CHAPTER 24

STATISTICS

INTRODUCTION

Statistics is a branch of mathematics which deals with the collection, classification, presentation, analysis and interpretation of numerical data.

Data

The word 'data' means information in the form of numerical figures or a set of given facts.

Ex. The percentage of marks scored by 10 pupils of a class in a test are :

36, 80, 65, 75, 94, 48, 12, 64, 88 and 98. The set of these figures is the data related to the marks obtained by 10 pupils in a class test. Statistical data are of two types (i) Primary, (ii) secondary.

Primary Data

The data which is collected by the investigator with a definite plan or design in mind is called Primary Data.

Secondary Data

When the data is gathered from some sources which already had stored for some purpose, then the data is called secondary data.

Data as such can be qualitative or quantitative in nature. If one speaks of honesty, beauty, colour, etc. the data is qualitative while height, weight, distance, marks, etc. are quantitative.

SOME BASIC TERMS

Raw data

The data which is collected for specific purpose and put as it is (without any arrangement) is called raw data or the data obtained in original form are called raw data or ungrouped data. Each entry in raw data is known as an observation.

Ex : Runs scored by batsman in a T-20 cricket match are:

17, 20, 15, 42, 25, 17, 23, 18, 5, 15

Range of raw data

The difference between the highest and lowest values of given data is called range. In the above case,

Highest Score = 42, Lowest score = 5

∴ Range = Highest Score – Lowest score = 42 – 5 = 37

Variable or Variate

A characteristics that varies in magnitude from obeservation to observation.

Ex. weight, height, income, age, etc., are variables.

Frequency

The number of times a particular observation occurs is called its frequency.

Ex. Marks obtained by 20 students of a class in a unit test are:

8, 5, 10, 6, 12, 18, 10, 5, 11, 13, 4, 3, 9, 11, 2, 10, 10, 19, 5, 7

By observation, we can see that 5, 10 and 11 occurs three times, four times and two times respectively. Hence, frequency of 5, 10 and 11 are 3, 4 and 2 respectively.

PRESENTATION OF DATA

Frequency Distribution Table

There are two types of frequency distribution table.

(i) Discrete (or Ungrouped) Frequency Distribution Table

Consider the marks obtained by 20 students in mathematics annual examination of class VIII (Marks are out of 100) :

60, 70, 59, 60, 50, 58, 62, 56, 59, 59, 58, 70, 58, 62, 50, 58, 58, 50, 62, 56

To make the data easily understandable we write it as in the following table.

Marks obtained	50	56	58	59	60	62	70
No. of students (Frequency)	3	2	5	3	2	3	2

The above table is called Discrete (or Ungrouped) Frequency Distribution Table or simply a Frequency Distribution Table.

(ii) Grouped Frequency Distribution Table

When our data is large e.g. we are given marks out of 100 of 60 students. To make the data more sensible, we condense the data into groups like 50 – 55, ..., 90 – 95 (Assuming that our data is from 50 to 92) as given below.

Marks obtained (Groups or class Intervals)	No. of Students (frequency)
50 – 55	8
55 – 60	4
60 – 65	12
65 – 70	5
70 – 75	2
75 – 80	6
80 – 85	8
85 – 90	5
90 – 95	10
Total	**60**

The above type of table is called Grouped Frequency Distribution Table.

Least number of any class interval is called the lower limit of that class interval. And greatest number of any class interval is called the upper limit of that class interval. Ex. 50 and 55 are the lower and upper limit respectively of the class interval 50 – 55.

class size

The difference of upper limit and lower limit of any class interval is called class size or class width of the whole grouped frequency distribution. Class size of the above grouped frequency distribution is (55 – 50 = 5).

NOTE :

Upper limit of any class is not included in that class. Hence the class inverval 60-65 means 'equal and greater than 60 but less than 65.'

CUMULATIVE FREQUENCY DISTRIBUTION TABLE

There are two types of cumulative frequency distribution table.

(i) Discrete Frequency Distribution

In a discrete frequency distribution, the cumulative frequency of a particular value of the variable is the total of all the frequencies of the values of the variable which are less than or equal to the particular value.

Ex : Consider marks obtained by 45 students of class VIII in an assessment is given below:

Marks	0	5	10	14	16	18	20
No. of students (Frequency)	1	4	8	9	12	6	5

Then the commutative distribution table is given by

No. of children	No. of families (frequency)	Cumulative frequency
1	5	5
2	6	11 (= 5 + 6)
3	4	15 (= 11 + 4)
4	3	18 (= 15 + 3)
5	2	20 (= 18 + 2)
Total	**20**	

(ii) Grouped Frequency Distribution

In a grouped frequency distribution, the cumulative frequency of a class is the total of all frequencies up to that particular class.

To calculate cumulative frequencies, the classes should be written in ascending order.

Ex: The following table shows the number of patients getting medical treatment in a hospital on a day.

Age (in years) [Class Interval]	No. of Patients [Frequency]	Cumulative Frequency
10 – 20	90	90
20 – 30	50	140 (= 90 + 50)
30 – 40	60	200 (= 140 + 60)
40 – 50	80	280 (= 200 + 80)
50 – 60	50	330 (= 280 + 50)
60 – 70	30	360 (= 330 + 30)
Total	**360**	

The above table showing the cumulative frequency with class interval is called grouped frequency distribution table.

EXCLUSIVE AND INCLUSIVE CLASS INTERVAL

Class interval of the form 10 – 20, 20 – 30, 30 – 40, ……..; in which upper limit of any class interval coincides with the lower limit of the just next class interval, is called **Exclusive class Interval**.

Class interval of the form 10 – 19, 20 – 29, 30 – 39, …….; in which upper limit of any class interval does not coincides with the lower limit of the just next class interval, is called **Inclusive class Interval.**

RELATIVE FREQUENCY DISTRIBUTION

Relative frequencies are very useful for the comparison of two or more frequency distributions. The relative frequency of any class is the percentage of the total frequency.

i.e., $\text{Relative frequency} = \left(\frac{\text{Class frequency}}{\text{Total frequency}}\right) \times 100$

GRAPHICAL REPRESENTATION OF DATA

Data can be represented graphically in the form of:

(A) Bar graph or Bar chart

(B) Histogram and

(C) Frequency polygon

(D) Pie Chart

(E) Cumulative Frequency Curve (Ogive)

(A) BAR GRAPH

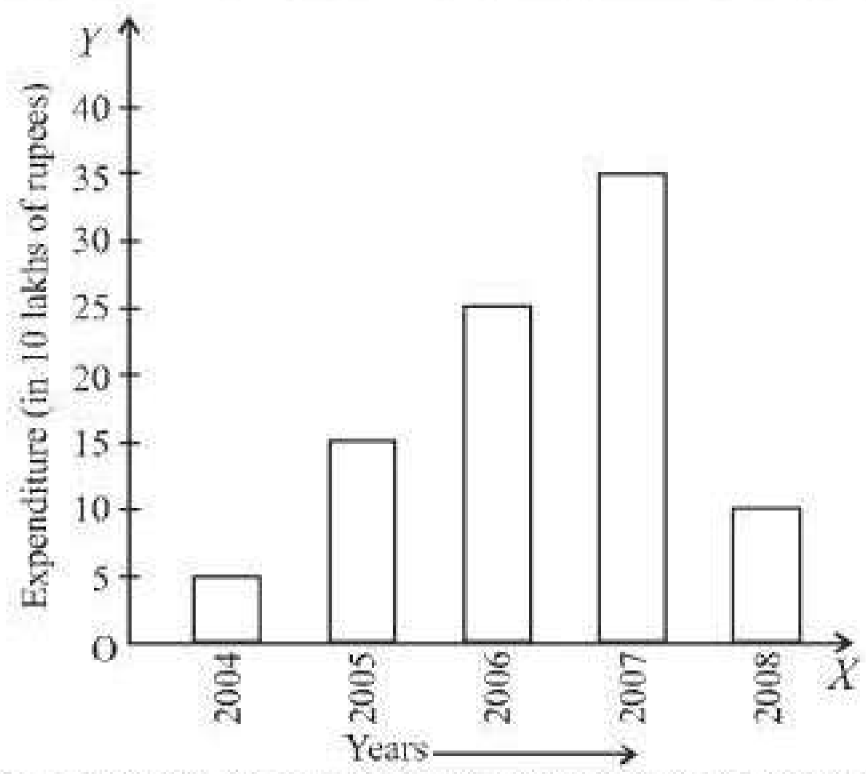
A bar graph is a pictorial representation of the numerical data by a number of bars of uniform width erected horizontally or vertically with equal spacing between them.

Bar graph is the simplest and popular graph to show ungrouped (or descrete) frequency distribution graphically. To draw bar graph, rectangular bars of uniform width with equal spaces in between them are drawn on a horizontal ray *OX*. Data (or items) are taken on *OX* ray. Each bar represent a data (or items). Height of each bar represents frequency of the corresponding data. Suitable scale of height of bars are shown on a vertical ray *OY*. See the following table :

Years	Expenditure (in 10 lakh of rupees) [frequency]
2004	5
2005	15
2006	25
2007	35
2008	10

The above table is the discrete frequency distribution table, which shows expenditure on health by ABC Pvt. Ltd. during five years (2004 to 2008).

Bar graph of the above discrete frequency distribution is shown below:



Bar graph of the expenditure of health by ABC Pvt. Ltd.

The important features of bar graphs are:

(a) It is used to represent unclassified frequency distributions.

(b) Frequency of a value of a variable is represented by a bar (rectangle) whose length (i.e.-height) is equal (proportional) to the frequency.

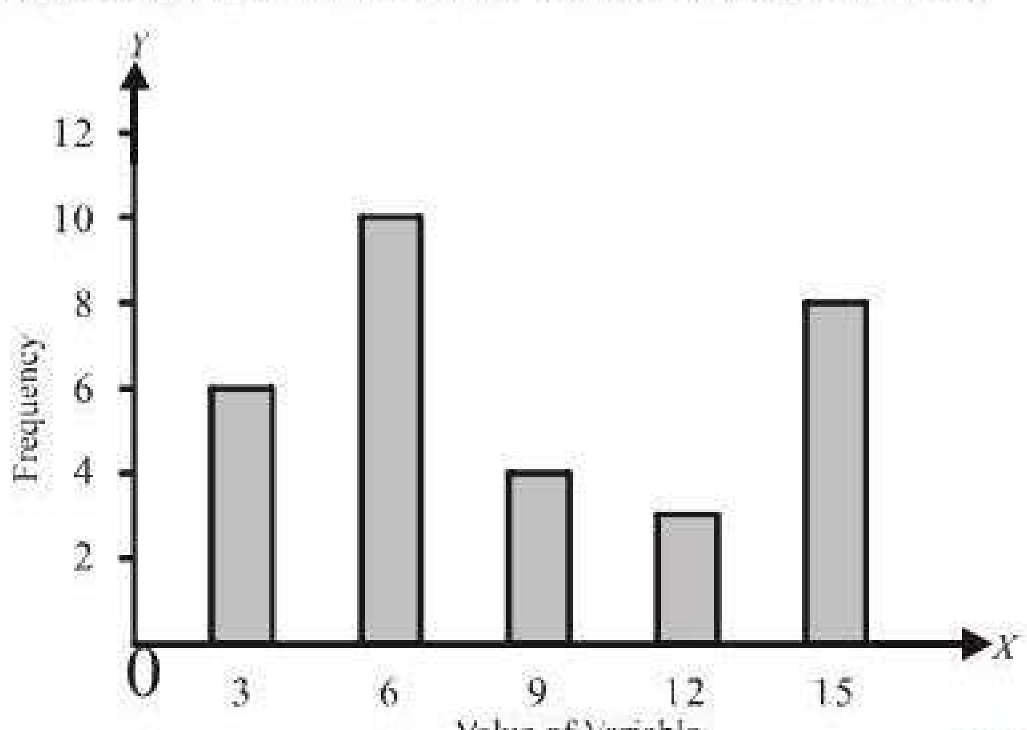
(c) The breadth of the bar is arbitrary and the breadth of all the bars are equal. The bars may or may not touch each other.

Example 1: Represent the following frequency distribution by bar graph :

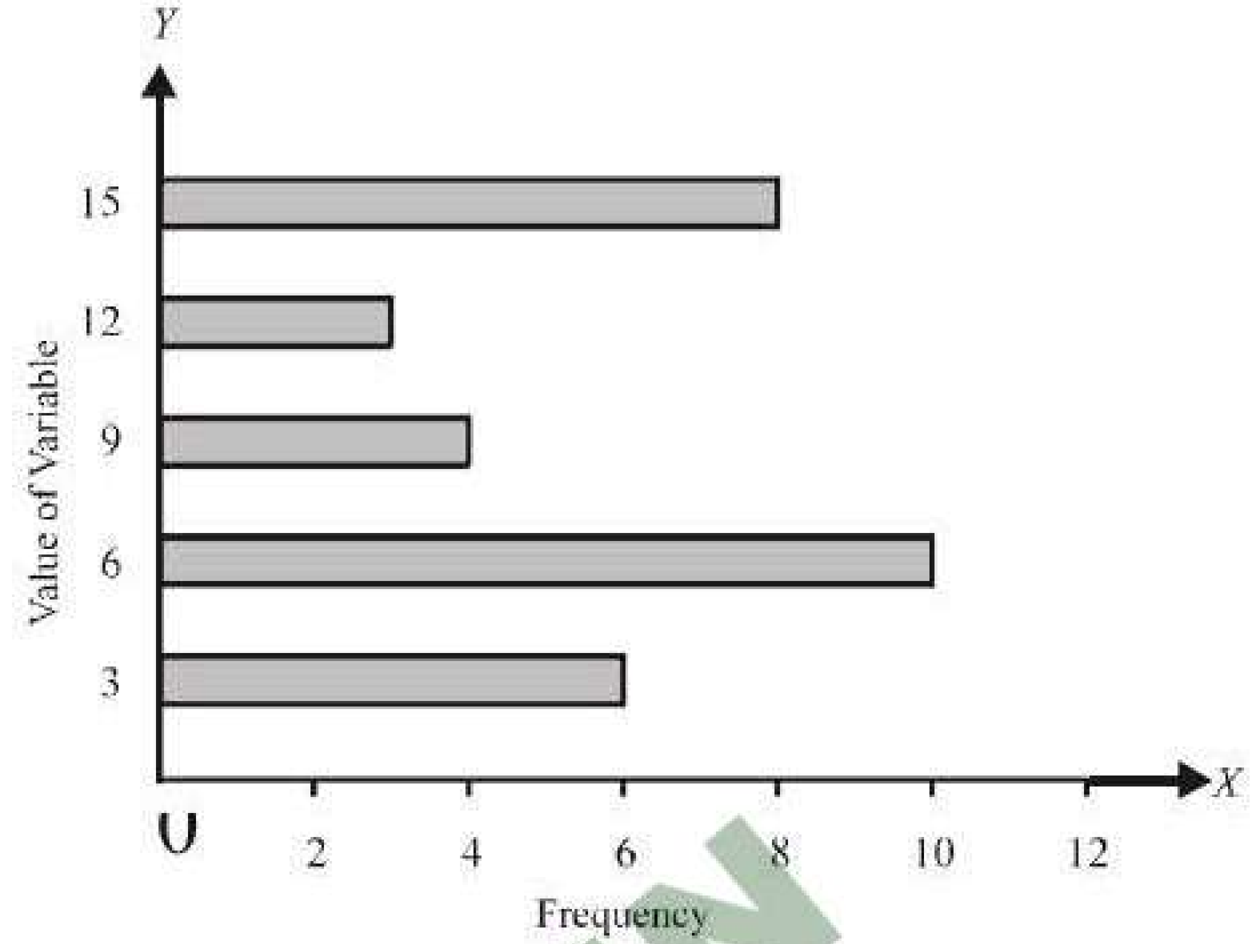
Value of variable	3	6	9	12	15
Frequency	6	10	4	3	8

Solution:

Either of the following bar graphs (fig (i) or fig (ii)) may be used to represent the above frequency distribution.



(fig. (i))



(fig. (ii))

(B) HISTOGRAM

A histogram is a graphical representation of a frequency distribution in the form of rectangles one after the other.

The bases of these rectangles represent the magnitudes of the variables of the class-boundaries. So these are taken along the x-axis. The heights of these represent the frequencies of the corresponding magnitudes of the variable of the class-boundaries.

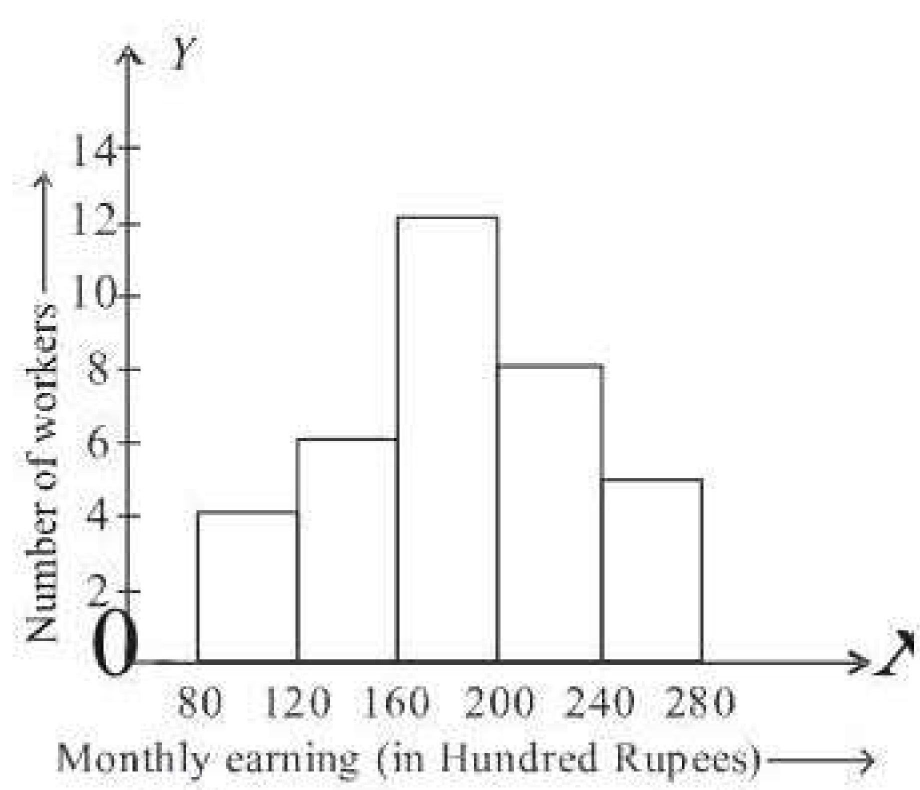
Monthly Earnings (in hundred rupees)	No. of Workers
80-120	4
120-160	6
160-200	12
200-240	8
240-280	5

To draw the histogram of the above frequency distribution we follow the following steps :

(a) On the horizontal axis, mark the class intervals with a uniform scale.

(b) On the vertical axis mark a scale to measure the height of bars which is equal to the frequencies, with a uniform scale.

(c) Construct rectangles with class intervals as bases and the corresponding frequencies as heights.



The above graph is the graph of histogram of the given grouped frequency distribution.

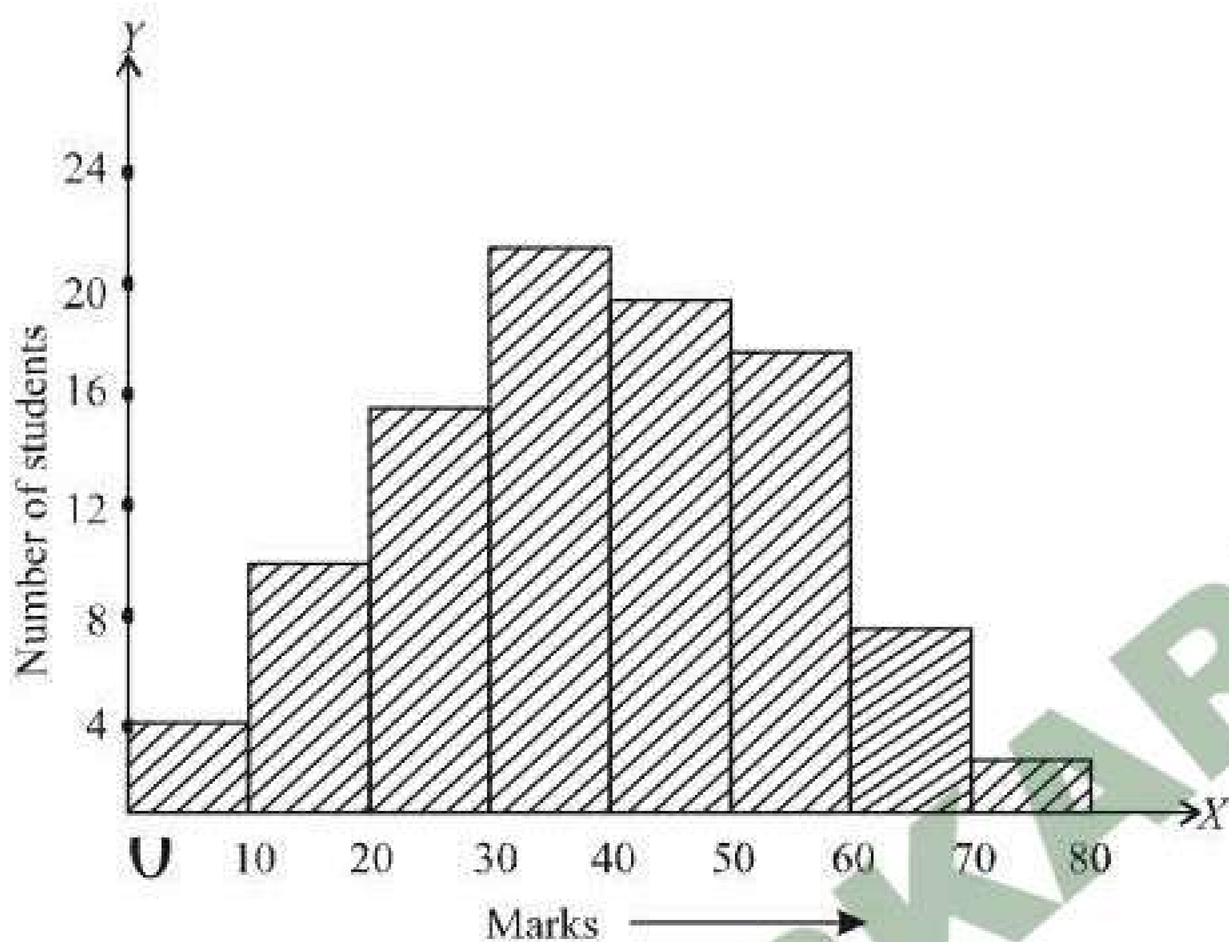
Note : Since the first class interval is start from 80 not from 0, hence we use a 'kink' or a break (z) on the x-ray just after 0.

Example 2 : The following table gives the marks scored by 100 students in an entrance examination.

Mark :	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80
No. of students (Frequency)	4	10	16	22	20	18	8	2

Represent this data in the form of a histogram.

Solution:



We represent the class limits along X-axis on a suitable scale and the frequencies along Y-axis on a suitable scale.

Taking class-intervals as bases and the corresponding frequencies as heights, we construct rectangles to obtain the histogram of the given frequency distribution as shown in figure.

(C) FREQUENCY POLYGON

It is another method of representing frequency distributions graphically. Also, it is used to represent classified or grouped data graphically.

It is a line graph of grouped frequency distribution plotted between class marks and frequencies. It can be obtained in two ways:

(a) By first drawing Histogram and

(b) Without drawing Histogram

(a) Steps of Drawing Frequency Polygon (By First Drawing Histogram):

(i) Draw the histogram from the given data

(ii) Obtain the mid-points of the upper horizontal sides of each rectangle.

(iii) Join these mid-points of the adjacent rectangles by dotted line segments.

(iv) Obtain the mid-point of two assumed class intervals of zero frequency. One before the first and other after the last class interval.

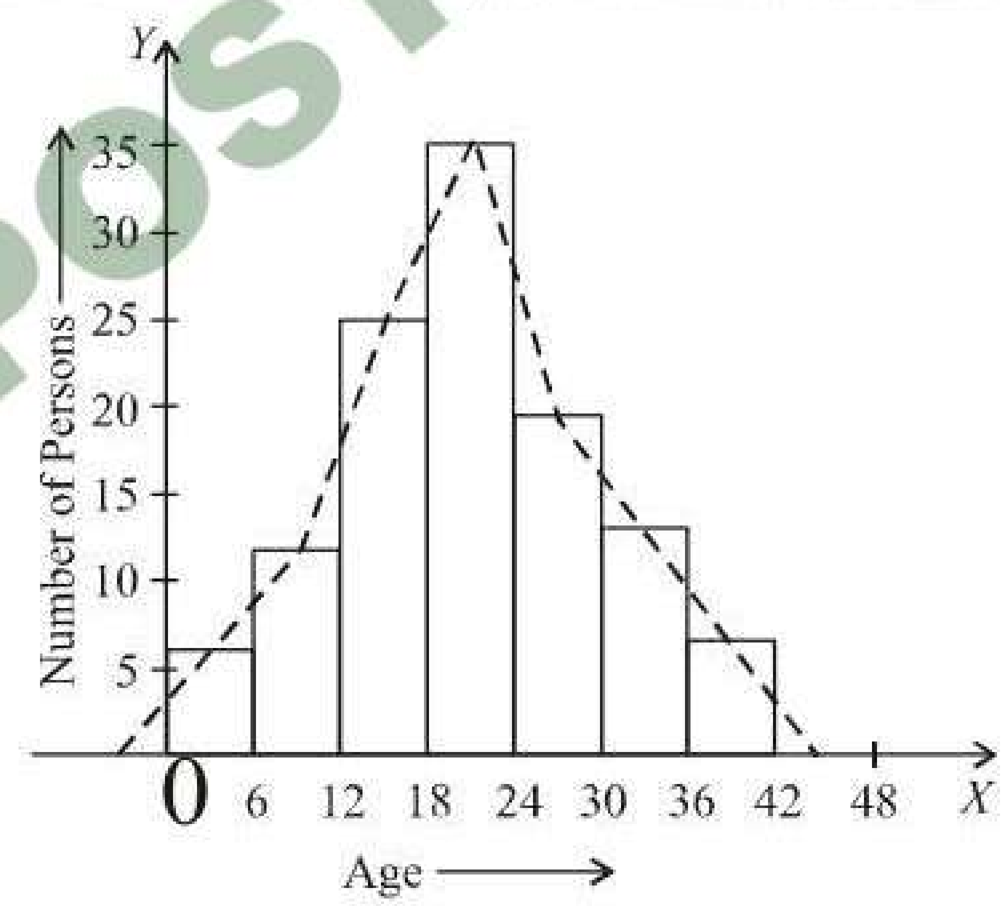
(v) Complete the polygon by joining the mid-point of class first to mid-point of its left adjacent class and mid-point of last class intervals to the mid point of its right adjacent class interval.

Example 3: Draw a histogram and a frequency polygon of the following data :

Age in years	0-6	6-12	12-18	18-24	24-30	30-36	36-42
No. of persons	6	11	25	35	18	12	6

Solution:

First we draw histogram of the given data, then we will locate mid-points of top horizontal side of rectangles.



Now, join these mid points by dotted line segments. Complete the polygon by joining the mid points of first interval to the mid-point of its left adjacent assumed class interval of zero frequency and mid-point of last class interval to its adjacent right assumed class interval of zero frequency.

(b) Steps of Drawing Frequency Polygon (Without Drawing Histogram)

(i) First calculate the class marks (mid points) $x_1, x_2, x_3, \ldots, x_4$ of the given class intervals.

(ii) Mark $x_1, x_2, x_3, \ldots, x_n$ along X-axis.

(iii) Mark respective frequencies $f_1, f_2, f_3, \ldots, f_n$ along Y-axis.

(iv) Plot the points $(x_1, f_1), (x_2, f_2), (x_3, f_3), \ldots (x_n, f_n)$

(v) Join points plotted in step (iv) by line segments.

(vi) Take two class intervals of zero frequency, one just before the first and other just after the last class interval given. Locate their mid points.

(vii) Complete the frequency polygon by joining the point (x , f) to the mid-point of the assumed class interval of zero frequency just before the first class interval and (x, f) to the assumed class interval of zero frequency just right to the last class interval.

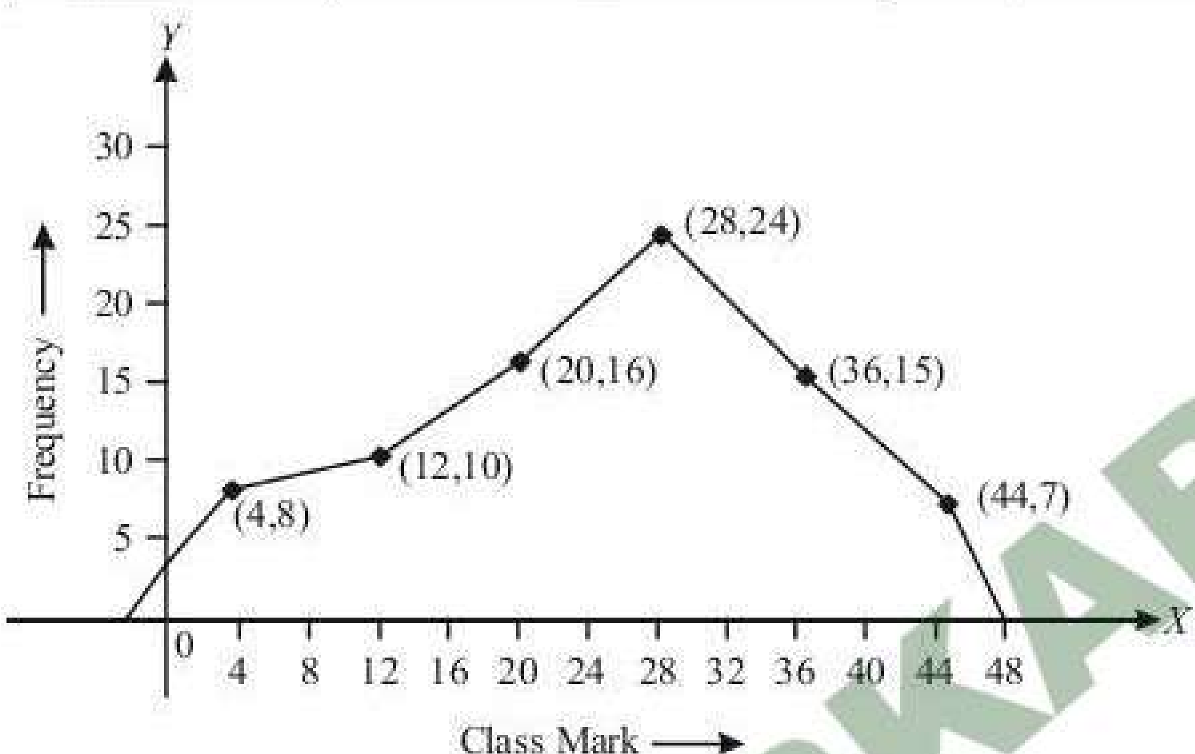
Example 4: Construct a frequency polygon for the following data without drawing the histogram:

Class Interval	0-8	8-16	16-24	24-32	32-40	40-48
Frequency	8	10	16	24	15	7

Solution:

Calculate class marks of given frequency distribution

Class Interval	Class Mark (or mid-point)	Frequency
0-8	4	8
8-16	12	10
16-24	20	16
24-32	28	24
32-40	36	15
40-48	44	7



Pie Chart or Circle Graph : A pie chart is a way of showing how something is shared or divided. It shows the relationship between a whole and its parts.

In pie charts, entire circle of appropriate radius represent total allocation. The circle is divided into sectors. The size of each sector is proportional to the information it represents.

$$\text{Central angle} = \frac{\text{Frequency}}{\text{Total Frequency}} \times 360^\circ$$

Following figure is the Pie chart which shows how 720 students usually come to school.

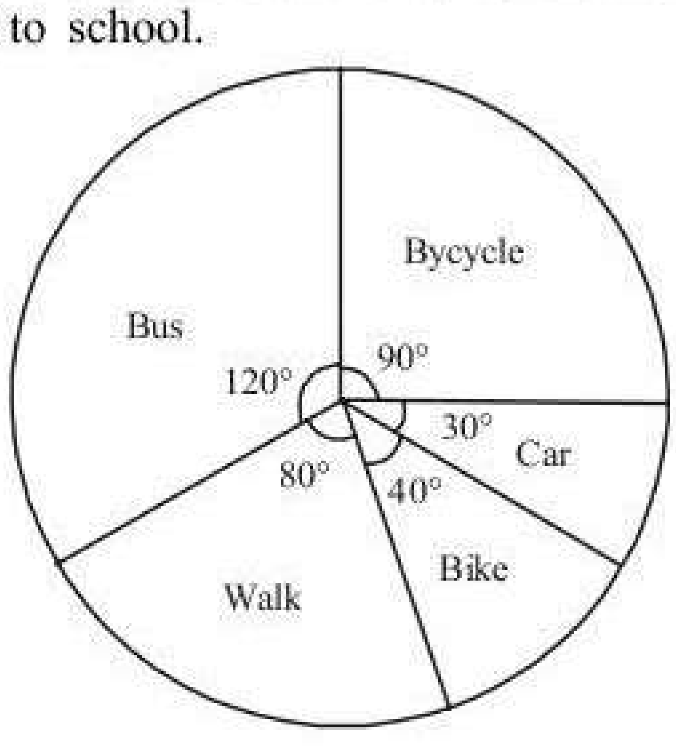


Fig.

The whole group of 720 students is represented by a circle. The total angle at the centre of the circle is 360°. The angle of 120° at the centre corresponds to students who come to school by bus. Since angle of 360° at the centre corresponds to whole group of 720° students.

$\therefore$ 1° corresponds to $\frac{720}{360} = 2$ students

$\therefore$ 120° corresponds to $\frac{720}{360} \times 120 = 240$ students

Similarly students coming by bike are $\frac{720}{360} \times 40 = 80$

Students walking to school are $\frac{720}{360} \times 80 = 160$ and so on.

Example 5: The way Mr. Goel spends his allowance is given below. Draw the pie chart.

Item	Percent
Hobby	10%
Lunch	20%
Recreation	45%
Savings	25%
Total	100%

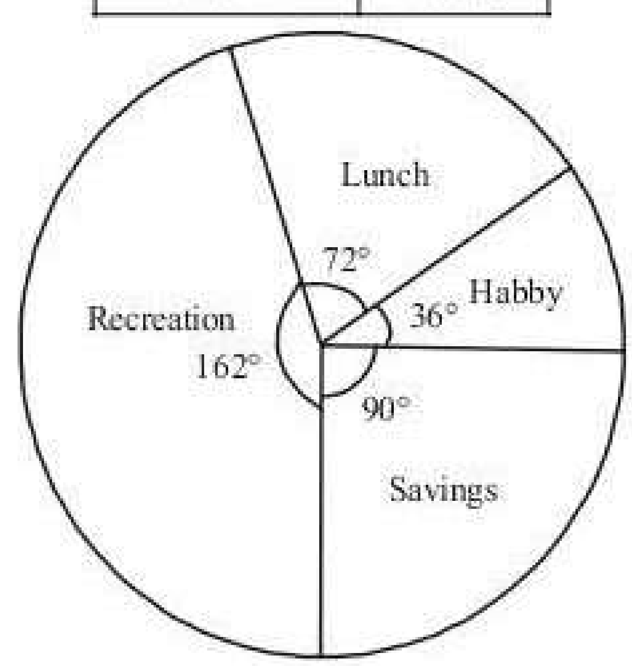


Fig.

Solution:

Item	%	Fractional Part	Central Angle
Hobby	10%	$\frac{10}{100} = \frac{1}{10}$	$\frac{1}{10} \times 360^\circ = 36^\circ$
Lunch	20%	$\frac{20}{100} = \frac{1}{5}$	$\frac{1}{5} \times 360^\circ = 72^\circ$
Recreation	45%	$\frac{45}{100} = \frac{9}{20}$	$\frac{9}{20} \times 360^\circ = 162^\circ$
Savings	25%	$\frac{25}{100} = \frac{1}{4}$	$\frac{1}{4} \times 360^\circ = 90^\circ$

1. Draw a circle of convenient radius.
2. Divide it into sectors of central angles 36°, 72°, 162° and 90°.
3. Write the items in the corresponding sectors.

The obtained **figure 5.4** given on prev. page is the required pie chart.

Cumulative Frequency Curve (OGIVE)

Cumulative frequency curve or an ogive is the graphical representation of a cumulative frequency distribution. There are two methods of constructing an ogive

(i) Less than method (ii) More than method

(i) Less Than Method

Following are the steps for construction of an ogive:

(a) Construct a cumulative frequency table of less than type.

(b) Mark upper class limits along X-axis.

(c) Mark the corresponding cumulative frequency along Y-axis.

(d) Plot the points (x_i, f_i), where x_i is the upper limit of a class and f_i is the corresponding c.f. and join them by a smooth free hand curve.

The curve we get is the **Less Than Type Ogive.**

Example *6:* **Draw an ogive for the following frequency distribution by less than method.**

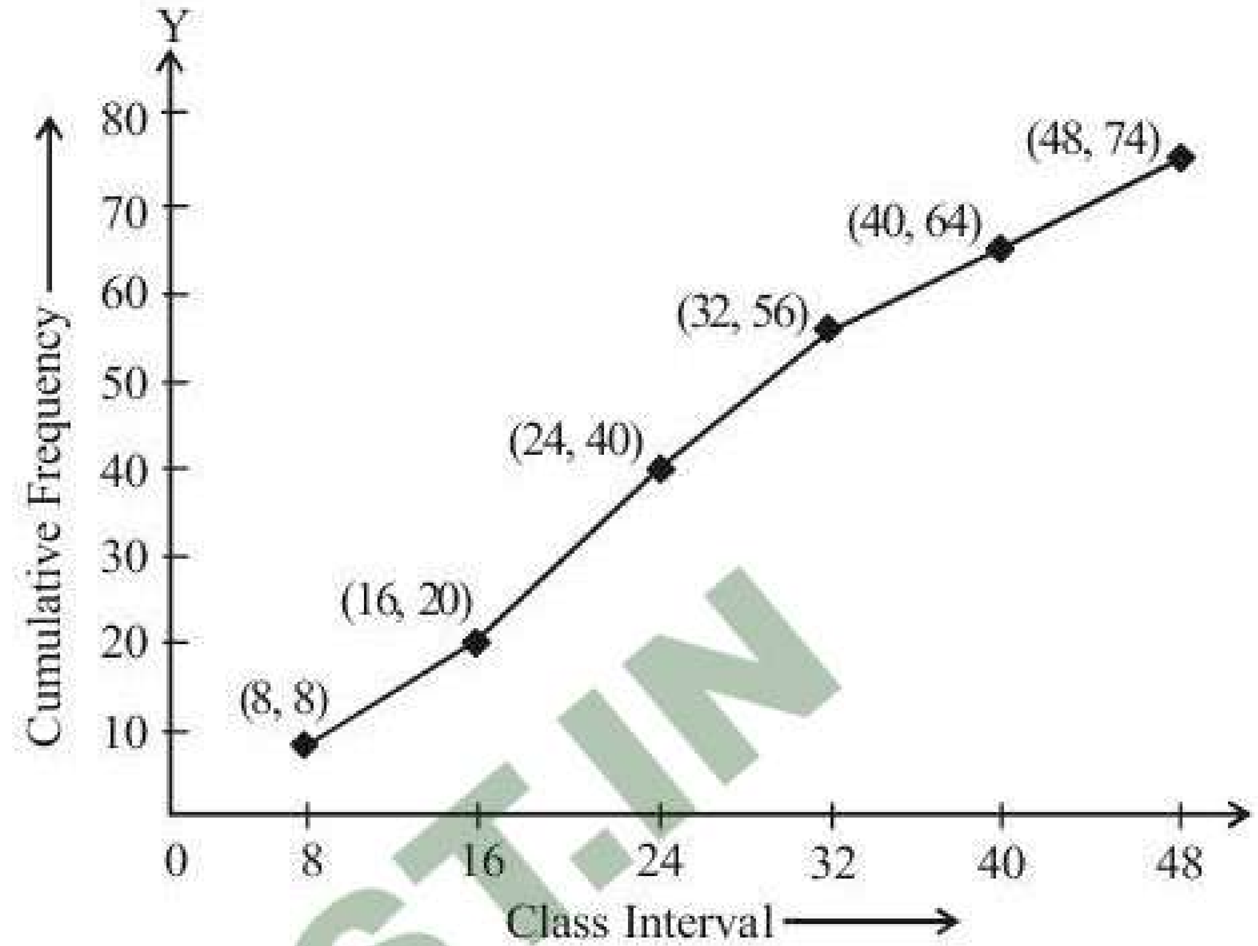
Class Interval	0-8	8-16	16-24	24-32	32-40	40-48
Frequency	8	12	20	16	8	10

Solution:

First we prepare a Less Than Type frequency distribution table.

Class Interval	Cumulative Frequency
Less than 8	8
Less than 16	8 + 12 = 20
Less than 24	20 + 20 = 40
Less than 32	40 + 16 = 56
Less than 40	56 + 8 = 64
Less than 48	64 + 10 = 74

Now, mark the upper class limit on X-axis and frequency along Y-axis. plot the points (8, 8), (16, 20), (24, 40), (32, 56), (40, 64) and (48, 74). Join the points by smooth free hand curve to get the required less than type ogive.



(ii) More Than Method

Following are the steps for constructing an ogive.

(a) Construct a more than type frequency distribution table

(b) Mark the lower class limit on X- axis.

(c) Mark the corresponding cumulative frequencies on Y- axis.

(d) Plot the points and join them by smooth free hand curve.

The obtained curve is the **More Than Type Ogive.**

Example *7:* **The frequency distribution of scores obtained by 230 candidates in a medical entrance test is as follows:**

Scores	400-450	450-500	500-550	550-600	600-650	650-700	700-750	750-800
No. of Candidate	20	35	40	32	24	27	18	34

Draw cumulative frequency curve or ogive by more than method.

Solution:

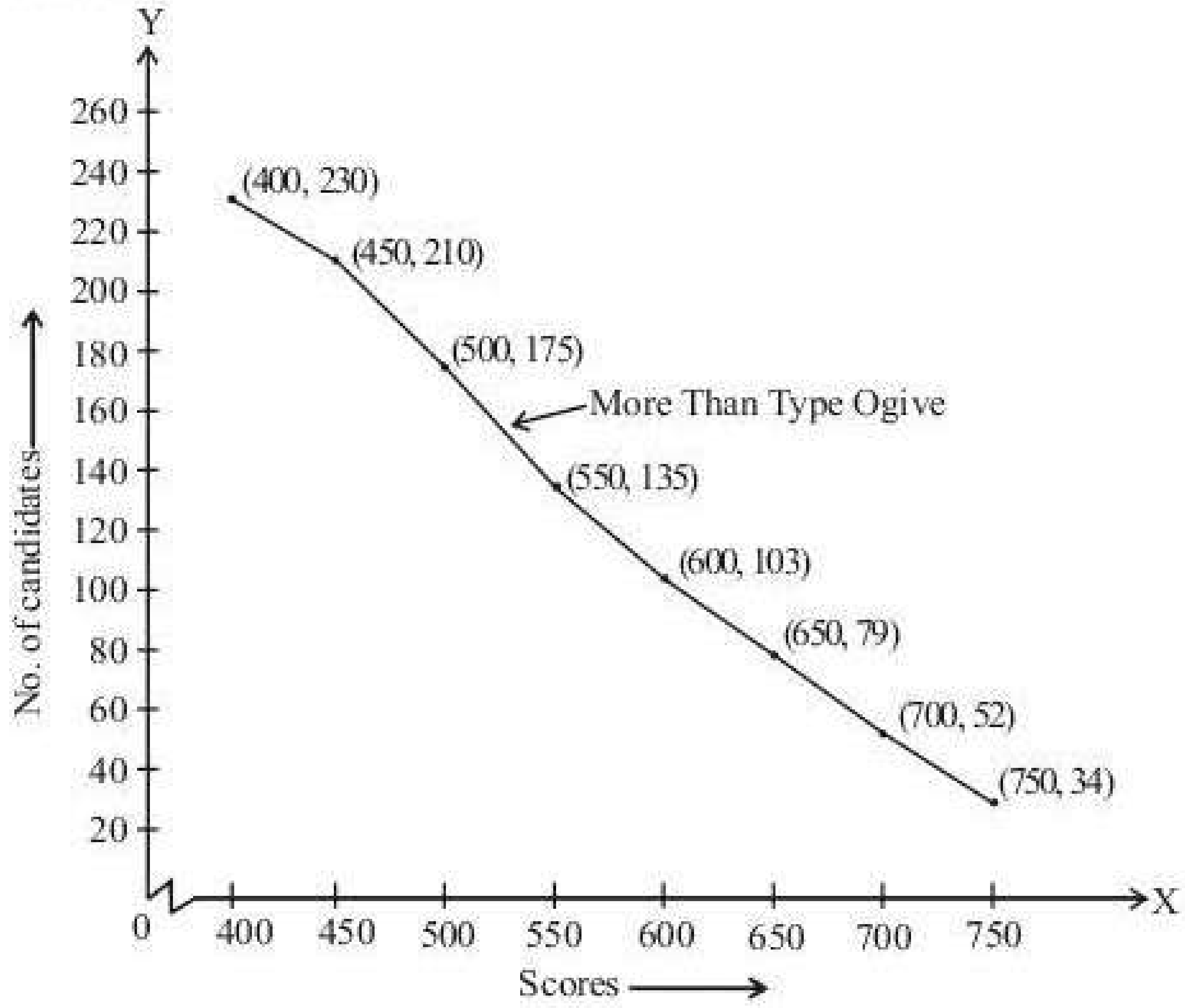
First convert the given frequency distribution table to More Than Type frequency distribution table.

Secors	No. of Candidates
More than or equal to 400	230
More than or equal to 450	230 – 20 = 210
More than or equal to 500	210 – 35 = 175
More than or equal to 550	175 – 40 = 135
More than or equal to 600	135 – 32 = 103
More than or equal to 650	103 – 24 = 79
More than or equal to 700	79 – 27 = 52
More than or equal to 750	52 – 18 = 34

Now mark the lower limits along X-axis and cumulative frequencies along Y-axis, and plot the points (400, 230), (450, 210), (500, 175), (550, 135), (600, 103), (650, 79), (700, 52), (750, 34).

Join the points listed above by smooth free hand curve to obtain the more than type ogive.

Scale:



MEASURES OF CENTRAL TENDENCY

If the data is very large, the user cannot get much information from these data or its associated frequency distribution. In this case, the information contained in data are represented by some numerical values, called averages. These averages are also called measures of central tendency or measures of location because they also give an idea about the concentration of the values in the central part of the distribution of data which describes the characteristics of the entire data or its associated frequency distribution.

Average can be categorised as follows:

(i) Mathematical averages
 (a) Arithmetic mean or mean
 (b) Geometric mean
 (c) Harmonic mean

(ii) Positional averages
 (a) Median
 (b) Mode

MEAN

Mean of a set of observations is their sum divided by number of observations. Mean is denoted by $\bar{x}$. The mean $\bar{x}$ of n observations $x_1, x_2, x_3, ..., x_n$ is given by

$$\bar{x} = \frac{x_1 + x_2 + x_3 + ... + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Example 8: Find mean of 21, 22, 24, 26, 32.

Solution:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5} = \frac{21 + 22 + 24 + 26 + 32}{5} = \frac{125}{5} = 25$$

Example 9: The mean of 6, 10, x and 12 is 8. Find the value of x.

Solution:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{6 + 10 + x + 12}{4} = \frac{28 + x}{4} \Rightarrow 8 = \frac{28 + x}{4}$$

$(\because \bar{x} = 8)$

$\Rightarrow 28 + x = 32 \Rightarrow x = 4$, value of x is 4

Some Important Results About Mean

(i) If each observation is increased by 'a', then the mean is also increased by 'a'. If is the mean of n observations $x_1, x_2, ..., x_n$, then the mean of observations $(x_1 + a), (x_2 + a), (x_3 + a), ..., (x_n + a)$ is $(\bar{x} + a)$.

(ii) If each observation is decreased by 'a', then mean is also decreased by 'a'. If $\bar{x}$ is the mean of n observations $x_1, x_2, ..., x_n$; then mean of observations $(x_1 - a), (x_2 - a), ..., (x_n - a)$ is $(\bar{x} - a)$.

(iii) If each observation is multiplied by a non-zero number 'a'. Then, mean is also multiplied by 'a'. If $\bar{x}$ is mean of n observations $x_1, x_2, ..., x_n$; then mean of $ax_1, ax_2, ..., ax_n$ is $a.\bar{x}$.

(iv) If each observation is divided by a non-zero number 'a', then mean is also divided by the non-zero number 'a'. If $\bar{x}$ is the mean of n observations $x_1, x_2, ..., x_n$, then mean of $\frac{x_1}{a}, \frac{x_2}{a}, ...\frac{x_n}{a}$, is $\frac{\bar{x}}{a}$.

(v) If is the mean of n observations $x_1, x_2, ... x_n$, then the algebraic sum of deviations from mean is zero. i.e.

$$\sum_{i=1}^{n}(x_i - \bar{x}) = 0$$

Combined Mean

If $\bar{x}_1, \bar{x}_2, ..., \bar{x}_k$ are the mean of k series having number of observations (or data) $n_1, n_2, ..., n_k$ respectively then the mean $\bar{x}$ of the composite series is given by

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + ... + n_k\bar{x}_k}{n_1 + n_2 + ... + n_k}$$

Weighted Arithmetic Mean

If $x_1, x_2, ..., x_n$ denote n values of the variable x and $w_1, w_2, ..., w_n$ denote respectively their weights, then their weighted mean is $\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$

Note : If the weight of each observation is 1, then weighted mean = arithmetic mean

Example 10: The mean income of a group of persons is ₹ 400. Another group of persons has mean income ₹ 480. If the mean income of all the persons in the two groups together is ₹ 430, then find ratio of the number of persons in the group.

Solution :

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1+n_2} \quad \because \bar{x}_1 = 400,\ \bar{x}_2 = 480,\ \bar{x} = 430$$

$$\therefore\ 430 = \frac{n_1(400)+n_2(480)}{n_1+n_2} \Rightarrow 30n_1 = 50n_2 \Rightarrow \frac{n_1}{n_2} = \frac{5}{3}$$

$\Rightarrow n_1 : n_2 = 5 : 3$

Hence, required ratio = 5 : 3

Mean of Discrete Frequency Distribution

Mean of discrete frequency distribution may be computed by any one of the following methods :

(i) Direct method

(ii) Assume mean method (Short–cut method)

(iii) Step–Deviation method

(i) Direct Method

For the discrete frequency distribution, we have

x_i	x_1	x_2	x_3	....	x_n
f_i	f_1	f_2	f_3	....	f_n

then arithmetic mean of values $x_1, x_2, ..., x_n$ is given by

$$\text{Mean } (\bar{x}) = \frac{f_1x_1 + f_2x_2 + ...+ f_nx_n}{f_1+f_2+.............+f_n} = \sum_{i=1}^{n}\frac{f_ix_i}{\Sigma f_i}$$

Example *11:* Find the arithmetic mean of the following frequency distribution :

x	4	7	10	13	16	19
f	7	10	15	20	25	30

Solution :

x_i	f_i	f_ix_i
4	7	28
7	10	70
10	15	150
13	20	260
16	25	400
19	30	570
	$\Sigma f_i = 107$	$\Sigma f_ix_i = 1478$

$$\bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{1478}{107} = 13.81$$

Example *12:* Find the value of *k* if mean of the following data is 14

x_i	5	10	15	20	25
f_i	7	k	8	4	5

Solution :

x_i	f_i	f_ix_i
5	7	35
10	k	$10k$
15	8	120
20	4	80
25	5	125
Total	$\Sigma f_i = 24+k$	$\Sigma f_ix_i = 360+10k$

$$\text{Now, } \bar{x} = \frac{\Sigma f_ix_i}{\Sigma f_i} \Rightarrow 14 = \frac{360+10k}{24+k}$$

$\Rightarrow 336 + 14k = 360 + 10k$

$\Rightarrow 14k - 10k = 360 - 336$

$\Rightarrow 4k = 24 \Rightarrow k = 6$

(ii) Assumed Mean Method (Short-cut Method) :

$\bar{x} = A + \frac{1}{N}\sum_{i=1}^{n} f_id_i$, where A is assumed mean, $d_i = x_i - A$

and $N = \sum_{i=1}^{n} f_i$

NOTE:

(i) The number 'A' is generaly chosen in such a way that the deviations (d_i) are small.

(ii) **Generally the middle data is considered as assumed mean.**

If there are two middle data, then you can take those middle data as assumed mean, whose frequency is greater. Let us understand this method by doing the following illustration.

Example *13:* Compute the mean height by Assumed mean method of plants from the following frequency distribution :

Height in cm	61	64	67	70	73
No. of plants	5	18	42	27	8

Solution :

Height (x_i)	**No. of plants (f_i)**	**$d_i = x_i - A$**	**f_id_i**
61	5	–6	–30
64	18	–3	–54
67 (A)	42	0	0
70	27	3	81
73	8	6	48
	$N = \Sigma f_i = 100$		$\Sigma f_id_i = 45$

Let assumed mean (A) = 67 (generally middle data)

We know, $\bar{x} = A + \frac{1}{N}\sum_{i=1}^{n} f_i d_i$

$\Rightarrow \bar{x} = 67 + \frac{45}{100} = 67 + 0.45 = 67.45$

(iii) Step Deviation Method

This method is used if the deviation d_i**'s** are divisible by any common number h.

$$\bar{x} = A + \left(\frac{\sum f_i u_i}{\sum f_i}\right) h \text{ where } u_i = \frac{x_i - A}{h}$$

NOTE:

(i) The step-deviation method will be convenient to apply if all the d_is have a common factor.

(ii) The mean obtained by all the three methods is the same.

Following illustrations will help to understand this method.

Example *14:* **Find the arithmetic mean by step-deviation method for the following frequency distribution :**

x_i	5	10	15	20	25	30	35	40	45	50
f_i	20	43	75	67	72	45	39	9	8	6

Solution :

First of all we construct the calculation table by taking 25 as assumed mean (A).

x_i	Frequency (f_i)	$u_i = \frac{x_i - 25}{5}$	$f_i u_i$
5	20	–4	–80
10	43	–3	–129
15	75	–2	–150
20	67	–1	–67
25 (A)	72	0	0
30	45	1	45
35	39	2	78
40	9	3	27
45	8	4	32
50	6	5	30
Sum	$N = \Sigma f_i = 384$		$\Sigma f_i u_i = -214$

Thus Arithmetic mean

$$(\bar{x}) = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times h = 25 + \left(\frac{-214}{384}\right) \times 5$$

$$= 25 - 2.786 = 22.214$$

Example *15:* **The measurements (in milli-metres) of the diameter of the heads of 107 screws are given below :**

Diameter in mm (x_i)	34	37	40	43	46
No. of screws (f_i)	17	19	23	21	27

Calculate mean diameter by step deviation method of the heads of the screws.

Solution :

Let us suppose assumed mean, $A = 40$

As $x_i - 40$, where $x_i = 34, 37, 40, 43, 46$, is divisible by 3, therefore $h = 3$

Calculation table is given as :

x_i	f_i	$u_i = \frac{x_i - 40}{3}$	$f_i u_i$
34	17	–2	–34
37	19	–1	–19
40(A)	23	0	0
43	21	1	21
46	27	2	54
	$\Sigma f_i = 107$		$\Sigma f_i u_i = 22$

We know, $\bar{x} = A + \left(\frac{\Sigma f_i u_i}{\Sigma f_i}\right) \times h$

$\Rightarrow \bar{x} = 40 + \frac{22 \times 3}{107} = 40 + 0.6 = 40.6\,mm$ nearly.

Mean of Continuous frequency distribution

In case of continuons frequency distribution with class intervals arithmetic mean may be computed by applying any of the methods discussed above

NOTE:

The values of $x_1, x_2, --, x_n$ are taken as the mid–points or class marks of the various classes. Let us understand the mean of Continuous frequency distribution by following illustrations.

(i) Direct Method

For the discrete frequency distribution

x_i	x_1	x_2	x_3	....	x_n
f_i	f_1	f_2	f_3	....	f_n

$$\text{Mean } (\bar{x}) = \frac{f_1x_1 + f_2x_2 + f_3x_3 + \ldots\ldots f_nx_n}{f_1 + f_2 + f_3 + \ldots\ldots + f_n}$$

Example *16:* **The following table shows the marks secured by 100 students in an examination.**

Marks	0 – 10	10 – 20	20 – 30	30 – 40	40 – 50
No. of students	15	20	35	20	10

Find mean marks obtained by a student.

Solution :

Marks	Mid value (x_i)	Frequency (f_i)	$f_i x_i$
0 – 10	5	15	75
10 – 20	15	20	300
20 – 30	25	35	675
30 – 40	35	20	700
40 – 50	45	10	450
Total		$\Sigma f_i = 100$	$\Sigma f_i x_i = 2400$

$$\text{Mean} = \bar{x} = \frac{\Sigma f_i x_i}{\Sigma f_i} = \frac{2400}{100} = 24$$

(ii) Assumed Mean Method

Mean $(\bar{x}) = A + \frac{\Sigma f_i d_i}{\Sigma f_i}$ where $d_i = x_i - A$, x_i = mid-point of i^{th} class interval

A = Assumed mean, which is generally the mid-value of middlemost class Interval.

NOTE:

If there are two middle class interval, then you can take mid-point of that middle class interval whose frequency is more as assumed mean.

Example 17: **Find the mean of the following frequency distribution by Assumed Mean Method.**

Class Interval	Frequency
0–4	6
4–8	3
8–12	6
12–16	16
16–20	3
20–24	14
24–28	10
28–32	8

Solution :

Class interval	Mid values (x_i)	Frequency (f_i)	Deviation (d_i) $= x_i - A$	$f_i d_i$
0–4	2	6	–12	–72
4–8	6	3	–8	–24
8–12	10	6	–4	–24
12–16	14 (A)	16	0	0
16–20	18	3	4	12
20–24	22	14	8	112
24–28	26	10	12	120
28–32	30	8	16	128
		$\Sigma f_i = 66$		$\Sigma f_i d_i = 252$

$$\text{Mean} = A + \frac{\Sigma f_i d_i}{\Sigma f_i} = 14 + \frac{252}{66} = 14 + 3.818 = \mathbf{17.818}$$

(iii) Step Deviation Method

$$\textbf{Mean } (\bar{x}) = A + \frac{\Sigma f_i u_i}{\Sigma f_i} \times h, \text{ where, } A = \text{Assumed mean},$$

$$u_i = \frac{x_i - A}{h}, \ h = \text{Class Size}$$

Example 18: **Calculate the mean for the following frequency distribution (By step deviation method).**

Class interval	0 – 80	80 – 160	160 – 240	240 – 320	320 – 400
frequency	22	35	44	25	24

Solution :

Class Interval	Mid - value (x_i)	f_i	$u_i = (x_i - A)/h$	$f_i u_i$
0 – 80	40	22	–2	–44
80 – 160	120	35	–1	–35
160 – 240	200 (A)	44	0	0
240 – 320	280	25	1	25
320 – 400	360	24	2	48
Total		$\Sigma f_i = 150$		$\Sigma f_i u_i = -6$

$$\text{Mean}(\bar{x}) = A + \frac{\Sigma f_i u_i}{\Sigma f_i} \times h = 200 + \left(\frac{-6}{150}\right) \times 80$$

$$= 200 - \frac{2 \times 8}{5} = 200 - \frac{16}{5} = \frac{1000 - 16}{5} = \frac{984}{5} = 196.8$$

NOTE:

(i) The mean obtained by all the there methods is the same.
(ii) The assumed mean method and step-deviation method are just simplified forms of the direct method.
(iii) If x_i and f_i are sufficiently small, then the direct method is an appropriate choice.
(iv) If x_i and f_i are numerically large numbers, then we can go for the assumed mean method or step-deviation method.
(v) If the class sizes are unequal, and x_i are large numberically, then we can go for the step-deviation method.

Geometric Mean (GM)

If $x_1, x_2, ..., x_n$ are n non-zero values of a variate X, then geometric mean is

$$GM = (x_1 . x_2 . ....x_n)^{1/n}$$

And if corresponding frequency of each variate is $f_1, f_2, ..., f_n$, then

$$GM = (x_1^{f1} . x_2^{f2} ... x_n^{fn})^{1/n}$$

When $N = f_1 + f_2 + ... + f_n$

Example 19: **The GM of the numbers 3, 3^2, 3^3,..., 3^n is**

(a) $3^{2/n}$ **(b)** $3^{(n-1)/2}$
(c) $3^{n/2}$ **(d)** $3^{(n+1)/2}$

Solution : **(d)**

$$GM = (3 \cdot 3^2 \cdot 3^3 \cdot ....3^n)^{1/n}$$

$$= 3^{(1+2+....+n)/n} = \left(3^{\frac{n(n+1)}{2}}\right)^{1/n} = 3^{\frac{n+1}{2}}$$

Harmonic Mean

If $x_1, x_2, , ... x_n$ are n non-zero values of a variate x, then harmonic mean is

$$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + ... + \frac{1}{x_n}}$$

And if corresponding frequencies of each variate is $f_1, f_2, ..., f_n$, then

$$HM = \frac{f_1 + f_2 + ... + f_n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + ... + \frac{f_n}{x_n}}$$

MEDIAN

The **median** is that value of the given number of observations, which divides it into exactly two parts.

Median is the middle most or the central value of the variate in a set of observations,when the observations are arranged either in ascending or in descending order of their magnitudes.

Methods to Find Median

When the data is arranged in ascending (or descending) order the median of ungrouped data is calculated as follows:

(i) When the number of observation (n) is odd, the median is the value of the $\frac{n+1}{2}$th observation.

(ii) When the number of observations (n) is even, the median is the mean of the $\frac{n}{2}$th and the $\frac{n}{2}+1$ th observations.

Median of the Discrete Frequency Distribution

Step-I : *Find the cumulative frequencies (c.f.)*

Step-II : *Find* $\frac{N}{2}$, *where* $N = \sum_{i=1}^{n} f_i$

Step-III : *See the cumulative frequency (c.f.) just greater than* $\frac{N}{2}$ *and determine the corresponding value of the variable.*

Step-IV : *The value obtained in step III is the median.*

Example 20: **Obtain the median for the following frequency distribution :**

x :	1	2	3	4	5	6	7	8	9
f :	8	10	11	16	20	25	15	9	6

Solution:

Calculation of Median

x	*f*	*cf*
1	8	8
2	10	18
3	11	29
4	16	45
5	20	65
6	25	90
7	15	105
8	9	114
9	6	120
	$N = 120$	

Here, $N = 120 \Rightarrow \frac{N}{2} = 60$

We find that the cumulative frequency just greater than $\frac{N}{2}$ is 65 and the value of x corresponding to 65 is 5. Therefore, Median = 5.

MEDIAN OF CONTINUOUS FREQUENCY DISTRIBUTION

Algorithm to Find the Median :

Step I : Make cumulative frequency table.

Step II : Choose the median class. Median class is the class whose cumulative frequency is greater than and nearest to $\frac{n}{2}$, where n is the sum of all frequencies.

Step III : Use this formula

$$\text{Median} = \ell + \left[\frac{\frac{n}{2} - c \cdot f}{f}\right] \times h$$

where, ℓ = lower limit of median class
n = sum of all frequencies (or sum of all observations)
cf = cumulative frequency of class preceeding the median class
f = frequency of the median class
h = class size

Example 21: The following table gives the distribution of the life time of 400 neon lamps.

Life time (in hours)	No. of Lamps
1500 – 2000	14
2000 – 2500	56
2500 – 3000	60
3000 – 3500	86
3500 – 4000	74
4000 – 4500	62
4500 – 5000	48

Solution :

Class - Intervals	Frequency	Cumulative frequency
1500 – 2000	14	14
2000 – 2500	56	70
2500 – 3000	60	130
3000 – 3500	86	216
3500 – 4000	74	290
4000 – 4500	62	352
4500 – 5000	48	400

Here n = 400 $\therefore \frac{n}{2} = \frac{400}{2} = 200$

$\therefore$ Median class is 3000 – 3500.

So, $\ell = 3000$, f = 86, h = 500, cf = 130

$$\therefore \text{Median} = \ell + \left[\frac{\frac{n}{2} - cf}{f}\right] \times h = 3000 + \left[\frac{\frac{400}{2} - 130}{86}\right] \times 500$$

$$= 3000 + \frac{70}{86} \times 500 = 3406.98$$

MODE

The **mode** is that value of the observation which occurs most frequently, i.e., an observation with the maximum frequency is called the mode.

NOTE:

It is not effected by presence of extremely large or small items.

Mode of Individual Observation

Mode of a frequency distribution is that value of the variable which has maximum frequency.

Example 22: Find the mode of the following data :
120, 110, 130, 110, 120, 140, 130, 120, 140, 120

Solution :

120 has the maximum frequency. Hence, the mode is 120.

Mode of a Discrete Series

Mode of a discrete series is the value of variable consisting highest frequency.

Example 23: Compute the modal value for the following frequency distribution

x	95	105	115	125	135	145	155	165	175
y	4	2	18	22	21	19	10	3	2

(a) 115 (b) 125
(c) 22 (d) None of these

Solution : **(b)**

From the given table, it is clear that 125 has the highest frequency 22. Hence, modal value of the given frequency distribution is 125.

MODE OF CONTINUOUS FREQUENCY DISTRIBUTION

$$\text{Mode} = \ell + \frac{f_1 - f_0}{2f_1 - f_0 - f_2} \times h, \text{ where}$$

ℓ = lower limit of the modal class*
h = size of the class interval
f_1 = frequency of the modal class
f_0 = frequency of the class preceeding the modal class
f_2 = frequency of the class succeeding the modal class.

NOTE : Modal class is the class having maximum frequency.

Example 24: The given distribution shows the number of runs scored by some top batsman of the world in one-day international cricket matches. Find the mode of the data.

Runs scored	No. of batsman
3000 – 4000	4
4000 – 5000	18
5000 – 6000	9
6000 – 7000	7
7000 – 8000	6
8000 – 9000	3
9000 – 10000	1
10000 – 11000	1

Solution :

Here, maximum frequency is 18. So, the modal class is 4000-5000.

So, $\ell = 4000$, $h = 1000$, $f_1 = 18$, $f_0 = 4$, $f_2 = 9$

$$\text{Mode} = \ell + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h = 4000 + \left(\frac{18-4}{2 \times 18 - 4 - 9}\right) \times 1000$$

$$= 4000 + \frac{14}{23} \times 1000 = 4000 + 608.7 = 4608.7$$

THE RELATIONSHIP BETWEEN MEAN, MEDIAN AND MODE

3 Median = Mode + 2 Mean

EXERCISE

1. The mean weight of all the students in a certain class is 60 kg. The mean weight of the boys from the class is 70 kg, while that of the girls is 55 kg. What is the ratio of number of boys to that of girls?
(a) 2 : 1 (b) 1 : 2
(c) 1:4 (d) 4:1
2. Students of two schools appeared for a common test carrying 100 marks. The arithmetic means of their marks for school I and II are 82 and 86 respectively. If the number of students of school II is 1.5 times the number of students of school I, what is the arithmetic mean of the marks of all the students of both the schools?
(a) 84.0
(b) 84.2
(c) 84.4
(d) This cannot be calculated with the given data
3. If AM of numbers $x_1, x_2 .... x_n$ is μ, then what is the AM of the numbers which are increased by 1, 2, 3, ...n respectively?
(a) $\mu+\left(\frac{n+1}{2}\right)$ (b) μ
(c) $\mu+\frac{n(n+1)}{2}$ (d) $\mu-\left(\frac{n+1)}{2}\right)$
4. If in a frequency distribution table with 12 classes, the width of each class is 2.5 and the lowest class boundary is 6.1, then what is the upper class boundary of the highest class?
(a) 30.1 (b) 27.6
(c) 30.6 (d) 36.1
5. Let $\bar{x}$ be the mean of n observations $x_1, x_2, ..., x_n$. If (a – b) is added to each observation, then what is the mean of new set of observations?
(a) 0 (b) $\bar{x}$
(c) $\bar{x}-(a-b)$ (d) $\bar{x}+(a-b)$
6.

X	1	2	3	4
Frequency	2	3	*f*	5

The frequency distribution of a discrete variable *X* with one missing frequency *f* is given above. If the arithmetic mean of *X* is $\frac{23}{8}$, what is the value of the missing frequency?
(a) 5 (b) 6
(c) 8 (d) 10
7. If the monthly expenditure pattern of a person who earns a monthly salary of ₹15000 is represented in a pie diagram, then the sector angle of an item on transport expenses measures 15°. What is his monthly expenditure on transport?
(a) ₹450 (b) ₹625
(c) ₹675
(d) Cannot be computed from the given data
8. If $\sum_{i=1}^{n}(x_i-2)=110, \sum_{i=1}^{n}(x_i-5)=20$, then what is the mean?
(a) 11/2 (b) 2/11
(c) 17/3 (d) 17/9
9. The distributions *X* and *Y* with total number of observations 36 and 64, and mean 4 and 3 respectively are combined. What is the mean of the resulting distribution *X* + *Y*?
(a) 3.26 (b) 3.32
(c) 3.36 (d) 3.42
10.

Class Interval	1-5	6-10	11-15	16-20
Frequency	3	7	6	5

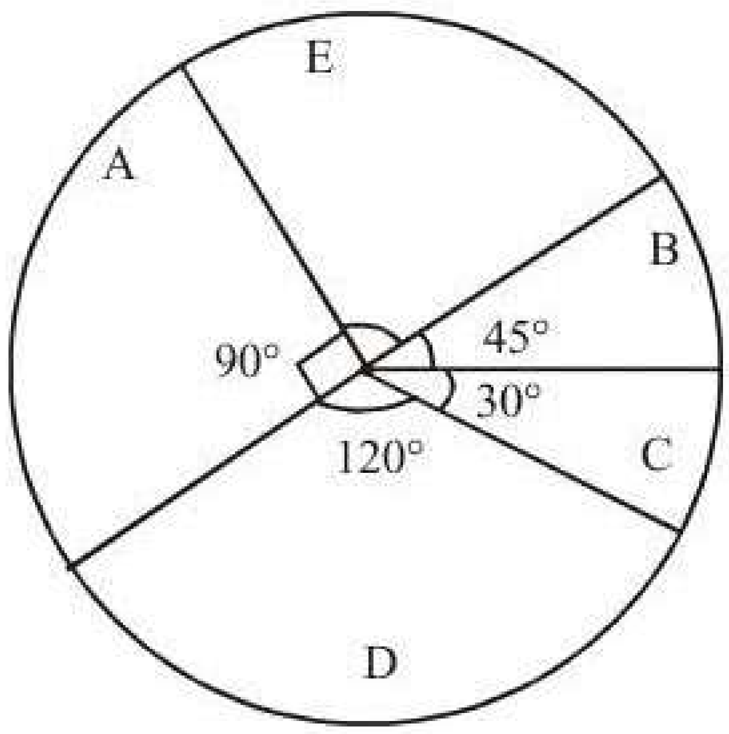
Consider the following statements in respect of the above frequency distribution.
I. The median is contained in the modal class.
II. The distribution is bell-shaped.
Which of the above statements is/are correct?
(a) Only I (b) Only II
(c) Both I and II (d) Neither I nor II

DIRECTIONS (Q. 11 & 12) : The following table gives the continuous frequency distribution of a continuous variable X

Class Interval	0-10	10-20	20-30	30-40	40-50
Frequency	5	10	20	5	10

11. What is the median of the above frequency distribution?
(a) 23 (b) 24
(c) 25 (d) 26
12. What is the mean of the above frequency distribution?
(a) 25 (b) 26
(c) 27 (d) 28
13. What is the cumulative frequency curve of statistical data commonly called?
(a) Cartogram (b) Histogram
(c) Ogive (d) Pictogram
14. The mean of 7 observations is 10 and that of 3 observations is 5. What is the mean of all the 10 observations?
(a) 15 (b) 10
(c) 8.5 (d) 7.5
15. Some measures of central tendency for *n* discrete observations are given below:
1. Arithmetic mean 2. Geometric mean
3. Harmonic mean 4. Median
A desirable property of a measure of central tendency is if every observation is multiplied by *c*, then the measure of central tendency is also multiplied by *c*, where *c* > 0. Which of the above measures satisfy the property?
(a) 1, 2 and 3 only (b) 1, 2 and 4 only
(c) 3 and 4 only (d) 1, 2, 3 and 4

DIRECTIONS (Qs. 16 to 18): Note : *Study the pie chart given below and answer the next 04 (four) questions that follow :*
The following pie chart gives the distribution of funds in a five year plan under the major heads of development expenditures: Agriculture (A), Industry (B), Education (C), Employment (D) and Miscellaneous (E)
The total allocation is 36,000 (in crores of rupees).



16. Which head is allocated maximum funds?
   (a) Agriculture (b) Industry
   (c) Employment (d) Miscellaneous
17. How much money (in crores) is allocated to Education?
   (a) 3000 (b) 6000
   (c) 9000 (d) 10800
18. The arithmetic mean of numbers a, b, c, d, e is M. What is the value of (a – M) + (b – M) + (c – M) + (d – M) + (e – M) ?
   (a) M (b) a + b + c + d + e
   (c) 0 (d) 5 M
19. The median of 27 observations of a variable is 18. Three more observations are made and the values of these observations are 16, 18 and 50. What is the median of these 30 observations ?
   (a) 18
   (b) 19
   (c) 25.5
   (d) Can not be determined due to insufficient data
20. Marks obtained by 7 students in a subject are 30, 55, 75, 90, 50, 60, 39. The number of students securing marks less than the mean marks is
   (a) 7 (b) 6
   (c) 5 (d) 4
21. What type of classification is needed to enumerate the female population of India?
   (a) Geographical (b) Chronological
   (c) Qualitative (d) Quantitative
22. If mean of y and $\frac{1}{y}$ is M, then what is the mean of y^3 and $\frac{1}{y^3}$?
   (a) $\frac{M(M^2-3)}{3}$ (b) M^3
   (c) $M^3 - 3$ (d) $M(4M^2 - 3)$
23. For less than ogive, the cumulative frequencies are plotted against which of the following?
   (a) Upper limits of class intervals
   (b) Lower limits of the class intervals
   (c) Mid-points of the class intervals
   (d) Either (b) or (c)
24. For the following frequency distribution:

Class interval	0-5	5-10	10-15	15-20	20-25	25-30
Frequency	10	15	30	80	40	20

If m is the value of mode, then which one of the following is correct?
   (a) $5 < m < 10$ (b) $10 < m < 15$
   (c) $15 < m < 20$ (d) $20 < m < 25$
25. An average Indian family allocates its monthly income under different heads as follows:

Items	Percentage Share
Food	40
House Rent	15
Saving	x
Transport	12
Miscellaneous	23

A pie diagram of this data is to be drawn. What is the value of x, if the angle which the sector representing saving makes at the centre is 36°?
   (a) 13 (b) 11
   (c) 10 (d) 8
26. Data on percentage distribution of area of land in acres owned by households in two districts of a particular state are as follows:

Land Holding	District A	District B
0.01 – 0.99	5.62	13.53
1.0 – 2.49	18.35	21.84
2.5 – 7.49	47.12	39.32
7.5 – 12.49	19.34	12.15
12.5 – 19.99	7.21	7.43
20.0 – 29.99	2.36	5.73

What is the appropriate diagram to represent the above data?
   (a) Pie diagram (b) Histogram
   (c) Bar chart (d) None of the above

27. The following table shows the percentage of male and female coffee drinkers and non-coffee drinkers in two towns A and B.

Attributes	Town A		Town B	
	Male	Female	Male	Female
Coffee drinkers	40%	5%	25%	15%
Non-coffee drinkers	20%	35%	30%	30%

If the total population of the towns A and B are 10000 and 20000 respectively, then what is the total number of female coffee drinkers in both towns?
(a) 8000 (b) 6000
(c) 3500 (d) 2500

DIRECTIONS (Qs. 28-30): *Read the following information carefully to answer the questions that follow.*
The arithmetic mean, geometric mean and median of 6 positive numbers a, a, b, b, c, c, where a < b < c are $\frac{7}{3}$, 2, 2, respectively.

28. What is the sum of the squares of all the six numbers?
(a) 40 (b) 42
(c) 45 (d) 48

29. What is the value of c?
(a) 1 (b) 2
(c) 3 (d) 4

30. What is the mode?
(a) 1 (b) 2
(c) 1, 2 and 4 (d) No mode

31. Consider the following data:

x	1	2	3	4	5
f	3	5	9	–	2

If the arithmetic mean of the above distribution is 2.96, then what is the missing frequency?
(a) 4 (b) 6
(c) 7 (d) 8

32. Consider the following statements in respect of a histogram:
I. The histogram consists of vertical rectangular bars with a common base such that there is no gap between consecutive bars.
II. The height of the rectangle is determined by the frequency of the class it represents.
Which of the statements given above is/are correct?
(a) Only I (b) Only II
(c) Both I and II (d) Neither I nor II

33. Which one of the following relations for the numbers 10, 7, 8, 5, 6, 8, 5, 8 and 6 is correct?
(a) Mean = Median (b) Mean = Mode
(c) Mean > Median (d) Mean > Mode

34. If m is the mean of p, q, r, s, t, u and v, then what is (p – m) + (q – m) + (r – m) + (s – m) + (t – m) + (u – m) + (v – m) equal to?
(a) 0 (b) s
(c) $\frac{(p+v)}{2}$ (d) None of these

DIRECTIONS (Qs. 35-36): *Read the following information carefully and answer the questions given below.*
The median of the following distribution is 14.4 and the total frequency is 20.

Class Interval	0 – 6	6 – 12	12 – 18	18 – 24	24 – 30
Frequency	4	x	5	y	1

35. What is x equal to?
(a) 4 (b) 5
(c) 6 (d) 7

36. What is the relation between x and y?
(a) 2x = 3y (b) 3x = 2y
(c) x = y (d) 2x = y

37. There are 45 male and 15 female employees in an office. If the mean salary of the 60 employees is ₹ 4800 and the mean salary of the male employees is ₹ 5000, then the mean salary of the female employees is
(a) ₹ 4200 (b) ₹ 4500
(c) ₹ 5600 (d) ₹ 6000

38. The mean of 7 observations is 7. If each observation is increased by 2, then the new mean is:
(a) 12 (b) 10
(c) 9 (d) 8

39. Which of the following are the examples of discrete variables?
I. Number of errors per page in a book.
II. Height of individuals measured in centimetre.
III. Waiting time to failure of electric bulbs.
IV. Number of leaves on branches of a tree.
Select the correct answer using the codes given below.
(a) Only I (b) I and IV
(c) III and IV (d) II and IV

40. Consider the following statements:
I. A frequency distribution condenses the data and reveals its important features.
II. A frequency distribution is an equivalent representation of original data.
Which of the above statements is/are correct?
(a) Only I (b) Only II
(c) Both I and II (d) Neither I nor II

41. Consider the following:
I. The arithmetic mean of two unequal positive numbers is always greater than their geometric mean.
II. The geometric mean of two unequal positive numbers is always greater than their harmonic mean.
Which of the above statements is/are correct?
(a) Only I (b) Only II
(c) Both I and II (d) Neither I nor II

42. Consider the following statements in respect of a discrete set of numbers.
    I. The arithmetic mean uses all the data is always uniquely defined.
    II. The median uses only one or two numbers from the data and may not be unique.
    Which of the above statements is/are correct?
    (a) Only I (b) Only II
    (c) Both I and II (d) Neither I nor II
43. The geometric mean of $(x_1, x_2, x_3, ..., x_n)$ is x and the geometric mean of $(y_1, y_2, y_3, ..., y_n)$ is y. Which of the following is/are correct?
    I. The geometric mean of $(x_1 y_1, x_2 y_2, x_3 y_3, ..., x_n y_n)$ is XY.
    II. The geometric mean of $\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \frac{x_3}{y_3}, ..., \frac{x_n}{y_n}\right)$ is $\frac{X}{Y}$.
    Select the correct answer using the code given below.
    (a) Only I (b) Only II
    (c) Both I and II (d) Neither I nor II
44. The following table gives 'less than' type frequency distribution of income per day.

Income (in ₹) less than	Number of persons
1500	100
1250	80
1000	70
750	55
500	32
250	12

What is the modal class?
(a) 250-500 (b) 500-750
(c) 750-1000 (d) None of these
45. Which of the following items of information is a good example of statistical data?
    (a) A table of logarithms of numbers
    (b) A list of names of 120 students of a class
    (c) A list of annual incomes of the members of a club
    (d) Holiday list of the offices of Government of India in the year 2013
46. Consider the following in respect of variate which takes values 2, 2, 2, 2, 7, 7, 7 and 7.
    I. The median is equal to mean.
    II. The mode is both 2 and 7.
    Which of the above statements is/are correct?
    (a) Only I
    (b) Only II
    (c) Both I and II
    (d) Neither I nor II
47. Consider the following statements pertaining to a frequency polygon of a frequency distribution of a continuous variable having seven class intervals of equal width.
    I. The original frequency distribution can be reconstructed from the frequency polygon.
    II. The frequency polygon touches the X-axis in its extreme right and extreme left.
    Which of the above statements is/are correct?
    (a) Only I (b) Only II
    (c) Both I and II (d) Neither I nor II
48. The mean of the following distribution is 18.

Class interval	Frequency
11 – 13	3
13 – 15	6
15 – 17	9
17 – 19	13
19 – 21	f
21 – 23	5
23 – 25	4

What is the value of f?
(a) 8 (b) 9
(c) 10 (d) 11
49. The class which has maximum frequency is known as
    (a) median class (b) mean class
    (c) modal class (d) None of these
50. Consider the following statements related to cumulative frequency polygon of a frequency distribution, the frequencies being cumulated from the lower end of the range :
    1. The cumulative frequency polygon gives an equivalent representation of frequency distribution table.
    2. The cumulative frequency polygon is a closed polygon with one horizontal and one vertical side. The other sides have non–negative slope.
    Which of the above statements is / are correct ?
    (a) Only 1 (b) Only 2
    (c) Both 1 and 2 (d) Neither 1 nor 2
51. Consider the following data :
    1. Number of complaints lodged due to road accidents in a state within a year for 5 consecutive years.
    2. Budgetary allocation of the total available funds to the various items of expenditure.
    Which of the above data is / are suitable for repsentation of a pie diagram ?
    (a) Only 1 (b) Only 2
    (c) Both 1 and 2 (d) Neither 1 nor 2
52. When we take class intervals on the X–axis and corresponding frequencies on the Y-axis and draw rectangles with the areas proprotional to the frequencies of the respective class intervals, the graph so obtained is called
    (a) bar diagram (b) frequency curve
    (c) ogive (d) None of the above

53. If x_i 's are the mid–points of the class intervals of grouped data, f_i's are the corresponding frequencies and $\overline{x}$ is the mean, then what is $\sum f_i(x_i - \overline{x})$ equal to?
(a) 0 (b) –1
(c) 1 (d) 2

54. Ten observations 6, 14, 15, 17, $x+1$, $2x-13$, 30, 32, 34 and 43 are written in ascending order. The median of the data is 24. What is the value of x?
(a) 15 (b) 18
(c) 20 (d) 24

55. If A, G and H are the arithmetic, geometric and harmonic means between a and b respectively, then which one of the following relations is correct? **(CDS)**
(a) G is the geometric mean between A and H
(b) A is the arithmetic mean between G and H
(c) H is the harmonic mean between A and G
(d) None of the above

56. The geometric mean of three positive numbers a, b, c is 3 and the geometric mean of another three positive numbers d, e, f, is 4. Also, at least three elelments in the set $\{a, b, c, d, e, f\}$ are distinct. Which one of the following inequalities gives the best information about M, the arithmetic mean of the six numbers ? **(CDS)**
(a) $M > 2\sqrt{3}$
(b) $M > 3.5$
(c) $M \geq 3.5$
(d) It is not possible to set any precise lower limit for M

57. There are five parties A, B, C, D and E in an election. Out of total 100000 votes cast, 36000 were cast to party A, 24000 to party B, 18000 to party C, 7000 to party D and rest to party E. What angle will be allocated for party E in the pie chart? **(CDS)**
(a) 15° (b) 54°
(c) 60° (d) 72°

DIRECTIONS (Qs. 58-61): *For the next four (4) items that follow: Consider the following frequency distribution :*

Class	Frequency
0-10	4
10-20	5
20-30	7
30-40	10
40-50	12
50-60	8
60-70	4

58. What is the mean of the distribution? **(CDS)**
(a) 37.2 (b) 38.1
(c) 39.2 (d) 40.1

59. What is the median class? **(CDS)**
(a) 20–30 (b) 30–40
(c) 40–50 (d) 50–60

60. What is the median of the distribution? **(CDS)**
(a) 37 (b) 38
(c) 39 (d) 40

61. What is the mode of the distribution? **(CDS)**
(a) 38.33 (b) 40.66
(c) 42.66 (d) 43.33

62. The mean and median of 5 observations are 9 and 8 respectively. If 1 is substracted from each observation, then the new mean and the new median will respectively be **(CDS)**
(a) 8 and 7
(b) 9 and 7
(c) 8 and 9
(d) Cannot be determined due to insufficient data

63. The age distribution of 40 children is as follows:

Age (in years)	5–6	6–7	7–8	8–9	9–10	10–11
No. of children	4	7	9	12	6	2

Consider the following statements in respect of the above frequency distribution:
1. The median of the age distribution is 7 years.
2. 70% of the children are in the age group 6–9 years.
3. The modal age of the children is 8 years.

Which of the above statements are correct?
(a) 1 and 2 only (b) 2 and 3 only
(c) 1 and 3 only (d) 1, 2 and 3

64. Suppose $x_1 = \lambda^8$ for $0 < 10$, where $\lambda > 1$. **(CDS)**
Which one of the following is correct?
(a) AM – Median (b) GM – Median
(c) GM – Median (d) AM – Median

65. If a variable takes discrete values a + 4, a – 3.5, a – 2.5, a – 3, a–2, a + 0.5, a + 5 and a – 0.5 where a > 0, then the median of the data set is **(CDS)**
(a) a – 2.5 (b) a – 1.25
(c) a – 1.5 (d) a – 0.75

66. If each of n numbers $x_i = i$ (i = 1, 2, 3,..... n) is replaced by $(i+1)x_i$, then the new mean is **(CDS)**
(a) $\frac{n+3}{2}$ (b) $\frac{n(n+1)}{2}$
(c) $\frac{(n+1)(n+2)}{3n}$ (d) $\frac{(n+1)(n+2)}{3}$

HINTS & SOLUTIONS

1. (b) Let there be x number of boys and y number of girls.
Total students = x + y
Total weight of the students = (x + y)60
Total weight for boys = x × 70
Total weight for girls = y × 55
Hence, (x + y)60 = 70x + 55y
60x + 60y = 70x + 55y
5y = 10x ⇒ y = 2x
$\frac{x}{y}=\frac{1}{2} \Rightarrow x:y=1:2$

2. (c) Let the number of students of school I = x
∴ Number of students of School II = 1.5 x
As given :
Mean of marks for school I = 82
and mean of marks for school II = 86
∴ Combined mean $= \frac{x \times 82 + 1.5x \times 86}{x+1.5x}$
$= \frac{x(82+129)}{2.5x} = \frac{211}{2.5} = 84.4$

3. (a) Since, AM of number $x_1, x_2, x_3, ...x_n$ is μ
∴ $n\mu = x_1 + x_2 + ... x_n$
Sum of new numbers
$= (x_1+1)+(x_2+2)+(x_3+3)+...+(x_n+n)$
$= (x_1+x_2+...+x_n)+(1+2+3+...+n)$
$= n\mu + \frac{n(n+1)}{2}$
∴ $AM = \mu + \frac{(n+1)}{2}$

4. (d) Given : lowest class boundary = 6.1
Class width = 2.5, Number of classes = 12.2
⇒ Upper class boundary of the highest class
= 6.1 + (2.5× 12) = 6.1 +30 = 36. 1

5. (d) Let $\bar{x}$ is the mean of n observation $x_1, x_2, ..., x_n$.
$\Rightarrow \bar{x} = \frac{x_1+x_2+x_3+....+x_n}{n}$
Now, (a – b) is added to each term.
∴ New mean
$= \frac{x_1+(a-b)+x_2+(a-b)+.....+x_n+(a-b)}{n}$
$= \frac{x_1+x_2+....+x_n}{n} + \frac{n(a-b)}{n}$
$= \bar{x} + (a-b)$

6. (b) Arithmetic mean $= \frac{2\times1+3\times2+3f+4\times5}{2+3+f+5}$
$\Rightarrow \frac{23}{8} = \frac{28+3f}{10+f}$
$\Rightarrow 230+23f = 224+24f$
$\Rightarrow f = 6$

7. (b) Since, monthly salary = ₹15000
and sector angle of expenses = 15°
∴ Amount $= \frac{15°}{360°} \times 15000 =$ ₹ 625

8. (c) $\sum_{i=1}^{n}(x_i-2)=110$
∴ $x_1+x_2+...+x_n-2n=110$
$\Rightarrow x_1+x_2+...+x_n=2n+110$ ...(i)
and $\sum_{i=1}^{n}(x_i-5)=20$
$\Rightarrow x_1+x_2+...+x_n-5n=20$
$\Rightarrow x_1+x_2+...+x_n=5n+20$ ...(ii)
From equations (i) and (ii), we get
$5n+20=2n+110$
$\Rightarrow 3n=90 \Rightarrow n=30$
Now, mean $= \frac{x_1+x_2+...x_n}{n} = \frac{5\times30+20}{30} = \frac{170}{30} = \frac{17}{3}$

9. (c) Required mean $= \frac{36\times4+64\times3}{36+64} = \frac{144+192}{100} = \frac{336}{100}$
$= 3.36$

10. (d)

Class Interval	f	cf
0.5-5.5	3	3
5.5-10.5	7	10
10.5-15.5	6	16
15.5-20.5	5	21
Total	21	50

$N = 21$
∴ $\frac{N}{2} = \frac{21}{2} = 10.5$
∴ Median class is 10.5-15.5
Hence, Median $= 10.5 + \frac{10.5-10}{6} \times 5$
$= 10.5 + 0.417 = 10.917$
Thus, median is not contained in the modal class and the distribution is not bell-shaped.

Solution (Qs. 11-12) :

Class Interval	f	cf	x	fx
0-10	5	5	5	25
10-20	10	15	15	150
20-30	20	35	25	500
30-40	5	40	35	175
40-50	10	50	45	450
Total	50	145	125	1300

$\therefore \quad \frac{N}{2} = \frac{50}{2} = 25$

11. (c) Median group is 20-30.

$\Rightarrow \text{Median} = 20 + \frac{25-15}{20} \times 10 = 20 + 5 = 25$

12. (b) $\text{Mean} = \frac{\sum fx}{\sum f} = \frac{1300}{50} = 26$

13. (c) The cumulative frequency curve of statistical data is called Ogive.

14. (c) Given mean of 7 observations is 10.

$\therefore \frac{x_1 + x_2 + ... + x_7}{7} = 10$

$\Rightarrow x_1 + x_2 + ... + x_7 = 70$ ...(1)

Also, mean of 3 observations is 5.

$\therefore \frac{x_8 + x_9 + x_{10}}{3} = 5$

$\Rightarrow x_8 + x_9 + x_{10} = 15$ ...(2)

So, from (1) and (2)

Required mean

$= \frac{x_1 + x_2 + ... + x_7 + x_8 + x_9 + x_{10}}{10}$

$= \frac{70+15}{10} = \frac{85}{10} = 8.5$

15. (b) "If every observation is multiplied by c, then the measure of central tendency is also multiplied by c, where $c > 0$. Arithmetic mean, Geometric mean and median satisfies above property.

16. (c) Agriculture : $\frac{90}{360} \times 36000 = 9000$

Miscellaneous : $\frac{75}{360} \times 36000 = 7500$

Industry : $\frac{45}{360} \times 36000 = 4500$

Education : 3000 Employment : 12000

Hence, Employment is allocated maximum funds.

17. (a) Education : 3000

18. (c) Given $M = \frac{a+b+c+d+e}{5}$

$\Rightarrow a+b+c+d+e = 5M$

$\Rightarrow a+b+c+d+e-5M = 0$

$\Rightarrow (a-M)+(b-M)+(c-M)+(d-M)+(e-M) = 0$

Hence, Required value = 0

19. (b) Median is middle of data. Observations are 27 and median is 18. So, sum of all the observation are $18 \times 27 = 486$.

Now, 16, 18 and 50 are additional three observations.

So, Total $= 486 + 16 + 18 + 50 = 570$.

and number of obs. are 30.

$\therefore \quad \text{Median} = \frac{570}{30} = 19$

20. (d) Given marks are 30, 55, 75, 90, 50, 60, 39.

Mean marks $= \bar{x}$

$\bar{x} = \frac{30+55+75+90+50+60+39}{7} = \frac{399}{7} = 57$

Hence, 4 students secured marks less than the mean marks.

21. (d) Quantitative is needed to enumerate the female population of India.

22. (d) Mean of y and $\frac{1}{y}$ = M

$\Rightarrow \frac{y + \frac{1}{y}}{2} = M \Rightarrow y + \frac{1}{y} = 2M$ ... (i)

Now, mean of y^3 and $\frac{1}{y^3}$ is

$\frac{y^3 + \frac{1}{y^3}}{2} = \frac{\left(y + \frac{1}{y}\right)^3 - 3\left(y + \frac{1}{y}\right)}{2}$

$\Rightarrow \frac{y^3 + \frac{1}{y^3}}{2} = \frac{(2M)^3 - 6M}{2}$

$= \frac{(2M)\left[(2M)^2 - 3\right]}{2} = M(4M^2 - 3)$

23. (a) For an ogive, the cumulative frequencies are plotted as a upper limit of class intervals.

24. (c) Here, maximum frequency is 80, hence mode will be between 15-20.

25. (c) We know that:

$\text{Central angle} = \frac{\frac{\text{Value of item}}{\text{Sum of values of items}}}{x} \times 360°$

$\therefore \quad 36° = \dfrac{x}{40+15+x+12+3} \times 360°$

$\Rightarrow \quad \dfrac{36°}{360°} = \dfrac{x}{90+x}$

$\Rightarrow \quad 90 + x = 10x$

$\Rightarrow \quad 9x = 90$

$\therefore \quad x = 10$

26. (c) Because there is a gap between two adjacent bars, so both the districts can be represented by bar chart.

27. (c) Total number of female coffee drinkers
$= 5\%$ of $10000 + 15\%$ of $20000 = 500 + 3000 = 3500$

Solutions (Q. Nos. 28-30):

$a < b < c$

Total numbers = 6

Increasing order a, a, b, b, c, c

$\therefore \quad \text{Median} = \dfrac{\left(\dfrac{6}{2}\right)\text{th term} + \left(\dfrac{6}{2}+1\right)\text{th term}}{2}$

$= \dfrac{\text{3rd term} + \text{4th term}}{2}$

$2 = \dfrac{b+b}{2} = b$

$\text{Arithmetic mean} = \dfrac{a+a+b+b+c+c}{6}$

$\Rightarrow \quad \dfrac{7}{3} = \dfrac{a+b+c}{3}$

$\Rightarrow \quad a + b + c = 7$

$\Rightarrow \quad a + c = 7 - 2 = 5 \quad \text{... (i)}$

$\text{Geometric mean} = \left(a^2 \times b^2 \times c^2\right)^{\frac{1}{6}}$

$\Rightarrow \quad 2 = (abc)^{\frac{1}{3}}$

$\Rightarrow \quad abc = 8$

$\Rightarrow \quad ac = \dfrac{8}{2} = 4 \quad \text{... (ii)}$

$\Rightarrow \quad c = \dfrac{4}{a}$

From equation (i),

$a + \dfrac{4}{a} = 5$

$\Rightarrow \quad \dfrac{a^2+4}{a} = 5$

$\Rightarrow \quad a^2 - 5a + 4 = 0$

$\Rightarrow \quad a^2 - 4a - a + 4 = 0$

$\Rightarrow \quad a(a-4) - 1(a-4) = 0$

$\Rightarrow \quad (a-4)(a-1) = 0$

if a = 1 then c = 4

a = 4 then c = 1

a = 1, c = 4 and b = 2

28. (b) Required sum $= 2(a)^2 + 2(b)^2 + 2(c)^2$
$= 2(1)^2 + 2(2)^2 + 2(4)^2$
$= 2 + 8 + 32 = 42$

29. (d) The value of c is 4.

30. (d) Mode = 3 (Median) – 2 (Mean)
$= 3(2) - 2\left(\dfrac{7}{3}\right) = \dfrac{18-14}{3} = \dfrac{4}{3}$

31. (b)

x	f	xf
1	3	3
2	5	10
3	9	27
4	f_1	$4f_1$
5	2	10
Total	$19 + f_1$	$50 + 4f_1$

$\therefore \quad \text{Mean} = \dfrac{\Sigma x_i f_i}{\Sigma f_i}$

$\Rightarrow \quad 2.96 = \dfrac{50+4f_1}{19+f_1}$ [given]

$\Rightarrow \quad 56.24 + 2.96 f_1 = 50 + 4f_1$

$\Rightarrow \quad 6.24 = 1.04 f_1$

$\Rightarrow \quad f_1 = 6$

32. (c) **Statement I :**
A graph which displays the data by using vertical bars of various heights in rectangular shapes to represent frequencies. Such that there is no gap between consecutive bars and also the height of the rectangle.
Statement II :
The height of the rectangle is determined by the frequency of the class it represents.
So, both the statements are correct.

33. (a) Given numbers are 10, 7, 8, 5, 6, 8, 5, 8 and 6
Arrange in ascending order
5, 5, 6, 6, 7, 8, 8, 8, 10
Total term, n = 9 (odd)
Now,

(i) $\text{Mean} = \dfrac{5+5+6+6+7+8+8+8+10}{9}$
$= \dfrac{63}{9} = 7$

(ii) $\text{Median} = \left(\dfrac{n+1}{2}\right)\text{th term} = \left(\dfrac{9+1}{2}\right)\text{th term}$
= 5th term = 7

(iii) Mode = 8 because of higher frequency term

$\therefore$ Mean = Median

34. (a) $\frac{p+q+r+s+t+u+v}{6} = m$

$\Rightarrow$ $p+q+r+s+t+u+v = 6m$

$\therefore$ $(p-m)+(q-m)+(r-m)+(s-m)+(t-m)+(u-m)+(v-m)$

$= (p+q+r+s+t+u+v) - 6m$

$= 6m - 6m = 0$

Solutions (Q. Nos. 35-36):

Class interval	Frequency	Cumulative frequency
0 – 6	4	4
6 – 12	x	4 + x
12 – 18	5	9 + x
18 – 24	y	9 + x + y
24 – 30	1	10 + x + y
	20	

According to question,

Here, $10 + x + y = 20$

$\Rightarrow$ $x + y = 20 - 10$

$\Rightarrow$ $x + y = 10$ ... (i)

35. (a) Median $= l + \left(\frac{\frac{n}{2} - cf}{f}\right) \times h$

$\Rightarrow$ $14.4 = 12 + \left[\frac{\frac{20}{2} - (4+x)}{5}\right] \times 6$

$\Rightarrow$ $14.4 = 12 + \frac{10 - 4 - x}{5} \times 6$

$\Rightarrow$ $14.4 - 12 = \frac{6 - x}{5} \times 6$

$\Rightarrow$ $2.4 = \frac{36 - 6x}{5}$

$\Rightarrow$ $12 = 36 - 6x$

$\Rightarrow$ $6x = 24$

$\therefore$ $x = 4$

36. (b) Now, putting the value of x in equation (i), then,

$4 + y = 10$

$\Rightarrow$ $y = 6$

$\frac{x}{y} = \frac{4}{6} = \frac{2}{3}$

$\Rightarrow$ $3x = 2y$

37. (a) Given that,

Number of male employees (M) = 45

Number of female employees (F) = 15

Mean salary of male employee $(\bar{x}_M)$ = ₹ 5000

Total number of employees = (M + F) = 45 + 15 = 60

Mean salary of employees $(\bar{x}_{MF})$ = ₹ 4800

Let mean salary of female employee is $\bar{x}_F$

By formula,

$$\bar{x}_{MF} = \frac{M\bar{x}_M + F\bar{x}_F}{(M+F)}$$

$\Rightarrow$ $4800 = \frac{45 \times 5000 + 15 \times \bar{x}_F}{60}$

$\Rightarrow$ $4800 \times 60 - 45 \times 5000 = 15 \times \bar{x}_F$

$\therefore$ $\bar{x}_F = 4800 \times 4 - 3 \times 5000$

$= 300(16 \times 4 - 50) = 300 \times 14 = 4200$

38. (c) Given that, mean of 7 observations = 7

$\Rightarrow$ $\frac{1}{7}\sum_{i=1}^{7} x_i = 7 \Rightarrow \sum_{i=1}^{7} x_i = 49$ ... (i)

According to question,

Each observation is increased by 2. Then the new mean,

$= \frac{1}{7}\sum_{i=1}^{7}(x_i + 2) = \frac{1}{7}\left(\sum_{i=1}^{7} x_i + 2 \times 7\right)$

$= \frac{1}{7}(49 + 14) = \frac{1}{7} \times 63 = 9$

39. (b) **Discrete Variable:**

It is a variable whose value is obtained by counting.

Examples:

(i) Number of students present.

(ii) Number of red marbles in a jar.

(iii) Students' grade level.

Continuous Variable:

It is a variable whose value is obtained by measuring.

Examples:

(i) Height of students in class.

(ii) Weight of students in class.

(iii) Time it takes to get to school.

(iv) Distance travelled between class.

So, statement I and IV are examples of discrete variables.

40. (c) A frequency distribution is a summary of the data set in which the interval of possible values is divided into sub-intervals known as class.

41. (c) The decreasing order of mean are:

Arithmetic mean > Geometric mean > Harmonic mean

42. (a) **Statement I:**

The Arithmetic Mean is obtained by sum of all the elements of the data set and dividing by the number of elements and it is always uniquely defined.

Statement II:

The median is the middle element when the data set is arranged in order of magnitude.

Mean, Median and mid-range always exist and are unique.

43. (c) Geometric mean of $(x_1, x_2, x_3, ..., x_n)$

$$= (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}} = X$$

Geometric mean of $(y_1, y_2, y_3, ..., y_n)$

$$= (y_1 \cdot y_2 \cdots y_n)^{\frac{1}{n}} = Y$$

$\therefore$ Geometric mean of $(x_1 y_1, x_2 y_2, ..., x_n y_n)$

$$= (x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}} \times (y_1 \cdot y_2 \cdots y_n)^{\frac{1}{n}}$$

$$= (x_1 y_1 \cdot x_2 y_2 \cdots x_n y_n)^{\frac{1}{n}} = XY$$

Geometric mean of $\left(\frac{x_1}{y_1}, \frac{x_2}{y_2}, \cdots \frac{x_n}{y_n}\right)$

$$= \frac{(x_1 \cdot x_2 \cdots x_n)^{\frac{1}{n}}}{(y_1 \cdot y_2 \cdots y_n)^{\frac{1}{n}}} = \left(\frac{x_1}{y_1} \cdot \frac{x_2}{y_2} \cdots \frac{x_n}{y_n}\right)^{\frac{1}{n}} = \frac{X}{Y}$$

44. (b)

Income less than	Class interval	Number of persons	Frequency
1500	1250 – 1500	100	20
1250	1000 – 1250	80	10
1000	750 – 1000	70	15
750	500 – 750	55	23
500	250 – 500	32	20
250	0 – 250	12	12

$$\text{Mode} = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

$$= 250 + \left(\frac{20 - 12}{40 - 12 - 23}\right) \times 250 = 250 + \frac{8}{5} \times 250$$

$$= 250 + 400 = 650$$

So, the modal class is 500-750.

45. (c) **Statistical data:**

In statistics and quantitative research methodology, a data sample is a set of data collected and or selected from a different sources and good example of statistical data – A list of annual incomes of the members of a club.

46. (c) I. Mean of all observations $= \frac{2 \times 4 + 7 \times 4}{8} = 4.5$

For median, first we arrange in ascending order = 2, 2, 2, 2, 7, 7, 7, 7

$$\therefore \text{ Median} = \frac{4\text{th} + 5\text{th}}{2} = \frac{2 + 7}{2} = 4.5$$

II. Mode is both 2 and 7, since frequency of occurrence is same, i.e. maximum frequency.

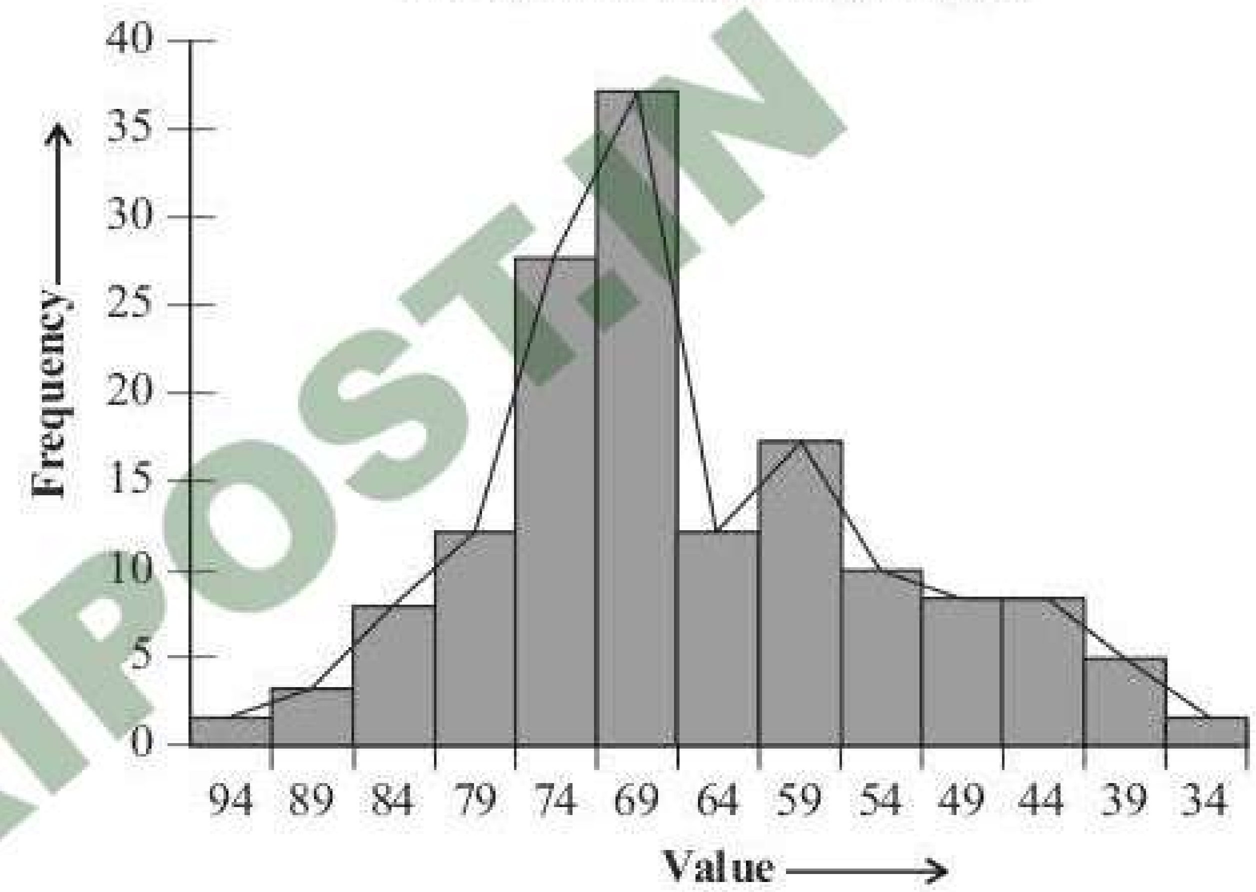
47. (c) **Statement I :**

Frequency polygons are a graphical device for understanding the shapes of distribution. They serve the same purpose as histograms. It is formed by joining the mid-points of histogram.

Statement II :

Frequency polygon touch the x-axis in its extreme left and extreme right of graph. See graph below:

Histogram / Frequency Polygon



48. (a)

Classes	Mid-values (x_i)	Frequency (f_i)	$d = x_i - 18$	$u_i = \frac{x_1 - 18}{2}$	$f_i u_i$
11 – 13	12	3	–6	–3	–9
13 – 15	14	6	–4	–2	–12
15 – 17	16	9	–2	–1	–9
17 – 19	18	13	0	0	0
19 – 21	20	f	2	1	f
21 – 23	22	5	4	2	10
23 – 25	24	4	6	3	12
		40 + f			f – 8

$$\text{Mean } (x) = A + \frac{\Sigma f_i u_i}{\Sigma f} \times h = 18 + \frac{f - 8}{40 + f} \times 2$$

Given mean = 18

$$18 = 18 + \frac{f - 8}{40 + f} \times 2$$

$$\Rightarrow 2f - 16 = 0 \Rightarrow 2f = 16$$

$$f = \frac{16}{2} = 8$$

$$f = 8$$

49. (c) The modal class means that the class which has maximum frequency.

50. (a) Here, Statement 1 is correct but Statement 2 is not correct.

51. (c) Both Statements 1 and 2 are suitable for represention of a pie diagram.

52. (d) In bar diagram, the frequency is shown by the height of the bar whereas in histogram the frequency is shown by the area of the bar. So obtained graph is histogram.

53. (a) $\sum f_i(x_i - \bar{x}) = 0$ because sum of product of deviations and frequencies from mean value will be 0.

54. (c) Observations can by arranged in ascending order. 6, 14, 15, 17, $x + 1$, $2x - 13$, 30, 32, 34 and 43.
Here, $n = 10$ [even]
∴ Median

$$= \frac{\text{Value of}\left(\frac{n}{2}\right)\text{th term} + \text{Value of}\left(\frac{n}{2}+1\right)\text{th term}}{2}$$

$$= \frac{\text{Value of}\left(\frac{10}{2}\right)\text{th term} + \text{Value of}\left(\frac{10}{2}+1\right)\text{th term}}{2}$$

$$= \frac{\text{Value of 5th term} + \text{Value of 6th term}}{2}$$

$$= \frac{x+1+2x-13}{2} = \frac{3x-12}{2}$$

But given, median = 24

$$\therefore \quad \frac{3x-12}{2} = 24$$

$\Rightarrow$ $3x - 12 = 24 \times 2 = 48$
$\therefore$ $3x = 48 + 12$
$\Rightarrow$ $3x = 60$
$\therefore$ $\Rightarrow$ $x = 20$
Hence, the value of x is 20.

55. (a) $A \geq G \geq H$ between a and b
So, G is the Geometric mean between A and H.

56. (b) G.M. of a,b,c = 3

$$\Rightarrow (abc)^{\frac{1}{3}} = 3 \Rightarrow abc = 27$$

Also, a, b, c are in Geometric progression
So $a = 1$, $b = 3$, $c = 9$
Geometric mean of $d,e,f = 4$

$$\Rightarrow (def)^{\frac{1}{3}} = 4$$

$\Rightarrow def = 64$
Also, d, e, f are in Geometric progression
So, $d = 2$, $e = 4$, $f = 8$
Set = {1, 3, 9, 2, 4, 8}
Arithmetic mean

$$= \frac{1+3+9+2+4+8}{6} = \frac{27}{6} = \frac{9}{2} = 4.5 > 3.5$$

So option (b) is correct

57. (b) Votes cast in favour of E
$= 100000 - (36000 + 24000 + 18000 + 7000) = 15{,}000$
Angle allocated for party E in

$$\text{Pie chart} = \frac{360°}{100000} \times 15000 = 54°$$

Solutions (Q. Nos. 58-60):

Class	Mid Values	Frequency F_i	$d_i = x_i - 35$	$U_i = \frac{x_i - 35}{10}$	$f_i U_i$	Cumulative Frequency
0 – 10	5	4	–30	–3	–12	4
10–20	15	5	–20	–2	–10	9
20–30	25	7	–10	–1	–7	16
30–40	35	10	0	0	0	26
40–50	45	12	10	1	12	38
50–60	55	8	20	2	16	46
60–70	65	4	30	3	12	50

$$\sum F_i U_i = 11$$

$$N = \sum f_i = 50$$

$$\frac{N}{2} = \frac{50}{2} = 25$$

58. (a) $\text{Mean} = A + h\left[\frac{\sum fiUi}{N}\right] = 35 + 10 \times \frac{11}{50} = 37.2$

59. (b) Median Class = 30 – 40
$l = 30$, F = 16, $f = 10$, $h = 10$

60. (c) $\text{Median} = l + \frac{\frac{N}{2} - F}{f} \times h$

$$= 30 + \frac{25-16}{10} \times 10 = 30 + 9 = 39$$

61. (d) $\text{Mode} = l + \frac{f - f_1}{2f - f_1 - f_2} \times h$

Here, the maximum frequency is 12.50 their class is 40 – 50, then $l = 40$ $f = 12$, $f_1 = 10$, $f_2 = 8$

$$\text{mode} = 40 + \frac{12-10}{2 \times 12 - 10 - 8} \times 10$$

$$= 40 + \frac{2}{6} \times 10 = 40 + 3.33 = 43.33$$

62. (a) Let 5 observations be x, y, z, p and q.

$$\text{Mean} = \frac{x+y+z+p+q}{5}$$

But mean = 9

$\Rightarrow\ x+y+z+p+q=45$

Also 1 is subtracted from each observation, then

$$\text{Mean} = \frac{(x-1)+(y-1)+(z-1)+(p-1)+(q-1)}{5}$$

$$= \frac{x+y+z+p+q-5}{5} = \frac{45-5}{5} = 8$$

New mean = 8

Since n = 5 is odd

$$\text{Median} = \frac{n+1}{2} = \frac{5+1}{2} = 3$$

i.e., 3rd observation.

$\Rightarrow$ z is median

But median = p

$\Rightarrow\ z = p$

New median = 8 – 1 = 7

Mean = 8

Median = 7

$\therefore$ Option (a) is correct.

63. (c) $\frac{N}{2} = \frac{40}{2} = 20$

Age (in years)	No. of children	Cumulative frequency
5–6	4	4
6–7	7	11
7–8	9	(20)
8 9	12	32
9–10	6	38
10–11	2	40
Total	40	

Median class i.e, (7 – 8)

$$\text{Median} = \ell + \left(\frac{\frac{N}{2} - f_{pc}}{fm}\right) \times h$$

$$= 7 + \left(\frac{20-11}{9}\right) \times 1 = 7 + 1 = 8$$

Median age = 8 years

Hence, statement-1 is not correct.

Total number of childrens in the age group 6 – 9 years = 7 + 9 + 12 = 28 percentage of children in the age group 6 – 9 years.

$$= \frac{28}{40} \times 100 = 70$$

Hence statement-2 is correct.

Mode ℓ class is (8 – 9)

$$\text{Mode} = \ell + \left(\frac{f_m - f_{mp}}{2f_m - f_{mp} - f_{ms}}\right) \times h$$

$$= 8 + \left(\frac{12-9}{2\times 12 - 9 - 6}\right) \times 1 = 8 + \frac{1}{3} = 8$$

(Nearest integer)

Hence model age = 8 years

Therefore statement-3 is correct.

64. (c)

65. (b) Arranging the data in ascending order.

(a – 3.5), (a – 3), (a – 2.5), (a – 2), (a – 0.5), (a + 0.5), (a + 4), (a + 5).

Total terms = 8

$$\text{Medium} = \frac{\text{4th term} + \text{5th term}}{2}$$

$$= \frac{a-2+a-0.5}{2} = \frac{2a-2.5}{2} = a - 1.25$$

So, option (b) is correct.

66. (d) $(i+1)x^i = (i+1)\,x^i$ where i = 1, 2, 3, ___n

$$\sum_{i=1}^{n} i(i+1) = 1.2 + 2.3 + 3.4 + 4.5 + \text{-------meters}$$

$$= \sum_{n=1}^{n} T_n = \sum n(n+1)$$

$$= \sum n^2 + \sum n$$

$$= \frac{(n+1)n(2n+1)}{6} + \frac{n(n+1)}{2}$$

$$\text{Mean} = \frac{1}{n}\left[\frac{(n+1)n(2n+1)}{6} + \frac{n(n+1)}{2}\right]$$

$$= \frac{(n+1)}{2}\left[\frac{2n+1}{3} + 1\right] = \frac{(n+1)}{2}\frac{(2n+4)}{3}$$

$$= \frac{(n+1)(n+2)}{3}$$

So, option (d) is correct.

■■■