1	Less than 24.5	1
24.5 — 29.5	2	1 + 2 = 3	Less than 29.5	3
29.5 — 34.5	26	3 + 26 = 29	Less than 34.5	29
34.5 — 39.5	22	29 + 22 = 51	Less than 39.5	51
39.5 — 44.5	20	51 + 20 = 71	Less than 44.5	71
44.5 — 49.5	15	71 + 15 = 86	Less than 49.5	86
49.5 — 54.5	14	86 + 14 = 100	Less than 54.5	100

6.2(ii) Drawing of Cumulative Frequency Polygon or Ogive

A cumulative frequency polygon or ogive is a graph of less than cumulative frequency distribution. It involves the following steps:

Step 1. Mark the class boundaries on x-axis and frequency (cumulative) on y-axis.

Step 2. Plot the points for the given frequencies corresponding to the upper class boundaries.

Step 3. Join the points by means of line segments.

Step 4. Drop perpendicular from the last point to x-axis to make a closed figure.

Example 2: Construct a cumulative frequency polygon for the given data.

Class limits	Frequency
4 — 6	2
7 — 9	4
10 — 12	8
13 — 15	3

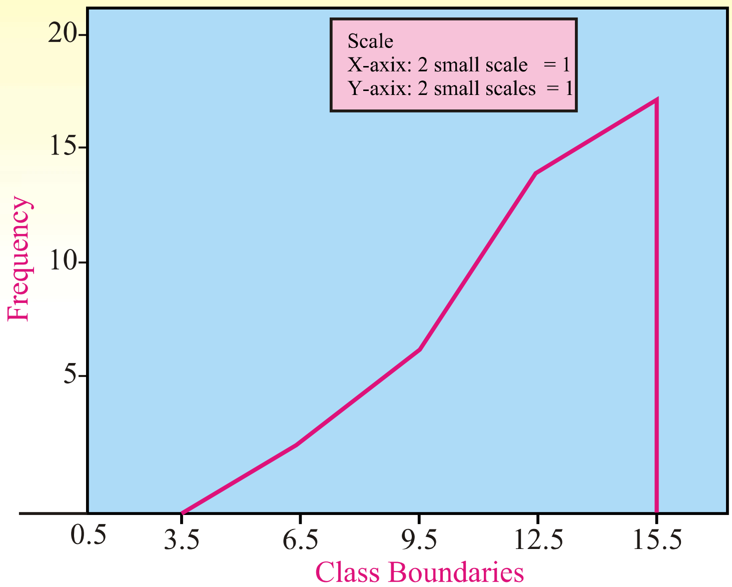
Solution: First we add one group before the first group. Then we make the class boundaries and also calculate the cumulative frequencies.

Class limits	Class Boundaries	Frequency	Cumulative frequency
1 — 3	**0.5 — 3.5**	**0**	**0**
4—6	3.5 — 6.5	2	0 + 2 = 2
7—9	6.5 — 9.5	4	2 + 4 = 6
10—12	9.5 — 12.5	8	6 + 8 = 14
13—15	12.5 — 15.5	3	14 + 3 = 17

Now we write the above frequency distribution in the form of Less than cumulative distribution as given below:

Class Boundaries	Cumulative frequency
Less than 3.5	0
Less than 6.5	2
Less than 9.5	6
Less than 12.5	14
Less than 15.5	17

The Cumulative frequency polygon follows:



EXERCISE 6.1

1. The following data shows the number of members in various families. Construct frequency distribution. Also find cumulative frequencies.

   9, 11, 4, 5, 6, 8, 4, 3, 7, 8, 5, 5, 8, 3, 4, 9, 12, 8, 9, 10, 6, 7, 7, 11, 4, 4, 8, 4, 3, 2, 7, 9, 10, 9, 7, 6, 9, 5, 7.

2. The following data has been obtained after weighing 40 students of class V. Make a frequency distribution taking class interval size as 5. Also find the class boundaries and midpoints.

   34, 26, 33, 32, 24, 21, 37, 40, 41, 28, 28, 31, 33, 34, 37, 23, 27, 31, 31, 36, 29, 35, 36, 37, 38, 22, 27, 28, 29, 31, 35, 35, 40, 21, 32, 33, 27, 29, 30, 23.

   Also make a less than cumulative frequency distribution. (Hint: Make classes 20—24, 25—29…..).

3. From the following data representing the salaries of 30 teachers of a school .Make a frequency distribution taking class interval size of Rs.100**,** 450, 500, 550, 580, 670, 1200,1150, 1120, 950, 1130, 1230, 890, 780, 760, 670, 880, 890, 1050, 980, 970, 1020, 1130, 1220, 760, 690, 710, 750, 1120, 760, 1240.

   (Hint: Make classes 450—549, 550—649,…..).

4. The following data shows the daily load shedding duration in hours in 30 localities of a certain city. Make a frequency distribution of the load shedding duration taking 2 hours as class interval size and answer the following questions.

   6, 12, 5, 7, 8, 3, 6, 7, 10,2, 14, 11, 12, 8, 6, 8, 9, 7, 11, 6, 9, 12, 13, 10, 14, 7, 6, 10, 11, 14, 12.

   (a) Find the most frequent load shedding hours?

(b) Find the least load shedding intervals?

(Hint: Make classes 2—3, 4—5, 6—7….)

5. Construct a Histogram and frequency Polygon for the following data showing weights of students in kg.

Weights	Frequency / No. of students
20—24	5
25—29	8
30—34	13
35—39	22
40—44	15
45—49	10
50—54	8

6.3 Measures of Central Tendency

Introduction

The purpose of frequency distribution and graphical techniques is to view, summarize and understand different aspects of data in a simple manner. But we are also interested to find a '*representative*' of the data under study. In other words to determine a specific value of the variable around which the majority of the observations tend to concentrate. This *representative* shows the tendency or behavior of the distribution of the variable under study. This value is called **average** or **the central value.** The measures or techniques that are used to determine this central value are called **Measures of Central Tendency**. The following measures of central tendency will be discussed in this section:

1. Arithmetic mean
2. Median
3. Mode
4. Geometric mean
5. Harmonic mean
6. Quartiles

All these measures are used under different situations depending upon the nature of the data.

6.3(i-a) Arithmetic Mean

Arithmetic Mean (or simply called Mean) is a measure that determines a value (observation) of the variable under study by dividing the sum of all values (observations) of the variable by their number of observations. We denote Arithmetic mean by $\overline{X}$. In symbols we define:

Arithmetic mean of n observations $\overline{X} = \frac{\sum X}{n} = \frac{\text{Sum of all values of observation}}{\text{No. of obsrvation}}$

Computation of Arithmetic Mean

There are two types of data, ungrouped and grouped. We, therefore have different methods to determine Mean for the two types of data. These are explained with the help of examples.

Ungrouped Data

For ungrouped data we use three approaches to find mean. These are as follows.

(i) Direct Method (By Definition)

The formula under this method is given by:

$$\bar{X} = \frac{\sum X}{n} = \frac{\text{Sum of all observations}}{\text{No. of observations}}$$

Example 1: The marks of seven students in Mathematics are as follows. Calculate the Arithmetic Mean and interpret the result.

Student No	1	2	3	4	5	6	7
Marks	45	60	74	58	65	63	49

Solution: Let X = marks of a student

$$\bar{X} = \frac{\sum X}{n} = \frac{x_1 + x_2 + x_3 + \text{........} + x_7}{7}$$

or $$\bar{X} = \frac{45 + 60 + 74 + 58 + 65 + 63 + 49}{7} = \frac{414}{7} = 59.14 \text{ marks}$$

Explanation: Since the unit of data is marks, so the result is also in marks. Hence we may say that, 'Each of the seven students obtains 59 marks on the average'.

(ii) Indirect, Short Cut or Coding Methods

There are two approaches under Indirect Method. These are used to find mean when data set consists of large values or large number of values. The purpose is to simplify the computation of Mean. These approaches exist in theory but are not used in practice as many Statistical softwares are available now to handle large data. However a student should have knowledge of the two approaches. These are:

(i) using an Assumed or Provisional mean

(ii) using a Provisional mean and changing scale of the variable.

Deviation is defined as difference of any value of the variable from any constant 'A'. For example we say,

Deviation from mean of X $= (x_i - \bar{X}) = (X_i - \bar{X})$ for i = 1, 2, …… n

Deviation from any constant A = $(x_i - A) = (X_i - A)$ for i = 1, 2, …… n

The formulae used under indirect methods are:

(i) $$\bar{X} = A + \frac{\sum_{i=1}^{n} D_i}{n}$$ (ii) $$\bar{X} = A + \frac{\sum_{i=1}^{n} u_i}{n} \times h$$

where

$D_i = (X_i - A)$, A is any assumed value of X called Assumed or Provisional mean.

$u_i = \frac{(X_i - A)}{h}$, "$h$" is the class interval size for unequal intervals.

Example 2: The salaries of five teachers are as follows. Find the mean salary using direct and indirect methods and compare the results. 11500, 12400, 15000, 14500, 14800.

Solution: We proceed as follows:

(a) Using Direct method

$$\bar{X} = \frac{\sum_{i=1}^{5} x_i}{5} = \frac{11500 + 12400 + 15000 + 14500 + 14800}{5}$$

$$= \frac{68200}{5} = 13640 \text{ Rupees.}$$

we assume $A = 13000$, $D_i = (x_i - 13000)$, $h = 100$ and $u_i = \frac{(x_i - A)}{100}$, the computations are shown in the following table:

X	$D_i = (x_i - 13000)$	$u_i = \frac{(x_i - A)}{100}$
11500	−1500	−15
12400	−600	−6
15000	2000	20
14500	1500	15
14800	1800	18
$\Sigma x_i = 68200$	$\Sigma D_i = 3200$	$\Sigma u_i = 32$

(i) Using Indirect methods

$$\bar{X} = 13000 + \frac{3200}{5} = 13000 + 640 = 13640 \text{ Rupees}$$

(ii) Using Indirect method

$$\bar{X} = 13000 + \frac{32}{5} \times 100 = 13640 \text{ Rupees}$$

Grouped Data

A data in the form of frequency distribution is called **grouped data**. For the grouped data we define formulae under Direct and Indirect methods as given below:

(a) Using Direct method,

$$\bar{X} = \frac{\sum f x_i}{\sum f} = \frac{\sum fX}{\sum f}$$

(b) Using Indirect method,

(i) $\bar{X} = A + \frac{\sum fD}{\sum f}$ (ii) $\bar{X} = A + \frac{\sum fu}{\sum f} \times h$

where '$X = x_i$' denotes the midpoint of a class or group if class intervals are given, and 'h' is the class interval size .

Example 3: Find the arithmetic mean using Direct method for the following frequency distribution.

(Number of heads) X	Frequency
1	3
2	8
3	5
4	3
5	1

Solution: We compute mean as follows:

X	f	fX
1	3	3
2	8	16
3	5	15
4	3	12
5	1	5
Total	$\Sigma f = 20$	$\Sigma fX = 49$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{49}{20} = 2.45 \text{ or 3 heads (since the variable is discrete).}$$

Example 4: For the following data showing weights of toffee boxes in gms. Determine the mean weight of boxes.

Classes / Groups	Frequency
0 —— 9	2
10 —— 19	10
20 —— 29	5
30 —— 39	9
40 —— 49	6
50 —— 59	7
60 —— 69	1
Total	$\Sigma f = 40$

Solution: First we calculate midpoints for each class and then find arithmetic mean.

Classes / Groups	f	Midpoints X	fX
0 —— 9	2	4.5	9
10 —— 19	10	14.5	145
20 —— 29	5	24.5	122.5
30 —— 39	9	34.5	310.5
40 —— 49	6	44.5	267
50 —— 59	7	54.5	381.5
60 —— 69	1	64.5	64.5
Total	$\Sigma f = 40$		$\Sigma fX = 1300$

$$\bar{X} = \frac{\sum fX}{\sum f} = \frac{1300}{40} = 32.5 \text{ gm}$$

Example 5: Find arithmetic mean using short formulae taking $X = 34.5$ as the provisional mean in example 4.

Solution: We use the following formulae

$$\text{(i)} \quad \bar{X} = A + \frac{\sum fD}{\sum f} \qquad\qquad \text{(ii)} \quad \bar{X} = A + \frac{\sum fu}{\sum f} \times h$$

Given $A = 34.5$, we note that the distribution has equal class interval size of 10. So we may take $h = 10$ and make the following calculations:

Classes/ Groups	f	Midpoints X	$D = X - 34.5$	$u = (X - A)/10$	fD	fu
0 —— 9	2	4.5	–30	–3	–60	–6
10 —— 19	10	14.5	–20	–2	–200	–20
20 —— 29	5	24.5	–10	–1	–50	–5
30 —— 39	9	34.5	0	0	0	0
40 —— 49	6	44.5	10	1	60	6
50 —— 59	7	54.5	20	2	140	14
60 —— 69	1	64.5	30	3	30	3
Total	40				$\Sigma fD = -80$	$\Sigma fu = -8$

Substituting the totals in the above formulae, we get

$$\text{(i)} \quad \bar{X} = 34.5 + \frac{-80}{40} = 34.5 - 2 = 32.5 \text{ gm}$$

(ii) $\bar{X} = 34.5 + \frac{-8}{40} \times 10 = 34.5 - 2 = 32.5$ gm.

Hence the three methods yield the same answer.

6.3(i) (b) Median

Median is the middle most observation in an arranged data set. It divides the data set into two equal parts. '$\tilde{x}$' is used to represent median. We determine Median by using the following formulae:

Ungrouped data

Case-1: When the number of observations is odd of a set of data arranged in order of magnitude the median (middle most observation) is located by the formula given below:

$$\tilde{X} = \text{size of } \left(\frac{n+1}{2}\right)\text{th} \quad \text{observation}$$

Case-2: When the number of observations is *even* of a set of data arranged in order of magnitude the median is the arithmetic mean of the two middle observations. That is, median is average of $\frac{n}{2}$ and $\left(\frac{n}{2}+1\right)$th values.

$$\tilde{X} = \frac{1}{2}\left[\text{size of } \left(\frac{n}{2}\text{ th} + \frac{n+2}{2}\text{ th}\right) \text{observations}\right]$$

Example 1: On 5 term tests in mathematics, a student has made marks of 82, 93, 86, 92 and 79. Find the median for the marks.

Solution: By arranging the grades in ascending order, the arranged data is

79 , 82 , 86 , 92 , 93

Since number of observations is odd *i.e.*, $n = 5$.

$$\tilde{X} = \text{size of } \left(\frac{5+1}{2}\right) \text{th observation}$$

$$\tilde{X} = \text{size of } 3^{rd} \text{ observation}$$

$$\tilde{X} = 86$$

Example 2: The sugar contents for a random sample of 6 packs of juices of a certain brand are found to be 2.3, 2.7, 2.5, 2.9, 3.1 and 1.9 milligram. Find the median.

Solution: Arrange the values by increasing order of magnitude

1.9 , 2.3 , 2.5 , 2.7 , 2.9 , 3.1

Since number of observations is even *i.e.*, $n = 6$.

$$\tilde{X} = \frac{1}{2}\left[\text{size of } \left(\frac{6}{2}\text{th} + \frac{6+2}{2}\text{ th}\right) \text{observations}\right]$$

$$\tilde{X} = \frac{1}{2}[\text{size of } (3^{rd} + 4^{th}) \text{ observations}]$$

$$\tilde{X} = \frac{2.5+2.7}{2} = 2.6 \text{ milligram.}$$

Grouped Data (Discrete)

The following steps are involved in determining median for grouped data(discrete):

- Make cumulative frequency column.
- Determine the median observation using cumulative frequency, i.e., the class containing $\left(\frac{n}{2}\right)^{th}$ observation.

Example 3: Find median for the following frequency distribution.

(Number of heads) X	Frequency
1	3
2	8
3	5
4	3
5	1

Solution: We first make cumulative frequency column as given below:

X	Frequency	Cumulative frequency
1	3	3
2	8	11
3	5	16
4	3	19
5	1	20
Total	$\Sigma f = 20$	

Now

Median = the class containing $\left(\frac{n}{2}\right)^{th}$ observation

Median = the class containing $\left(\frac{20}{2}\right)^{th}$ observation

Median = the class containing $(10)^{th}$ observation

Median = 2.

Grouped Data (Continuous)

The following steps are involved in determining median for grouped data (continuous):

- Determine class boundaries
- Make cumulative frequency column
- Determine the median class using cumulative frequency, i.e., the class containing $\left(\frac{n}{2}\right)^{th}$ observation
- Use the formula,

$$\text{Median} = l + \frac{h}{f}\left\{\frac{n}{2} - c\right\}$$

where l : lower class boundary of the median class,
h : class interval size of the median class,
f : frequency of the median class,
c : cumulative frequency of the class preceding the median class.

Example 4: The following data is the time taken by 40 students to solve a problem is recorded. Find the median time taken by the students.

138	164	150	132	144	125	149	157
146	158	140	147	136	148	152	144
168	126	138	176	163	119	154	165
146	173	142	147	135	153	140	135
161	145	135	142	150	156	145	128

Solution:

(*a*)

Class Intervals	Frequency	Class Boundaries	Cumulative Frequency	
118 –– 126	3	117.5 – 126.5	3	
127 –– 135	5	126.5 – 135.5	8	
136 –– 144	9	135.5 – 144.5	17	
145 –– 153	12	144.5 – 153.5	29	Median class
154 –– 162	5	153.5 – 162.5	34	
163 –– 171	4	162.5 – 171.5	38	
172 –– 180	2	171.5 – 180.5	40	
Total	$\Sigma f = 40$	—	—	

Now

median class = class containing $\left(\frac{n}{2}\right)^{th}$ observation

median class = class containing $\left(\frac{40}{2}\right)^{th} = 20^{th}$ observation

median class = 144.5 – 153.5

l = lower class boundary of median class = 144.5

c = cumulative frequency preceding the median class = 17

f = frequency of median class = 12

h = size of median class interval = 9

So, $\tilde{X} = l + \frac{h}{f}\left(\frac{n}{2} - c\right) = 144.5 + \frac{9}{12}(20 - 17) = 146.8$

6.3(i) (c) Mode

Mode is defined as the most frequent occurring observation in the data. It is the observation that occurs maximum number of times in given data. The following formula is used to determine Mode:

(i) Ungrouped data and Discrete Grouped data

Mode = the most frequent observation

(ii) Grouped Data (Continuous)

The following steps are involved in determining mode for grouped data:

- Find the group that has the maximum frequency.
- Use the formula

$$\text{Mode} = l + \frac{f_m - f_1}{2f_m - f_1 - f_2} \times h\ ,$$

Where l : lower class boundary of the modal class or group,

h : class interval size of the modal class,

f_m : frequency of the modal class,

f_1 : frequency of the class preceding the modal class and

f_2 : frequency of the class succeeding the modal class.

Example 1: Find the modal size of shoe for the following data:

4, 4.5, 5, 6, 6, 6, 7, 7.5, 7.5, 8, 8, 8, 6, 5, 6.5, 7.

Solution: We note the most occurring observation in the given data and find that,

mode = 6.

Example 2: Find Mode for the following frequency distribution.

(Number of heads) X	Frequency
1	3
2	8
3	5
4	3
5	1

Solution: Since the given data is discrete grouped data so that,

mode = 2,

(Since for X = 2 the frequency is maximum, means 2 heads appear the maximum number of times i e .,8) .

Example 3: For the following data showing weights of toffee boxes in gm. Determine the modal weight of boxes.

Classes / Groups	Frequency
0 —— 9	2
10 —— 19	10
20 —— 29	5
30 —— 39	9
40 —— 49	6
50 —— 59	7
60 —— 69	1

Solution: Since the data is continuous so we proceed as follows:

(a) Determine the class boundaries first,

(b) Find the class with maximum frequency

Classes/Groups	Class Boundaries	Frequency	
0 ---- 9	−0.5 ---- 9.5	2	
10 ---- 19	9.5 ---- 19.5	10	Modal class
20 ---- 29	19.5 ---- 29.5	5	
30 ---- 39	29.5 ---- 39.5	9	
40 ---- 49	39.5 ---- 49.5	6	
50 ---- 59	49.5 ---- 59.5	7	
60 ---- 69	59.5 ---- 69.5	1	
Total		$\sum f = 40$	

From the above table we get, modal class or group = 9.5 — 19.5.

$f_m = 10,\ l = 9.5,\ \text{h} = 10,\ f_1 = 2$ and $f_2 = 5$.

$$Mode = 9.5 + \frac{10 - 2}{2(10) - 2 - 5} \times 10$$

$$Mode = 9.5 + \frac{80}{13} = 9.5 + 6.134 = 15.654 \text{ gm}$$

6.3(i) (d) Geometric Mean

Geometric mean of a variable X is the nth positive root of the product of the $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ observations. In symbols, we write

$$\text{G.M.} = (x_1, x_2, x_3, \ldots\ x_n)^{1/n}$$

The above formula can also be written by using logarithm.

For Ungrouped data

$$\text{G.M.} = Anti \log\left(\frac{\sum \log X}{n}\right)$$

For Grouped data

$$\text{G.M.} = Anti \log\left(\frac{\sum f \log X}{\sum f}\right)$$

Example 1: Find the geometric mean of the observations 2, 4, 8 using (i) basic formula and (ii) using logarithmic formula.

Solution: (i) Using basic formula

$$G.M = (2\times4\times8)^{1/3} = (64)^{1/3} = 4.$$

X	**log X**
2	0.3010
4	0.6021
8	0.9031
Total	Σ log X = 1.8062

$$\text{G.M.} = \textit{Anti}\log\left(\frac{1.8062}{3}\right)$$

$$= \textit{Anti}\log(0.6021) = 4.00003 = 4$$

Example 2: Find the geometric mean for the following data:

Marks in percentage	Frequency/ (No of Students)
33 ---- 40	28
41 ---- 50	31
51 ---- 60	12
61 ---- 70	9
71 ---- 75	5

Solution: We proceed as follows:

Classes	f	X	log X	f log X
33 ---- 40	28	36.5	1.562293	43.7442
41 ---- 50	31	45.5	1.658011	51.39835
51 ---- 60	12	55.5	1.744293	20.93152
61 ---- 70	9	65.5	1.816241	16.34617
71 ---- 75	5	73	1.863323	9.316614
Total	$\Sigma f = 85$			Σf log X = 141.7369

$$\text{G.M} = \textit{Anti}\log\left(\frac{141.7369}{85}\right)$$

$$= \textit{Anti}\log(1.66749) = 46.50\ \%\ \text{marks}$$

6.3(i) (e) Harmonic Mean

Harmonic mean refers to the value obtained by reciprocating the mean of the reciprocal of $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ observations. In symbols, for ungrouped data,

$$\text{H.M.} = \frac{n}{\sum \frac{1}{X}}$$

and for grouped data,

$$\text{H.M.} = \frac{n}{\sum \frac{f}{X}}$$

Example 1: For the following data find the Harmonic mean.

X	12	5	8	4

Solution:

X	$^1/_X$
12	0.0833
5	0.2
8	0.125
4	0.25
Total	0.6583

$$\text{H.M.} = \frac{4}{0.6583} = 6.076$$

Example 2: Find the Harmonic mean for the following data.

Classes	No. of Students
33 —— 40	28
41 —— 50	31
51 —— 60	12
61 —— 70	9
71 —— 75	5

Solution:

Classes	f	X	$f\,/\,X$
33 —— 40	28	36.5	0.767123
41 —— 50	31	45.5	0.681319
51 —— 60	12	55.5	0.216216
61 —— 70	9	65.5	0.137405
71 —— 75	5	73	0.068493
Total	$\Sigma f = 85$		$\Sigma \frac{f}{X} = 1.870556$

$$\text{H.M.} = \frac{\Sigma f}{\Sigma \frac{f}{X}} = \frac{85}{1.8706} = 45.441$$

6.3(ii) Properties of Arithmetic Mean

(i) Mean of a variable with similar observations say constant k is the constant k itself.

(ii) Mean is affected by change in origin.

(iii) Mean is affected by change in scale.

(iv) Sum of the deviations of the variable X from its mean is always zero.

Example 1: Find the mean of observations: 34, 34, 34, 34,34, 34.

Solution: Since the variable say X here is taking same observation so by property (i)

$$\bar{X} = 34 .$$

Example 2: A variable X takes the following values 4, 5, 8, 6, 2. Find the mean of X. Also find the mean when (a) 5 is added to each observation (b) 10 is multiplied with each observation (c) Prove sum of the deviation from mean is zero.

Solution: Given the values of X,

X: 4 5 8 6 2.

We may introduce two new variables Y and Z under (*a*) and (*b*) respectively. So we are given that (*a*) $Y = X + 5$ (*b*) $Z = 10X$. The following table shows the desired result:

	X	$Y = X + 5$	$Z = 10X$	$X - \bar{X}$
	4	9	40	−1
	5	10	50	0
	8	13	80	3
	6	11	60	1
	2	7	20	−3
Total	$\Sigma X = 25$	$\Sigma Y = 50$	$\Sigma Z = 250$	$\sum(X - \bar{X}) = 0$

From the above table we get,

$$\bar{X} = \frac{25}{5} = 5 \quad ; \quad \bar{Y} = \frac{50}{5} = 10 \quad ; \quad \bar{Z} = \frac{250}{5} = 50$$

We note that (a) $\bar{Y} = 10 = 5 + 5 = \bar{X} + 5$

(b) $\bar{Z} = 50 = 10\ (5) = 10\bar{X}$

Which shows that mean is affected by change in origin and scale.

(c) From the last column of the table we note that $\sum(X - \bar{X}) = 0$, the sum of the deviations from mean is zero.

6.3 (iii) Calculation of Weighted Mean and Moving Averages

a. The Weighted Arithmetic Mean

The relative importance of a number is called its weight. When numbers x_1, x_2, ... x_n are not equally important, we associate them with certain weights, w_1, w_2, w_3, ..., w_n depending on the importance or significance.

$$\bar{x}_w = \frac{w_1 x_1 + w_2 x_2 + ... + w_n x_n}{w_1 + w_2 + ... + w_n} = \frac{\Sigma wx}{\Sigma w}$$

is called the weighted arithmetic mean.

Example 1: The following table gives the monthly earnings and the number of workers in a factory, compute the weighted average.

No. of employees	Monthly earnings. Rs.
4	800
22	45
20	100
30	30
80	35
300	15

Solution: Number of employees are treated as a weight (w) and monthly earnings as variable (x)

No. of employees (w)	Monthly earning in Rs. (x)	(xw)
4	800	3200
22	45	990
20	100	2000
30	30	900
80	35	2800
300	15	4500
$\Sigma w = 456$	———	$\Sigma xw = 14390$

$$\bar{x}_w = \frac{\Sigma xw}{\Sigma w} = \frac{14390}{456} = 31.5$$

b. Moving Averages

Moving averages are defined as the successive averages (arithmetic means) which are computed for a sequence of days/months/years at a time. If we want to find 3-days moving average, we find the average of first 3-days, then dropping the first day and add the succeeding day to this group. Place the average of each 3-days against the mid of 3-days. This process continues until all the days, beginning from first to the last, are exhausted.

Example 2: Calculate three days moving average for the following record of attendance:

Week	Sun	Mon	Tue	Wed	Thu	Fri	Sat
1	24	55	28	45	51	54	60

Solution:

Week and Days	Attendance	3-days moving	
		Total	Average
Sun.	24	—	—
Mon.	55	107	107/3 = 35.67
Tue.	28	128	128/3 = 42.67
Wed.	45	124	124/3 = 41.33
Thu.	51	150	150/3 = 50.00
Fri.	54	165	165/3 = 55.00
Sat.	60	—	—

By adding the first three values, we get 107, which is placed at the center of these values *i.e.*, Monday and then dropping the first observation *i.e.*, 24 and adding the next 3 values, we get 128 and placed at the middle of these three values and so on. For average values, divide 3 days moving total by "3" which shows in last column of the table.

6.3(iv) Graphical Location of Median, Quartiles and Mode

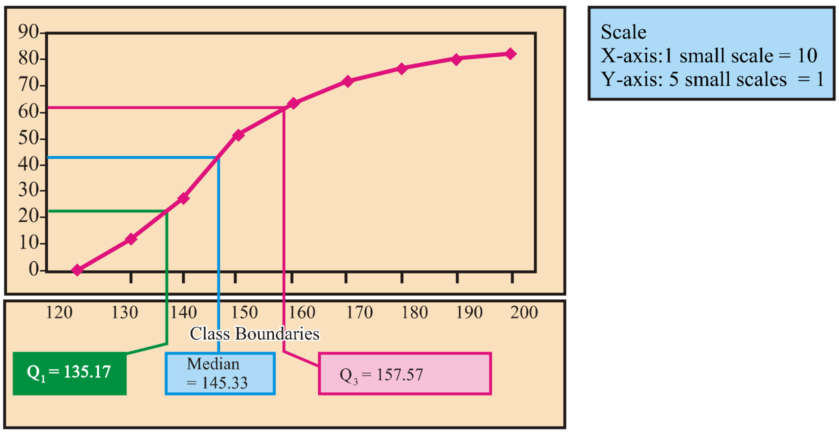
We explain the graphical location of Median, Quartiles and Mode by the help of following examples:

Example 1: For the following distribution, locate Median and Quartiles on graph:

Class Boundaries	Cumulative frequency
Less than 120	0
Less than 130	12
Less than 140	27
Less than 150	51
Less than 160	64
Less than 170	71
Less than 180	76
Less than 190	80
Less than 200	82

Solution:

We locate Median and Quartiles by using the following cumulative frequency polygon.



Finding

Q_1:

a) Find $(n/4)^{th}$ observation which is 82/4=20.5.

b) On the graph locate 20.5 on y-axis and draw a horizontal line segment parallel to x-axis touching the polygon.

c) Draw a vertical line segment from this point touching x-axis.

d) Read the value of first quartile at the point where the line segment meets x-axis which is 135.17.

Finding

Q_2 or Median :

a) Find $2(^{n}/4)^{th}$ observation which is 2(82/4)=41.

b) On the graph locate 41 on y-axis and draw a horizontal line segment parallel to x-axis touching the polygon.

c) Draw a vertical line segment from this point touching x-axis.

d) Read the value of Median at the point where the line segment meets x-axis which is 145.33.

Finding

Q_3 :

a) Find $3(n/4)^{th}$ observation which is 3(82/4)=61.5.

b) On the graph locate 61.5 on y-axis and draw a horizontal line segment parallel to x-axis touching the polygon.

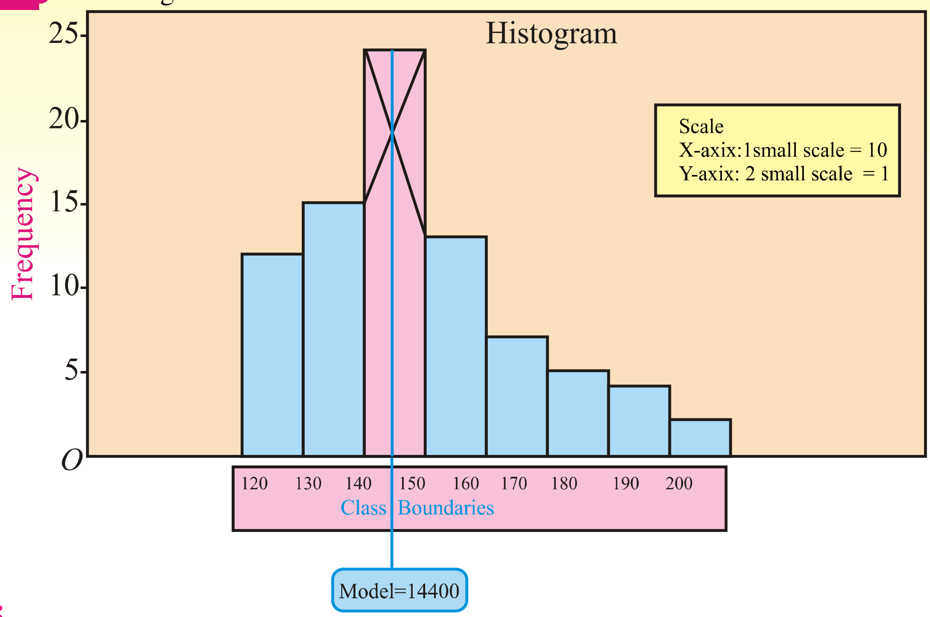
c) Draw a vertical line segment from this point touching x-axis.

d) Read the value of Median at the point where the line segment meets x-axis which is 157.57.

Example 2: For the following distribution locate Mode on graph.

Salaries in Rupees	No. of teachers
120 —— 130	12
130 —— 140	15
140 —— 150	24
150 —— 160	13
160 —— 170	7
170 —— 180	5
180 —— 190	4
190 —— 200	2

Solution: On Histogram the mode is located on X-axis as shown below:



Steps:

- Determine the rectangle having the highest peak indicating the modal class.
- Draw a line segment from the top left corner of the rectangle to the top left corner of the succeeding rectangle.
- Draw another line segment from the top right corner of the rectangle to the top right corner of the preceding rectangle.
- Drop perpendicular from the top of the rectangle to the x-axis passing through the point of intersection of the two line segments.
- Read the value at the point where the perpendicular meets the x-axis. This is the Mode of the data which is 144.

EXERCISE 6.2

1. What do you understand by measures of central tendency.
2. Define Arithmetic mean, Geometric mean, Harmonic mean, mode and median.
3. Find arithmetic mean by direct method for the following set of data:

   (i) 12, 14, 17, 20, 24, 29, 35, 45.

   (ii) 200, 225, 350, 375, 270, 320, 290.
4. For each of the data in Q. no 3., compute arithmetic mean using indirect method.

5. The marks obtained by students of class XI in mathematics are given below. Compute arithmetic mean by direct and indirect methods.

Classes / Groups	Frequency
0—9	2
10—19	10
20—29	5
30—39	9
40—49	6
50—59	7
60—69	1

6. The following data relates to the ages of children in a school. Compute the mean age by direct and short-cut method taking any provisional mean. (Hint. Take A = 8)

Class limits	Frequency
4—6	10
7—9	20
10—12	13
13—15	7
Total	50

Also Compute Geometric mean and Harmonic mean.

7. The following data shows the number of children in various families. Find mode and median.

9, 11, 4, 5, 6, 8, 4, 3, 7, 8, 5, 5, 8, 3, 4, 9, 12, 8, 9, 10, 6, 7, 7, 11, 4, 4, 8, 4, 3, 2, 7, 9, 10, 9, 7, 6, 9, 5.

8. Find Modal number of heads for the following distribution showing the number of heads when 5 coins are tossed. Also determine median.

X (number of heads)	Frequency (number of times)
1	3
2	8
3	5
4	3
5	1

9. The following frequency distribution the weights of boys in kilogram. Compute mean, median, mode.

Class Intervals	Frequency
1—3	2
4—6	3
7—9	5
10—12	4
13—15	6
16—18	2
19—21	1

10. A student obtained the following marks at a certain examination: English 73, Urdu 82, Mathematics 80, History 67 and Science 62.
   (i) If the weights accorded these marks are 4, 3, 3, 2 and 2, respectively, what is an appropriate average mark?
   (ii) What is the average mark if equal weights are used?

11. On a vacation trip a family bought 21.3 liters of petrol at 39.90 rupees per liter, 18.7 liters at 42.90 rupees per liter, and 23.5 liters at 40.90 rupees per liter. Find the mean price paid per liter.

12. Calculate simple moving average of 3 years from the following data:

Years	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Values	102	108	130	140	158	180	196	210	220	230

13. Determine graphically for the following data and check your answer by using formulae.
   (i) Median and Quartiles using cumulative frequency polygon.
   (ii) Mode using Histogram.

Class Boundaries	Frequency
10—20	2
20—30	5
30—40	9
40—50	6
50—60	4
60—70	1

6.4 Measures of Dispersion

Statistically, Dispersion means the spread or scatterness of observations in a data set. The spread or scatterness in a data set can be seen in two ways:

(i) The spread between two extreme observations in a data set.
(ii) The spread of observations around an average say their arithmetic mean.

The purpose of finding Dispersion is to study the behavior of each unit of population around the average value. This also helps in comparing two sets of data in more detail.

The measures that are used to determine the degree or extent of variation in a data set are called **Measures of Dispersion**.

We shall discuss only some important absolute measures of dispersion now.

(i) Range

Range measures the extent of variation between two extreme observations of a data set. It is given by the formula:

Range = $X_{max} - X_{min} = X_m - X_0$

where $X_{\max} = X_m$: the maximum, highest or largest observation.

$X_{\min} = X_0$: the minimum ,lowest or smallest observation.

The formula to find range for grouped continuous data is given below:

Range = (Upper class boundary of last group)-(lower class boundary of first group

Find Range for the following weights of students:

110, 109, 84, 89, 77, 104, 74, 97, 49, 59, 103, 62.

Solution: Given that $X_m = 110$, $X_0 = 49$, Range = 110 – 49 = 61

Example 2: Find the Range for the following distribution.

Classes / Groups	f
10 — 19	10
20 — 29	7
30 — 39	9
40 — 49	6
50 — 59	7
60 — 69	1
Total	40

Solution: We find class boundaries and class marks for the given data as follows:

Class limits	Class Boundaries	Frequency
10 — 19	9.5—19.5	10
20 — 29	19.5—29.5	7
30 — 39	29.5—39.5	9
40 — 49	39.5—49.5	6
50 — 59	49.5—59.5	7
60 — 69	59.5—69.5	1
		$\Sigma f = 40$

Range = 69.5 – 9.5 = 60

(ii) Variance

Variance is defined as the mean of the squared deviations of x_i (i = 1, 2, ….., n) observations from their arithmetic mean. In symbols,

$$\text{Variance of } X = \text{Var}(X) = S^2 = \frac{\sum (X-\bar{X})^2}{n}$$

(iii) Standard Deviation

Standard deviation is defined as the positive square root of mean of the squared deviations of X_i (i = 1, 2, ….., n) observations from their arithmetic mean. In symbols we write,

$$\text{Standard Deviation of X} = \text{S.D(X)} = S = \sqrt{\frac{\sum (X-\bar{X})^2}{n}}$$

Computation of Variance and Standard Deviation

We use the following formulae to compute Variance and Standard Deviation for Ungrouped and Grouped Data.

Ungrouped Data

The formula of Variance is given by:

$$\text{Var}(X) = S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2$$

and Standard deviation

$$\text{S.D}(X) = S = \sqrt{\left[\frac{\sum X^2}{n} - \left(\frac{\sum X}{n}\right)^2\right]}$$

Example 3: The marks of six students in Mathematics are as follows. Determine Variance and Standard deviation.

Student No	1	2	3	4	5	6
Marks	60	70	30	90	80	42

Solution: Let X = marks of a student. We make the following computations for finding Variance and Standard deviation.

X	X^2	$X-\bar{X}$	$(X-\bar{X})^2$
60	3600	–2	4
70	4900	8	64
30	900	–32	1024
90	8100	28	784
80	6400	18	324
42	1764	–20	400
$\Sigma X = 372$	$\Sigma X^2 = 25664$	$\Sigma (X-\bar{X}) = 0$	$\Sigma (X-\bar{X})^2 = 2600$

So, $\bar{X} = \frac{372}{6} = 62$ marks

and $\text{Var}(X) = S^2 = \frac{2600}{6} \approx 433.3333$ (square marks)

Using computational formula

$$\text{Var}(X) = S^2 = \frac{25664}{6} - \left(\frac{372}{6}\right)^2$$

$$\approx 4277.3333 - 3844 = 433.3333 \text{ (square marks)}$$

$$\text{S.D}(X) = S \approx \sqrt{4277.3333 - 3844} = \sqrt{433.3333}$$

$$\approx 20.81666 \text{ Marks}$$

Grouped Data

The formula of Variance is given by:

$$S^2 = \frac{\Sigma fX^2}{\Sigma f} - \left(\frac{\Sigma fX}{\Sigma f}\right)^2$$

and standard deviation

$$S = \sqrt{\left[\frac{\Sigma fX^2}{\Sigma f} - \left(\frac{\Sigma fX}{\Sigma f}\right)^2\right]}$$

Example 4: For the following data showing weights of toffee boxes in gm. determine the variance and standard deviation by using direct methods.

X (gm)	f
4.5	2
14.5	10
24.5	5
34.5	9
44.5	6
54.5	7
64.5	1

Solution: We make the following computations:

X	f	$X-\bar{X}$	$(X-\bar{X})^2$	$f(X-\bar{X})^2$	fX	fX^2
4.5	2	–28	784	1568	9	40.5
14.5	10	–18	324	3240	145	2102.5
24.5	5	–8	64	320	122.5	3001.25
34.5	9	2	4	36	310.5	10712.25
44.5	6	12	144	864	267	11881.5
54.5	7	22	484	3388	381.5	20791.75
64.5	1	32	1024	1024	64.5	4160.25
Total			$\Sigma(X-\bar{X})^2 = 2600$	$\Sigma f(X-\bar{X})^2 = 10440$	$\Sigma fX = 1300$	$\Sigma fX^2 = 52690$

Using definitional formula

$$S^2 = \frac{10440}{40} = 261 \text{ sq. gm.}$$

Using computational formula we get,

$$S^2 = \frac{52690}{40} - \left(\frac{1300}{40}\right)^2 = 1317.25 - (32.5)2 = 1317.25 - 1056.25 = 261 \text{ sq. gm.,}$$

and standard deviation is given by,

$$S = \sqrt{\frac{10440}{40}} = \sqrt{261} = 16.155 \text{ gm.}$$

$$S = \sqrt{\frac{52690}{40} - \left(\frac{1300}{40}\right)^2} = \sqrt{261} = 16.155 \text{ gm.}$$

Example 5: Compare the variation about mean for the two groups of students who obtained the following marks in Statistics:

X = Marks (section A)	Y = Marks section B
60	62
70	62
30	65
90	68
80	67
40	48

Solution: In order to compare variation about mean we compute standard deviation for the two groups as follows:

X	Y	$X-\bar{X}$	$(X-\bar{X})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
60	62	–2	4	0	0
70	62	8	64	0	0
30	65	–32	1024	3	9
90	68	28	784	6	36
80	67	18	324	5	25
40	48	–20	400	–14	196
$\Sigma X = 370$	$\Sigma Y = 372$		$\Sigma(X-\bar{X})^2 = 2600$		$\Sigma(Y-\bar{Y})^2 = 266$

Mean for group A = $\bar{X} = \frac{\sum X}{n} = \frac{370}{6} = 61.67 \simeq 62$ Marks

Mean for group B = $\bar{Y} = \frac{\sum Y}{n} = \frac{372}{6} = 62$ Marks

$$S.D\ (X) = \sqrt{\frac{\sum(X-\bar{X})^2}{n}} = \sqrt{\frac{2600}{6}} = \sqrt{433.333} = 20.82 \text{ Marks}$$

$$S.D\ (Y) = \sqrt{\frac{\sum(Y-\bar{Y})^2}{n}} = \sqrt{\frac{266}{6}} = \sqrt{44.333} = 6.66 \text{ Marks}$$

Comment: We note that the variation in Group B is smaller than that of Group A. This implies that the marks of students in Group B are closer to their Mean than that of Group A.

EXERCISE 6.3

1. What do you understand by Dispersion?
2. How do you define measures of dispersion?
3. Define Range, Standard deviation and Variance.
4. The salaries of five teachers in Rupees are as follows.

   11500, 12400, 15000, 14500, 14800.

   Find Range and standard deviation.
5. a- Find the standard deviation "*S*" of each set of numbers:

   (*i*) 12, 6, 7, 3, 15, 10, 18, 5

   (*ii*) 9, 3, 8, 8, 9, 8, 9, 18.

   b- Calculate variance for the data: 10, 8, 9, 7, 5, 12, 8, 6, 8, 2.
6. The length of 32 items are given below. Find the mean length and standard deviation of the distribution.

Length	20–22	23–25	26–28	29–31	32–34
Frequency	3	6	12	9	2

7. For the following distribution of marks calculate Range.

Marks in percentage	**Frequency/ (No of Students)**
31 — 40	28
41 — 50	31
51 — 60	12
61 — 70	9
71 — 75	5

MISCELLANEOUS EXERCISES

1. Multiple Choice Questions

Three possible answers are given for the following questions. Tick (✓) the correct answer.

(i) A grouped frequency table is also called
(a) data (b) frequency distribution
(c) frequency polygon

(ii) A histogram is a set of adjacent
(a) squares (b) rectangles
(c) circles

(iii) A frequency polygon is a many sided
(a) closed figure (b) rectangle
(c) square

(iv) A cumulative frequency table is also called
(a) frequency distribution (b) data
(c) less than cumulative frequency distribution

(v) In a cumulative frequency polygon frequencies are plotted against
(a) midpoints (b) upper class boundaries
(c) class limits

(vi) Arithmetic mean is a measure that determines a value of the variable under study by dividing the sum of all values of the variable by their
(a) number (b) group
(c) denominator

(vii) A Deviation is defined as a difference of any value of the variable from a
(a) constant (b) histogram
(c) sum

(viii) A data in the form of frequency distribution is called
(a) Grouped data (b) Ungrouped data
(c) Histogram

(ix) Mean of a variable with similar observations say constant k is
(a) negative (b) k itself
(c) zero

(x) Mean is affected by change in
(a) value (b) ratio
(c) origin

(xi) Mean is affected by change in
(a) place (b) scale
(c) rate

(xii) Sum of the deviations of the variable X from its mean is always
(a) zero (b) one
(c) same

(xiii) The n[th] positive root of the product of the $x_1, x_2, x_3, ..., x_n$ observations is called

(a) Mode (b) Mean

(c) Geometric mean

(xiv) The value obtained by reciprocating the mean of the reciprocal of $x_1, x_2, x_3, ..., x_n$ observations is called

(a) Geometric mean (b) Median

(c) Harmonic mean

(xv) The most frequent occurring observation in a data set is called

(a) mode (b) median

(c) harmonic mean

(xvi) The measure which determines the middlemost observation in a data set is called

(a) median (b) mode

(c) mean

(xvii) The observations that divide a data set into four equal parts are called

(a) deciles (b) quartiles

(c) percentiles

(xviii) The spread or scatterness of observations in a data set is called

(a) average (b) dispersion

(c) central tendency

(xix) The measures that are used to determine the degree or extent of variation in a data set are called measures of

(a) dispersion (b) central tendency

(c) average

(xx) The extent of variation between two extreme observations of a data set is measured by

(a) average (b) range

(c) quartiles

(xxi) The mean of the squared deviations of x_i $(i = 1, 2, ...,n)$ observations from their arithmetic mean is called

(a) variance (b) standard deviation

(c) range

(xxii) The positive square root of mean of the squared deviations of X_i $(i = 1, 2, ..., n)$ observations from their arithmetic mean is called

(a) harmonic mean (b) range

(c) standard deviation

2. Write short answers of the following questions.

(i) Define class limits.

(ii) Define class mark.

(iii) What is cumulative frequency?

(iv) Define a frequency distribution.

(v) What is a Histogram?

(vi) Name two measures of central tendency.
(vii) Define Arithmetic mean.
(viii) Write three properties of Arithmetic mean.
(ix) Define Median.
(x) Define Mode?
(xi) What do you mean by Harmonic mean?
(xii) Define Geometric mean.
(xiii) What is Range?
(xiv) Define Standard deviation.

SUMMARY

- **Range** is the difference between *maximum* and *minimum* observation.
- The minimum and the maximum values defined for a class or group are called **class limits**.
- The total of frequency up to an upper class limit or boundary is called the **cumulative frequency**.
- A **frequency distribution** is a tabular arrangement classifying data into different groups.
- A **Histogram** is a graph of adjacent rectangles constructed on *XY*-plane. A **cumulative frequency polygon** or **ogive** is a graph of less than cumulative frequency distribution.
- **Arithmetic mean** is a measure that determines a value of the variable under study by dividing the sum of all values of the variable by their number.
- A **Deviation** is defined as 'a difference of any value of the variable from any constant'. $D_i = x_i - A$.
- **Geometric mean** of a variable X is the n^{th} positive root of the product of the $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ observations.
- **Harmonic mean** refers to the value obtained by reciprocating the mean of the reciprocal of $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ observations.
- **Mode** is defined as the most frequent occurring observation of the variable or data.
- **Median** is the measure which determines the middlemost observation in a data set.
- Statistically, **Dispersion** means the spread or scatterness of observations in a data set.
- **Range** measures the extent of variation between two extreme observations of a data set.
- **Variance** is defined as the mean of the squared deviations of x_i (i = 1, 2, ….., n) observations from their arithmetic mean.
- **Standard deviation** is defined as the positive square root of mean of the squared deviations of x_i (i = 1, 2, ….., n) observations from their arithmetic mean.