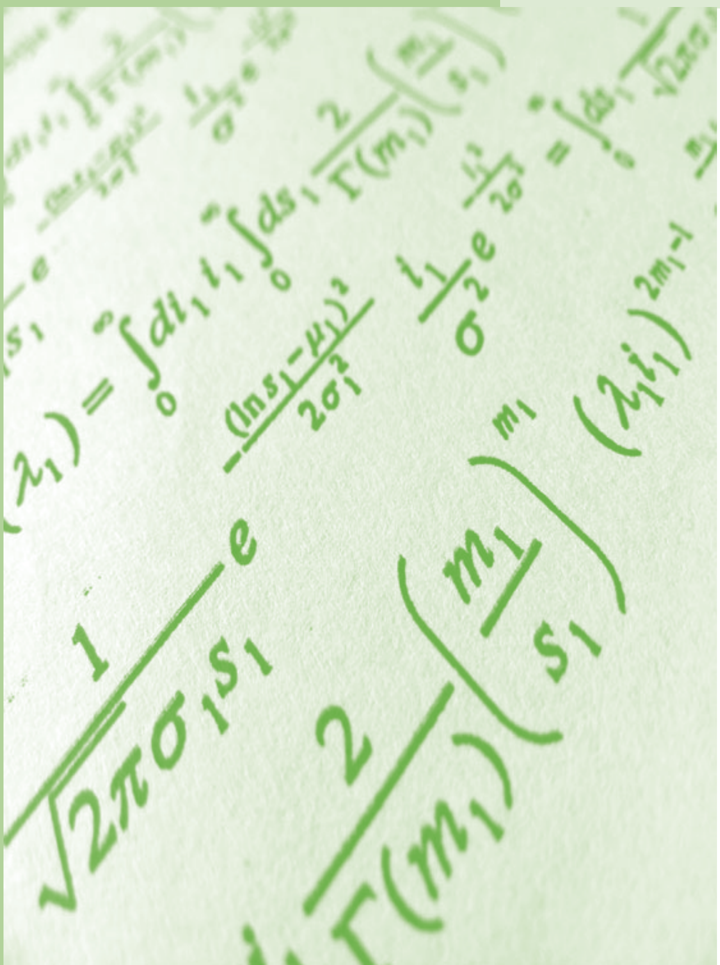
Chapter

12 Statistics



REMEMBER

Before beginning this chapter, you should be able to:

- Define data, types of data, tabulation of data, and statistical graphs
- Understand range, quartiles, deviation, etc

KEY IDEAS

After completing this chapter, you would be able to:

- Understand types of data and statistical groups
- Study measures of central tendencies for grouped and ungrouped data–mean, median and mode
- Evaluate empirical relation between mean, median and mode
- Learn about standard deviation, coefficient of variation, quartiles, estimation of median and quartiles from ogives, estimation of mode from histogram

INTRODUCTION

The word 'statistics' is derived from the Latin word 'status' which means political state.

Political states had to collect information about their citizens to facilitate governance and plan for development. Then, in course of time, statistics came to mean a branch of mathematics which deals with collection, classification, presentation and analysis of numerical data.

In this chapter, we shall learn about the classification of data, i.e., grouped data and ungrouped data, measures of central tendency, and their properties.

Data

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

For example, the percentage of marks scored by 10 students 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 scored by the 10 students in a class test.

Types of Data

Statistics is basically the study of numerical data. It includes methods of collection, classification, presentation, analysis of data and inferences from data. Data as 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. Data can also be classified as: raw data, and grouped data.

Raw Data

Data obtained from direct observation is called raw data.

The marks obtained by 10 students in a monthly test is an example of raw data or ungrouped data.

In fact, little can be inferred from this data. So, to make this data clearer and more meaningful, we group it into ordered intervals.

Grouped Data

To present the data in a more meaningful way, we condense the data into convenient number of classes or groups, generally not exceeding 10 and not less than 5. This helps us in perceiving at a glance, certain salient features of data.

Some Basic Definitions

Before getting into the details of tabular representation of data, let us review some basic definitions:

Observation Each numerical figure in a data is called an observation.

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

Tabulation or Presentation of Data

A systematical arrangement of the data in a tabular form is called 'tabulation' or 'presentation' of the data. This grouping results in a table called 'frequency table' that indicates the number of scores within each group.

Many conclusions about the characteristics of the data, the behaviour of variables, etc., can be drawn from this table.

The quantitative data that is to be analyzed statistically can be divided into three categories:

Individual Series Any raw data that is collected, forms an individual series.

Examples:

1. The weights of 5 students:

   32, 40, 65, 48 and 54 (in kg)

2. Percentage of marks scored by 10 students in a test:

   48, 59, 63, 72, 48, 72, 84, 98, 90 and 60

Discrete Series A discrete series is formulated from raw data. Here, the frequency of the observations are taken into consideration.

Example:

Given below is the data showing the number of computers in 12 families of a locality:

1, 1, 2, 3, 2, 1, 4, 3, 2, 2, 1, 1

Arranging the data in ascending order:

1, 1, 1, 1, 1, 2, 2, 2, 2, 3, 3, 4

To count, we can use tallymarks. We record tallymarks in bunches of five, the fifth one crossing the other four diagonally, i.e., $\cancel{\vert\vert\vert\vert}$

Thus, we may prepare the following frequency table.

Number of Computers	Tally Marks	Number of Families (Frequency)
1	$\cancel{\vert\vert\vert\vert}$	5
2	$\vert\vert\vert\vert$	4
3	$\vert\vert$	2
4	$\vert$	1

Continuous Series When the data contains large number of observations, we put them into different groups called 'class intervals', such as 1–10, 11–20, 21–30.

Here, 1–10 means data whose values lie between 1 and 10 including both 1 and 10.

This form is known as **'inclusive form'**. Also, 1 is called the **'lower limit'** and 10 is called the **'upper limit'**.

Example:

Given below are the marks (out of 50) obtained by 30 students in an examination.

43	19	25	32	48
17	29	9	15	50
7	24	20	37	44
22	2	50	27	25
18	42	16	1	33
25	35	45	35	28

Taking class intervals 1–10, 11–20, 21–30, 31–40 and 41–50, we prepare a frequency distribution table for the above data.

First, we write the marks in ascending order as:

1	2	7	9	15	16	17	18	19	20
22	24	25	25	25	27	28	29	32	33
35	35	37	42	43	44	45	48	50	50

Now, we can prepare the following frequency distribution table:

Class Interval	Tally Marks	Frequency
1–10	$\vert\vert\vert\vert$	4
11–20	$\cancel{\vert\vert\vert\vert}\ \vert$	6
21–30	$\cancel{\vert\vert\vert\vert}\ \vert\vert\vert$	8
31–40	$\cancel{\vert\vert\vert\vert}$	5
41–50	$\cancel{\vert\vert\vert\vert}\ \vert\vert$	7

Now, with this idea, let us review some more concepts about tabulation.

Class Interval

A group into which the raw data is condensed is called a class interval.

Each class is bounded by two figures which are called the class limits. The figure on the LHS is called lower limit and the figure on the RHS is called upper limit of the class. Thus, 0–10 is a class with lower limit being '0' and the upper limit being '10'.

Class Boundaries

In an exclusive form, the lower and upper limits are known as class boundaries or true lower limit and true upper limit of the class.

Thus, the boundaries of 15–25 in exclusive form are 15 and 25.

The boundaries in an inclusive form are obtained by subtracting 0.5 from the lower limit and adding 0.5 to the upper limit.

Thus, the boundaries of 15–25 in the inclusive form are 14.5–25.5.

Class Size

The difference between the true upper limit and the true lower limit is called 'class size'.

Hence, in the above example, the class size = 25 − 15 = 10.

Class Mark or Mid-value

Class mark $= \frac{1}{2}$(upper limit + lower limit)

Thus, the class mark of 15–25 is: $\frac{1}{2}(25 + 15) = 20$.

Statistical Graphs

Information provided by a numerical frequency distribution is easily understood when represented by diagrams or graphs. Diagrams act as visual aids and leave a lasting impression on the mind. This enables the investigator to make quick conclusions about the distribution.

There are different types of graphs or diagrams to represent statistical data. Some of them are:

1. Bar chart or bar graph (for unclassified frequency distribution)
2. Histogram (for classified frequency distribution)
3. Frequency polygon (for classified frequency distribution)
4. Frequency curve (for classified frequency distribution)
5. Cumulative frequency curve (for classified frequency distribution).
   (i) Less than cumulative frequency curve.
   (ii) Greater than cumulative frequency curve.

Bar Graph

The important features of bar graphs are:

1. Bar graphs are used to represent unclassified frequency distributions.
2. Frequency of a value of a variable is represented by a bar (rectangle) whose length (i.e., height) is equal (or proportional) to the frequency.
3. 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 12.1

Represent the following frequency distribution by a bar graph:

Value of variable	2	4	6	8	10
Frequency	5	8	4	2	7

SOLUTION

Either of the following bar graphs (Figs. 12.1 (a) or (b) may be used to represent the above frequency distribution.

The first graph takes value of the variable along the X-axis and the frequency along the Y-axis, whereas the second one takes the frequency along the X-axis and the value of the variable on the Y-axis.

All the rectangles (i.e., bars) should be of same width and uniform spaces should be left between any two consecutive bars.

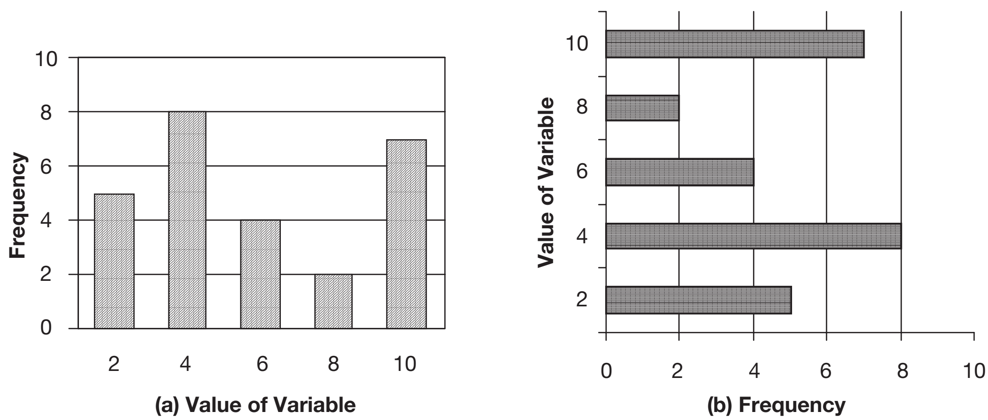


Figure 12.1

Classified or grouped data is represented graphically by histograms. A histogram consists of rectangles each of which has its breadth proportional to the size of concerned class interval and its height proportional to the corresponding frequency. In a histogram, two consecutive rectangles have a common side.

Hence, in a histogram, we do the following:

1. We represent class boundaries along the X-axis.
2. Along the Y-axis, we represent class frequencies.
3. We construct rectangles with bases along the X-axis and heights along the Y-axis.

EXAMPLE 12.2

Construct a histogram for the following frequency distribution.

Class Interval	20–30	30–40	40–50	50–60	60–70
Frequency	5	8	3	7	4

SOLUTION

Here, the class intervals are continuous.

The following histogram is drawn according to the method described above.

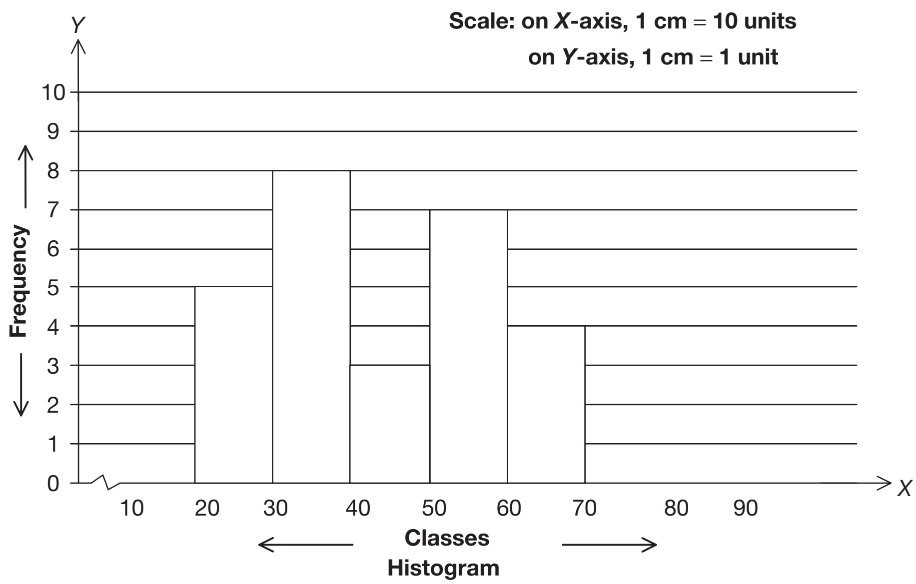


Figure 12.2

Remarks

The following points may be noted.

1. A link mark (——⋀——) made on the horizontal axis, between the vertical axis and first vertical rectangle, if there is a gap between 0 and the lower boundary of first class interval.
2. We may shade all rectangles. A heading for the histogram may also be given.

Important Observations

1. If the class intervals are discontinuous, the distribution has to be changed into continuous intervals, and then the histogram has to be drawn.
2. Bar graphs are used for unclassified frequency distributions, whereas histograms are used for classified frequency distribution. The breadths of rectangles in a bar graph are arbitary, while those in histogram are determined by class size.

Frequency Polygon

Frequency polygons are used to represent classified or grouped data graphically. It is a polygon whose vertices are the midpoints of the top sides of the rectangles, forming the histogram of the frequency distribution.

To draw a frequency polygon for a given frequency distribution, the mid-values of the class intervals are taken on *X*-axis and the corresponding frequencies on *Y*-axis and the points are plotted on a graph sheet.

These points are joined by straight line segments which form the frequency polygon.

EXAMPLE 12.3

Construct a frequency polygon for the following data:

Class Interval	12–17	18–23	24–29	30–35	36–41	Total
Frequency	10	7	12	8	13	50

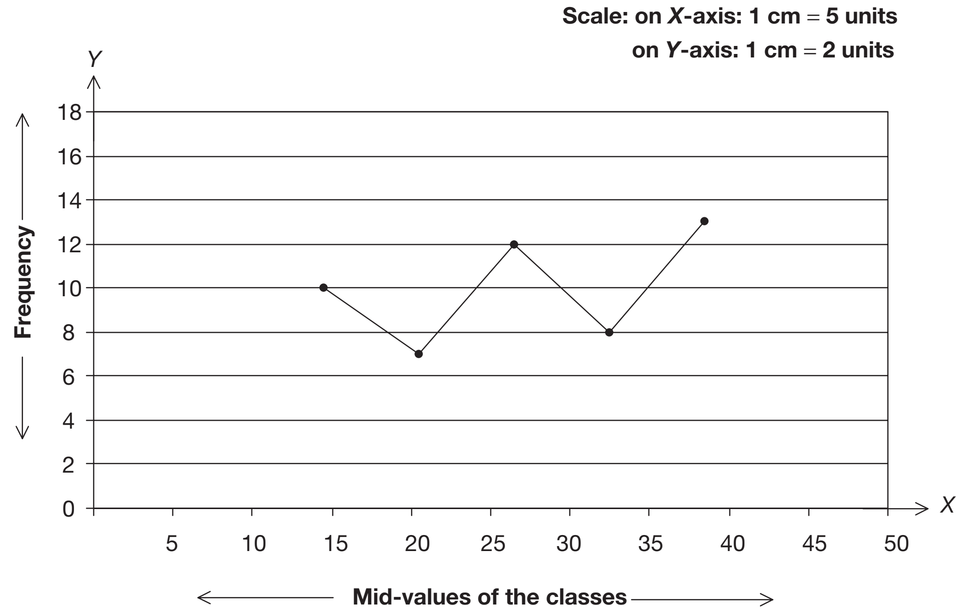
SOLUTION

Here, the class intervals are discontinuous.

Hence, first we convert the class intervals to continuous class intervals, and then find the mid-points of each class intervals. We do this by adding 0.5 to each upper limit and subtracting 0.5 from each lower limit.

Class Interval	Exclusive Class Interval	Mid-value of Class	Frequency
12–17	11.5–17·5	14.5	10
18–23	17.5–23.5	20.5	7
24–29	23.5–29.5	26.5	12
30–35	29.5–35.5	32.5	8
36–41	35.5–41.5	38.5	13

Now, taking the mid-values of class intervals on the *X*-axis and the corresponding frequencies on the *Y*-axis, we draw the frequency polygon as shown in the Fig. 12.3.



Frequency Polygon

Figure 12.3

Frequency Curve

Frequency curves are used to graphically represent classified or grouped data.

As the class-interval in a frequency distribution decreases, the points of the frequency polygon become closer, and closer, and then the frequency polygon tends to become a frequency curve. So, when the number of scores in the data is sufficiently large and the class-intervals become smaller (ultimately tending to zero), the limiting form of frequency polygon becomes frequency curve.

EXAMPLE 12.4

Draw a frequency curve for the following data

Mid-values	5	10	15	20	25	30	35	40
Frequency	2	4	7	5	10	12	6	4

SOLUTION

For the given (i,e., a table) showing the mid-values of classes and frequencies is made.

Now, taking the mid-values of the classes along the *X*-axis and the corresponding frequency along the *Y*-axis, we mark the points obtained from the above table in a graph sheet and join them with a smooth curve, which gives the following frequency curve:

Mid-value of Classes	Frequency
5	2
10	4
15	7
20	5
25	10
30	12
35	6
40	4

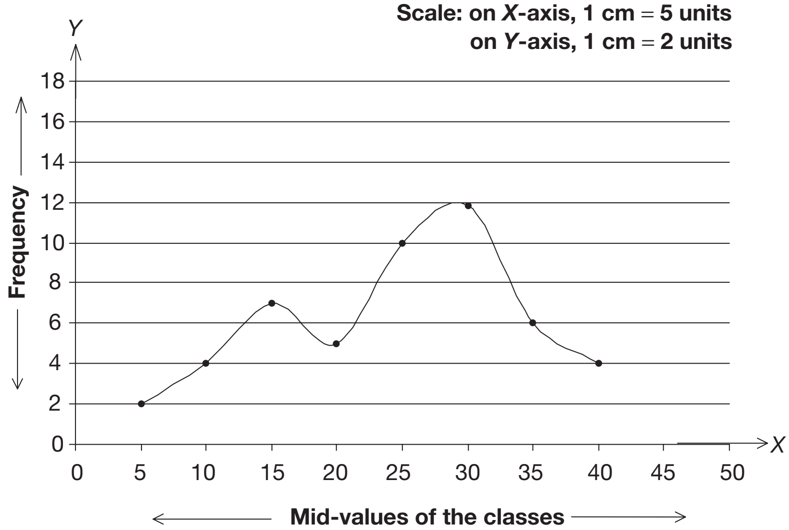


Figure 12.4

Cumulative Frequency Curves

The curves drawn for cumulative frequencies, less than or more than the true limits of the classes of a frequency distribution, are called 'cumulative frequency curves'. The curve drawn for the 'less than cumulative frequency distribution' is called the 'less than cumulative frequency curve' and the curve drawn for the greater than cumulative frequency distribution is called the 'greater than cumulative frequency curve'.

From these curves, we can find the total frequency above or below a particular value of the variable.

EXAMPLE 12.5

For the given distribution, draw the less than and greater than cumulative frequency curves.

Class	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100
Frequency	2	4	5	7	17	12	6	4	3

SOLUTION

Less than cumulative frequency distribution:

Upper Boundaries of the Classes	Frequency	Less than Cumulative Frequency
20	2	2
30	4	6
40	5	11
50	7	18
60	17	35
70	12	47
80	6	53
90	4	57
100	3	60

Greater than cumulative frequency distribution:

Lower Boundaries of the Classes	Frequency	Greater than Cumulative Frequency
10	2	60
20	4	58
30	5	54
40	7	49
50	17	42
60	12	25
70	6	13
80	4	7
90	3	3

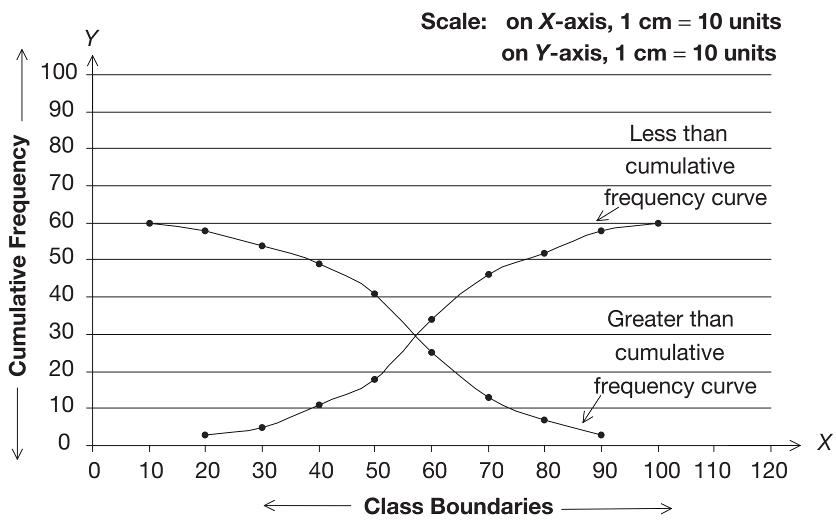


Figure 12.5

Measures of Central Tendencies for Ungrouped Data

Till now, we have observed that data collected in statistical enquiry or investigation happens in the form of raw data.

If data is very large, users cannot get much information. For this reason, data is grouped together to obtain some conclusions.

The measure of central tendency is a value which represents the total data, that is, it is the value in a data around which the values of all the other observations tend to concentrate.

The most commonly used measures of central tendency are the:

1. Arithmetic mean
2. Median
3. Mode

These measures give an idea about how the data is clustered or concentrated.

Arithmetic Mean or Mean (AM)

The arithmetic mean (or simply the mean) is the most commonly used measure of central tendency.

Arithmetic Mean for Raw Data The arithmetic mean of a statistical data is defined as the quotient obtained when the sum of all the observations or entries is divided by the total number of items.

If $x_1, x_2, ..., x_n$ are the n items, then:

$$\text{AM} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}, \text{ or briefly } \frac{\sum x}{n}.$$

AM is usually denoted by $\bar{x}$.

EXAMPLE 12.6

Find the mean of the first 10 natural numbers.

SOLUTION

Given data is 1, 2, 3, ..., 10

$$\therefore \text{ Arithmetic mean (AM)} = \frac{\text{Sum of observations}}{\text{Total number of observations}} = \frac{1+2+3+\cdots+10}{10} = \frac{55}{10} = 5.5.$$

Mean of Discrete Series Let $x_1, x_2, x_3, \ldots, x_n$ be n observations with respective frequencies $f_1, f_2, \ldots, f_n$.

This can be considered as a special case of raw data where the observation x_1 occurs f_1 times, x_2 occurs f_2 times, and so on.

$$\therefore \text{ The mean of the above data } = \frac{f_1x_1 + f_2x_2 + \cdots + f_nx_n}{f_1 + f_2 + \cdots + f_n}.$$

It can also be represented by $\bar{x} = \dfrac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$

Weighted Arithmetic Mean

When the variables $x_1, x_2, ..., x_n$ do not have same importance, and the weights $w_1, w_2, ..., w_n$ are given to each of the variables, the weighted arithmetic mean is given by $\bar{X}_w = \dfrac{\sum x_i w_i}{\sum w_i}$.

EXAMPLE 12.7

The salaries of 100 workers of a factory are given below.

Salary (in ₹)	Number of Workers
6000	40
8000	25
10000	12

Salary (in ₹)	Number of Workers
12000	10
15000	8
20000	4
25000	1
Total	100

Find the mean salaries of the workers of the factory.

SOLUTION

The mean $\bar{x}$ is given by:

$$\bar{x} = \frac{\begin{array}{c}(6000 \times 40) + (8000 \times 25) + (10000 \times 12) + (12000 \times 10) \\ + (15000 \times 8) + (20000 \times 4) + (25000 \times 1)\end{array}}{40 + 25 + 12 + 10 + 8 + 4 + 1}$$

$\Rightarrow \quad \bar{x} = ₹9050$

$\therefore$ The mean salary of the workers is ₹9050.

Some Important Results About AM

1. The algebraic sum of deviations taken about the mean is zero. That is, $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.
2. The value of the mean depends on all the observations.
3. The AM of two numbers a and b is $\frac{a+b}{2}$.
4. Combined mean: If $\bar{x}_1$ and $\bar{x}_2$ are the arithmetic means of two series with n_1 and n_2 observations, respectively, then the combined mean is:

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$

   The above result can be extended to any number of groups of data.
5. If $\bar{x}$ is the mean of $x_1, x_2, \ldots, x_n$, then the mean of $x_1 + a, x_2 + a, x_3 + a, \ldots, x_n + a$ is $\bar{x} + a$, for all values of a.
6. If $\bar{x}$ is the mean of $x_1, x_2, \ldots, x_n$, then the mean of $ax_1, ax_2, \ldots, ax_n$ is $a\bar{x}$ and that of $\frac{x_1}{a}, \frac{x_2}{a}, \ldots, \frac{x_n}{a}$ is $\frac{\bar{x}}{a}$.
7. The mean of the first n natural numbers is $\left(\frac{n+1}{2}\right)$.
8. The mean of the squares of the first n natural numbers $= \frac{(n+1)(2n+1)}{6}$.
9. The mean of the cubes of the first n natural numbers $= \frac{n(n+1)^2}{4}$.

EXAMPLE 12.8

If the average wage of 50 workers is ₹100 and the average wage of 30 of them is ₹120, then find the average wage of the remaining workers.

(a) ₹80 **(b)** ₹70 **(c)** ₹85 **(d)** ₹75

SOLUTION

We know that,

$$\text{Average} = \frac{\text{Sum of the quantities}}{\text{Number of the quantities}}$$

$$\therefore \frac{\text{Wage of 50 workers}}{50} = 100$$

$$\Rightarrow \quad \text{Wage}_{50} = 50 \times 1000 = ₹5,000$$

Similarly,

$$\frac{\text{Wage}_{30}}{30} = 120 \quad \Rightarrow \quad \text{Wage}_{30} = 30 \times 120 = ₹3,600$$

Now,

$$\text{Wage of 20 workers} = \text{Wage}_{50} - \text{Wage}_{30}$$
$$= 5000 - 3000$$
$$= ₹1,400$$

$$\Rightarrow \quad \text{Average wage}_{20} = \frac{1400}{20} = ₹70$$

∴ Hence, option (b) is the correct answer.

EXAMPLE 12.9

x	2	4	6	8
f	3	5	6	y

The mean of the above data is 5.5. Find the missing frequency (y) in the above distribution.

(a) 6 **(b)** 8 **(c)** 15 **(d)** 11

SOLUTION

Given AM = 5.5

$\Sigma f \cdot x = 2 \times 3 + 4 \times 5 + 6 \times 6 + 8xy$

$= 6 + 20 + 36 + 8y$

$\Sigma f \cdot 2 = 62 + 8y$ (1)

Now $\Sigma f = 3 + 5 + 6 + y$

$\Sigma f = 14 + y$ (2)

Using Eqs. (1) and (2)

$$\text{AM} = \frac{\sum f \cdot x}{\sum f} \quad \Rightarrow \quad 5.5 = \frac{62 + 8y}{14 + y}$$

$$\Rightarrow \quad 5.5(14 + y) = 62 + 8y$$

$$\Rightarrow \quad 77 + 5.5y = 62 + 8y$$

$$\Rightarrow \quad 8y - 5.5y = 77 - 62$$
$$\Rightarrow \quad 2.5y = 15$$
$$\Rightarrow \quad y = 6$$

$\therefore$ Hence, option (a) is the correct answer.

Median

Another measure of central tendency of a given data is the median.

Definition

If the values x_i in the raw data are arranged either in the increasing or decreasing order of magnitude, then the middle-most value in this arrangement is called the median.

Thus, for the raw (i.e., ungrouped) data, the median is computed as follows:

1. The values of the observations are arranged in order of magnitude.
2. The middle-most value is taken as the median. Hence, depending on the number of observations (odd or even), we determine median as follows.

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

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

EXAMPLE 12.10

Find the median of the following data: 2, 7, 3, 15, 12, 17 and 5.

SOLUTION

Arranging the given numbers in the ascending order, we have 2, 3, 5, 7, 12, 15, 17.

Here, the middle term is 7.

$\therefore$ Median = 7.

EXAMPLE 12.11

Find the median of the data 5, 8, 4, 12, 16 and 10.

SOLUTION

Arranging the given data in ascending order, we have 4, 5, 8, 10, 12, 16.

As the given number of values is even, we have two middle values. Those are $\left(\frac{n}{2}\right)$th and $\left(\frac{n}{2}+1\right)$th observations. Those are 8 and 10.

$\therefore$ Median of the data = Average of 8 and 10.

$$= \frac{8+10}{2} = 9.$$

Some Important Facts About Median

1. The median does not take into consideration all the items.
2. The sum of absolute deviations taken about the median is the least.
3. The median can be calculated graphically, but not the mean.
4. The median is not effected by extreme values.
5. The sum of deviations taken about median is less than the sum of absolute deviations taken from any other observation in the data.

EXAMPLE 12.12

A sequence, a, ax, ax^2, …, ax^n, has odd number of terms. Find its median.

(a) ax^{n-1} **(b)** $ax^{\frac{n}{2}-1}$ **(c)** $ax^{\frac{n}{2}}$ **(d)** $ax^{\frac{n}{2}+1}$

SOLUTION

$a, ax, ax^2, ax^3, \ldots, ax^n$

As there are odd number of terms, the median is:

$\left(\dfrac{n+1+1}{2}\right)$th term is $\left(\dfrac{n+2}{2}\right)$th term

Median $= a\left(x^{\frac{n+2}{2}-1}\right) = a \cdot x^{\frac{n}{2}}$.

Mode

The third measure of central tendency of a data is the mode.
The most frequently found value in the data is called the mode.
This is the measure which can be identified in the simplest way.

EXAMPLE 12.13

Find the mode of 0, 5, 2, 7, 2, 1, 1, 3, 2, 4, 5, 7, 5, 1 and 2.

SOLUTION

Among the given observations, the most frequently found observation is 2. It occurs 4 times.
$\therefore$ Mode = 2.

Notes

1. For a given data, the mode may or may not exist. In a series of observations, if no item occurs more than once, then the mode is said to be ill-defined.
2. If the mode exists for a given data, it may or may not be unique.
3. Data having unique mode is uni-model, while data having two modes is bi-model.

Properties of Mode

1. It can be graphically calculated.
2. It is not effected by extreme values.
3. It can be used for open-ended distribution and qualitative data.

Empirical Relationship Among Mean, Median, and Mode

For a moderately symmetric data, the above three measures of central tendency can be related by the formula, Mode = 3Median − 2Mean.

EXAMPLE 12.14

Find the mode when median is 12 and mean is 16 of a data.

SOLUTION

Mode = 3Median − 2Mean

= (3 × 12) − (2 × 16) = 36 − 32 = 4.

Observations

1. For a symmetric distribution,

   Mean = Median = Mode.
2. Given any two of the mean, median and mode, the third can be calculated.
3. This formula is to be applied in the absence of sufficient data.

Measure of Central Tendencies for Grouped Data

We studied the measure of central tendencies of ungrouped or raw data. Now, we study the measures of central tendencies (i,e., mean, median and mode) for grouped data.

Mean of Grouped Data

If the frequency distribution of 'n' observations of a variable x has k classes, x_i is the mid-value, and f_i is the frequency of ith class, then the mean $\overline{x}$ of grouped data is defined as:

$$\overline{x} = \frac{f_1x_1 + f_2x_2 + \cdots + f_kx_k}{f_1 + f_2 + \cdots + f_k} = \frac{\sum_{i=1}^{k} f_ix_i}{\sum_{i=1}^{k} f_i}$$

(or) simply, $\overline{x} = \dfrac{\sum f_ix_i}{N}$; where $N = \sum_{i=1}^{k} f_i$.

In grouped data, it is assumed that the frequency of each class is concentrated at its mid-value.

EXAMPLE 12.15

Calculate the arithmetic mean (AM) of the following data.

Percentage of marks	0–20	20–40	40–60	60–80	80–100
No. of students	2	12	13	15	8

SOLUTION

Let us write the tabular form as given below:

Percentage of Marks	Number of Students (f_i)	Midpoints of Classes (x_i)	$f_i x_i$
0–20	2	10	20
20–40	12	30	360
40–60	13	50	650
60–80	15	70	1050
80–100	8	90	720
	$\Sigma f_i = N = 50$		$\Sigma f_i x_i = 2800$

$$\therefore \text{ Mean } = \overline{x} = \frac{\sum fx}{N} = \frac{2800}{50} = 56.$$

Short-cut Method for Finding the Mean of Group Data (Deviation Method)

Sometimes, when the frequencies are large in number, the calculation of mean using the given formula is cumbersome. This can be simplified if the class interval of each class of grouped data is the same. Under the assumption of equal class interval, we get the following formula for the mean of grouped data:

$$\overline{x} = A + \frac{1}{N}\left(\sum_{i=1}^{k} f_i u_i\right) \times c$$

Where, A = assumed value from among mid-values

C = length of class interval

K = number of classes of the frequency distribution

N = Sum of frequencies $= \sum_{i=1}^{k} f_i$

$u_i = \frac{x_i - A}{C}$, $i = 1, 2, 3, \ldots, k$ and

x_i = mid-value of the ith class

u_i is called the deviation or difference of the mid-value of the ith class from the assumed value, divided by the class interval.

Using this method, the previous example can be worked out as follows:

Percentage of Marks	Number of Students (f_i)	Mid-values (x_i)	Deviation $u_i = \frac{x_i - A}{C}$	$f_i u_i$
0–20	2	10	−3	−6
20–40	12	30	−2	−24
40–60	13	50	−1	−13
60–80	15	70 (A)	0	0
80–100	8	90	1	8
	$N = 50$			$\Sigma f_i u_i = -35$

Here, $A = 70$; $N = 50$; $C = 20$; $\Sigma f_i u_i = -35$

$$\therefore \text{ AM } = A + \frac{1}{N}(\Sigma f_i u_i) \times c = 70 + \frac{1}{50}(-35) \times 20 = 70 - 14 = 56.$$

Median of Grouped Data

Before finding out how to obtain the median of grouped data, first we review what a median class is.

If 'n' is the number of observations, then from the cumulative frequency distribution, the class in which $\left(\frac{n}{2}\right)$th observation lies is called the median class.

Formula for calculating median:

$$\text{Median } (M) = L + \frac{\frac{n}{2} - F}{f}(C)$$

Where, L = Lower boundary of the median class, i.e., class in which $\left(\frac{n}{2}\right)$th observation lies.

n = Sum of frequencies

F = cumulative frequency of the class just preceding the median class

f = frequency of median class

C = length of class interval

EXAMPLE 12.16

Following is the data showing weights of 40 students in a class. Find its median.

Weight	45	46	47	48	49	50	51	52	53
No. of students	6	2	3	4	7	4	7	4	3

SOLUTION

To find the median, we prepare less than cumulative frequency table as given below.

Weight in kgs	No. of Students	Cumulative Frequency (*F*)
45	6	6
46	2	8
47	3	11
48	4	15
49	7	22
50	4	26
51	7	33
52	4	37
53	3	40

Here, $n = 40$ which is even.

$\therefore$ Median = value of $\frac{40}{2}$, or the 20th observation.

From the column of cumulative frequency, the value of 20th observation is 49.

$\therefore$ Median = 49 kg.

Note In the above example, we do not have any class interval. As there is no class interval, we cannot use the formula.

EXAMPLE 12.17

Find the median of the following data.

Class interval	0–10	10–20	20–30	30–40	40–50
Frequency	7	6	5	8	9

SOLUTION

To find the median, we prepare the following table:

Class Interval	Frequency	Cumulative Frequency
0–10	7	7
10–20	6	13 (*F*)
20–30	5 (*f*)	18
30–40	8	26
40–50	9	35
Total	35	

$$\text{Median } = L + \frac{\left(\frac{n}{2} - F\right)}{f} \times C \quad \text{Here, } n = 35 \quad \Rightarrow \quad \frac{n}{2} = \frac{35}{2} = 17.5$$

This value appears in the class 20–30.

L = lower boundary of median class (20−30) = 20

$F = 13$; $f = 5$ and $C = 10$ (class length)

$$\therefore \text{ Median } = 20 + \frac{\left(\frac{35}{2} - 13\right)}{5} \times 10 = 20 + \frac{9}{2 \times 5} \times 10 = 29.$$

Mode of Grouped Data

The formula for determining the mode of grouped data is $L_1 + \dfrac{\Delta_1 C}{\Delta_1 + \Delta_2}$.

Where, L_1 = lower boundary of model class (class with highest frequency).

$\Delta_1 = f - f_1$ and $\Delta_2 = f - f_2$, where f is the frequency of model class

f_1 = frequency of the previous class of the model class

f_2 = frequency of the next class of the model class.

Rewriting the formula:

$$\text{Mode } = L_1 + \frac{(f - f_1)C}{(f - f_1) + (f - f_2)}$$

$$\text{Mode } = L_1 + \frac{(f - f_1)C}{2f - (f_1 + f_2)}$$

EXAMPLE 12.18

The following information gives the monthly salaries of 100 employees. Find the mode of the data.

Salary (₹)	2000–3499	3500–4999	5000–6499	6500–7999	8000–9499
Number of Persons	35	37	21	12	5

SOLUTION

Here, the given classes are not continuous.

Hence, we first rewrite it as follows.

Salary	Adjusted Salary (₹)	No. of Persons
2000–3499	1999.5–3499.5	35 (f_1)
3500–4999	3499.5–4999.5	37 (f)
5000–6499	4999.5–6499.5	21 (f_2)
6500–7999	6499.5–7999.5	12
8000–9499	7999.5–9499.5	5

From the above table, it can be known that the maximum frequency occurs in the class interval 3500–4999.

$\therefore f = 37; f_1 = 35; f_2 = 29;$

$L_1 = 3499 \cdot 5;\ C = 1500$

$$\therefore \text{ Mode } = L_1 + \frac{(f - f_1)C}{(f - f_1) + (f - f_2)} = 3499.5 + \frac{1500(2)}{2+8} = 3499.5 + 300 = 3799.5.$$

EXAMPLE 12.19

Mode for the following distribution is 22 and $10 > y > x$. Find y.

Class interval	0–10	10–20	20–30	30–40	40–50	Total
Frequency	5	8	10	x	y	30

(a) 2 **(b)** 5 **(c)** 3 **(d)** 4

HINTS

(i) 10 is the model class.

(ii) Using mode $= L + \dfrac{\Delta_1}{\Delta_1 + \Delta_2} \times C$, we can get the value of x.

Range

The difference between the maximum and the minimum values of the given observations is called the range of the data.

Given $x_1, x_2, \ldots, x_n$ (n individual observations)

$$\text{Range} = (\text{Maximum value}) - (\text{Minimum value}).$$

EXAMPLE 12.20

Find the range of {2, 7, 6, 4, 3, 8, 5, 12}.

HINTS

By arranging the given data in the ascending order.

We have, {2, 3, 4, 5, 6, 7, 8, 12}

$\therefore$ Range = (Maximum value) − (Minimum value) = 12 − 2 = 10.

Note The range of the class interval is the difference of the actual limits of the class.

Calculation of Variance and Standard Deviation for Raw Data

Standard deviation (SD) is referred to as the root mean squared deviation about the mean. It is denoted by the symbol σ and read as sigma.

Variance is denoted by σ^2, which is the square of the standard deviation.

$\therefore$ Varience = $(\text{SD})^2$, or SD $= \sqrt{\text{varience}}$

$$\text{SD}(\sigma) = \sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2}{n}}$$

where, $x_1, x_2, \ldots, x_n$ are n observations with mean as $\bar{x}$.

Alternatively, the above formula can also be written as

$$\text{SD}(\sigma) = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}.$$

EXAMPLE 12.21

Calculate variance and standard deviation of the following data:

10, 12, 8, 14, 16

SOLUTION

$$\text{AM } \bar{x} = \frac{10+12+8+14+16}{5} = \frac{60}{5} = 12$$

$$\text{Varience } = \frac{(10-12)^2+(12-12)^2+(8-12)^2+(14-12)^2+(16-12)^2}{5}$$

$$= \frac{4+0+16+4+16}{5} = \frac{40}{5} = 8$$

$$\text{SD}(\sigma) = \sqrt{\text{variance}} = \sqrt{8}.$$

Calculation of Variance and SD for a Grouped Data

1. $N = \Sigma f =$ The sum of the frequencies

2. $\text{AM } (\bar{x}) = \dfrac{\sum fx}{N}$

3. D = Deviation from the AM $(x - \bar{x})$

4. Standard deviation (SD) $= \sigma = \sqrt{\dfrac{\sum fD^2}{N}}$

EXAMPLE 12.22

Calculate SD for the given data.

f	1	2	3	4
x	5	10	15	20

SOLUTION

f	x	fx	$D = x - \bar{x}$	D^2	fD^2
1	5	5	−10	100	100
2	10	20	−5	25	50
3	15	45	0	0	0
4	20	80	5	25	100
$\Sigma f = 10$		$\Sigma fx = 150$			$\Sigma fD^2 = 250$

$$\text{AM } (\bar{x}) = \frac{\sum fx}{N} = \frac{150}{10} = 15$$

$$\text{SD } = \sigma = \sqrt{\frac{\sum fD^2}{N}} = \sqrt{\frac{250}{10}} = 5.$$

EXAMPLE 12.23

Find the SD for the given data.

CI	0–10	10–20	20–30	30–40
Frequency	4	3	2	1

CI	f	x Mid-point	fx	$D = x - \bar{x}$	D^2	fD^2
0–10	4	5	20	−10	100	400
10–20	3	15	45	0	0	0
20–30	2	25	50	10	100	200
30–40	1	35	35	20	400	400
	$\Sigma f = 10$		$\Sigma fx = 150$			$\Sigma fD^2 = 1000$

$$\text{AM } = \bar{x} = \frac{\sum fx}{\sum f} = \frac{150}{10} = 15$$

$$\text{SD } = \sigma = \sqrt{\frac{\sum fD^2}{N}} = \sqrt{\frac{1000}{10}} = \sqrt{100} = 10.$$

EXAMPLE 12.24

If the standard deviation of $2x_i + 3$ is 8, then the variance of $\frac{3x_i}{2}$.

(a) 24 **(b)** 36 **(c)** 6 **(d)** 18

SOLUTION

SD $(2x_i) + 3 = 8$

SD of $2x_i = 8$ ($\because$ SD dose not alter when term is decreased by fixed constant.)

SD of $x_i = 4$

$\text{Var}(x_i) = \text{SD}^2 = 16$

$$\text{Var}\left(\frac{3x_i}{2}\right) = \left(\frac{9}{4}\right)\text{Var}(x_i) \quad (\because \text{Var}(ax + b) = a^2\,\text{Var}(x))$$

$$= \left(\frac{9}{4}\right)(16) = 36$$

$\therefore$ Hence, option (b) is the correct answer.

Coefficient of Variation (CV)

For any given data, let SD be the standard deviation and AM be the arithmetic mean, then coefficient of variation (CV) of the data is defined as:

$$\text{CV} = \frac{\text{SD}}{\text{AM}} \times 100.$$

This is a relative measure and helps in measuring the consistency. Smaller the CV, greater is the consistency.

EXAMPLE 12.25

In a series of observations, find the coefficient of variation given SD = 12.5 and AM = 50.

SOLUTION

$$\text{CV} = \frac{\text{SD}}{\text{AM}} \times 100 = \frac{12.5}{50} \times 100 = 25.$$

$\therefore$ Coefficient of Variation = 25.

Quartiles

In a given data, the observations that divide the given set of observations into four equal parts are called quartiles.

First Quartile or Lower Quartile

When the given observations are arranged in ascending order, the observation which lies midway between the lower extreme and the median is called the first quartile, or the lower quartile, and is denoted as Q_1.

Third Quartile or Upper Quartile

Of the data when the given observations are arranged in ascending order, the observation that lies in midway between the median and the upper extreme observation is called the third quartile, or the upper quartile, and is denoted by Q_3.

We can find Q_1 and Q_3 for an ungrouped data containing n observations as follows:

We arrange the given n observations or items in ascending order, then

lower or first quartile, Q_1 is $\left(\frac{n}{4}\right)$th item or observation if n is even and $\left(\frac{n+1}{4}\right)$th item or observation when n is odd.

EXAMPLE 12.26

Find Q_1 for 8, 12, 7, 5, 16, 10, 21 and 19.

SOLUTION

Arranging the given observations in ascending order.

We have, 5, 7, 8, 10, 12, 16,19, 21

Here, $n = 8$ (n is even)

$\therefore$ First quartile, $Q_1 = \left(\frac{n}{4}\right)$th item $= \left(\frac{8}{4}\right)$th item = 2nd item of the data, i.e., 7.

$\therefore$ $Q_1 = 7$.

EXAMPLE 12.27

Find Q_1 of the observations 21, 12, 9, 6, 18, 16 and 5.

SOLUTION

Arranging the observations in ascending order, we have 5, 6, 9, 12, 16, 18, 21.

Here, $n = 7$ (odd)

$\therefore$ $Q_1 = \left(\frac{n+1}{4}\right)$th item, i.e., $\left(\frac{7+1}{4}\right)$th item = 2nd item.

$\therefore$ $Q_1 = 6$.

EXAMPLE 12.28

The marks of 10 students in a class are 38, 24, 16, 40, 25, 27, 17, 32, 22, and 26. Find Q_1.

SOLUTION

The given observations, when arranged in ascending order, we have 16, 17, 22, 24, 25, 26, 27, 32, 38, 40

Here, $n = 10$ (even)

$$\therefore\ Q_1 = \left(\frac{n}{4}\right)\text{th item } = \left(2\frac{1}{2}\right)\text{th item of the data}$$

$$\therefore\ Q_1 = \text{2nd item } + \frac{1}{2}\text{(3rd} - \text{2nd) item } = 17 + \frac{1}{2}(22-17) = 17 + \frac{5}{2} = 19.5$$

$$\therefore\ Q_1 = 19.5$$

Third quartile: Q_3 is $\left(\frac{3n}{4}\right)$th item, when n is even and $3\left(\frac{n+1}{4}\right)$th item when n is odd.

EXAMPLE 12.29

Find Q_3 for 7, 16, 19, 10, 21 and 12.

SOLUTION

By arranging the data in ascending order, we have 7, 10, 12, 16, 19, 21.

Here, $n = 6$ (even)

$$\therefore\quad Q_3 = 3\left(\frac{n}{4}\right)\text{th item } = 4\frac{1}{2}\text{ th item}$$

$$\Rightarrow\quad Q_3 = \left[\text{4th} + \frac{1}{2}(\text{5th} - \text{4th})\right]\text{ item} = 16 + \frac{1}{2}(19-16)$$

$Q_3 = 17.5$.

Semi-inter Quartile Range or Quartile Deviation (QD)

Quartile deviation is given by the formula, QD $= \dfrac{Q_3 - Q_1}{2}$

EXAMPLE 12.30

Find semi-inter quartile range of the following data.

X	2	5	6	8	9	10	12
Frequency	1	8	12	16	11	9	3

SOLUTION

X	Frequency	Cumulative Frequency
2	1	1
5	8	9
6	12	21
8	16	37
9	11	48
10	9	57
12	3	60
	$N = 60$	

Here, $N = 60$

$$\therefore \quad Q_1 = \left(\frac{N}{4}\right)\text{th item} = \left(\frac{60}{4}\right)\text{th item} = 15\text{th item}$$

$\therefore\ Q_1 = 6$ (as the 15th item lies in 21 in the cumulative frequency)

$$Q_3 = 3\left(\frac{N}{4}\right)\text{th item} = 45\text{th item}$$

$\therefore\ Q_3 = 9$ (as 45th item lies in 48 in the cumulative frequency)

Semi-inter quartile range (QD) $= \dfrac{Q_3 - Q_1}{2} = \dfrac{9-6}{2} = 1.5.$

Note For an individual data, the second quartile Q_2 coincides with median.

$\Rightarrow \quad Q_2 =$ Median of the data.

EXAMPLE 12.31

The heights of 31 students in a class are given below.

Height (in cm)	126	127	128	129	130	131	132
Number of Students	7	3	4	2	5	6	4

1. Find the median of the above frequency distribution.
2. Find the semi-interquartile range of the above frequency distribution.

HINTS

1. Find the less than cumulative frequency, then find the median by using formulae.
2. Find the less than cumulative frequency, then find the inter-quartile range by using formulae

Estimation of Median and Quartiles from Ogive

1. Prepare the cumulative frequency table with the given data.
2. Draw ogive.
3. Let, total number of observations = sum of all frequencies = N.
4. Mark the points A, B and C on Y-axis, corresponding to $\dfrac{N}{4}, \dfrac{N}{2}$ and $\dfrac{3N}{4}$.
5. Mark three points (P, Q, R) on ogive corresponding to $\dfrac{3N}{4}, \dfrac{N}{2}$ and $\dfrac{N}{4}$.
6. Draw vertical lines from the points R, Q and P to meet X-axis Q_1, M and Q_3.
7. Then, the abscissas of Q_1, M and Q_3 gives lower quartile, median and upper-quartile.

EXAMPLE 12.32

The following table shows the distribution of the weights of a group of students.

Weight in Kg	30–35	35–40	40–45	45–50	50–55	55–60	60–65
No. of students	5	6	7	5	4	3	2

SOLUTION

Class Interval	Number of Students (Frequency)	Cumulative Frequency
30–35	5	5
35–40	6	11
40–45	7	18
45–50	5	23
50–55	4	27
55–60	3	30
60–65	2	32
	$N = 32$	

$N = 32$

$$\Rightarrow \quad \frac{N}{4} = 8, \frac{N}{2} = 16 \text{ and } \frac{3N}{4} = 24$$

From the graph:

Lower quartile $(Q_1) = 38$

Upper quartile $(Q_3) = 52$

Median $(M) = 44$

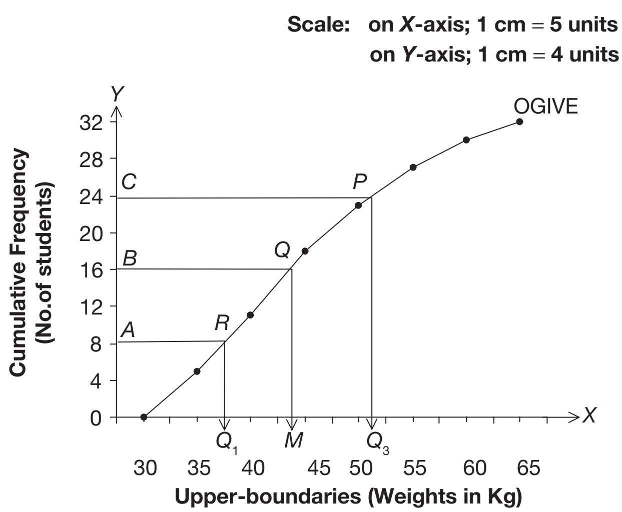


Figure 12.6

Estimation of Mode from Histogram

1. Draw a histogram to represent the given data.
2. From the upper corners of the highest rectangle, draw segments to meet the opposite corners of adjacent rectangles, diagonally as shown in the given example. Mark the intersecting point as *P*.
3. Draw *PM* perpendicular to *X*-axis, to meet *X*-axis at *M*.
4. Abscissa of *M* gives the mode of the data.

EXAMPLE 12.33

Estimate mode of the following data from the histogram.

CI	0–10	10–20	20–30	30–40	40–50	50–60	60–70
Frequency	10	16	17	20	15	13	12

From the graph, mode (M) = 34

SOLUTION

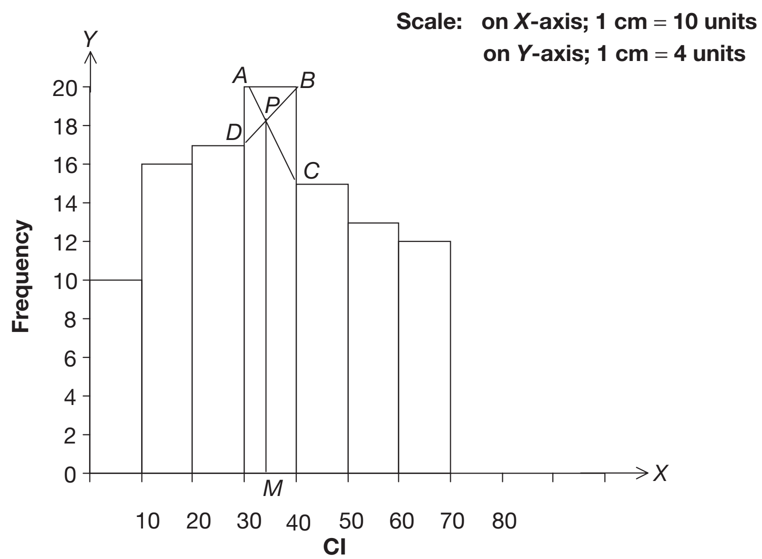


Figure 12.7

TEST YOUR CONCEPTS

Very Short Answer Type Questions

1. The class mark of a class is 25, and if the upper limit of that class is 40, then its lower limit is _______.
2. Consider the data: 2, 3, 2, 4, 5, 6, 4, 2, 3, 3, 7, 8, 2, 2. The frequency of 2 is _______.
3. 1–5, 6–10, 11–15, …, are the classes of a distribution, the upper boundary of the class 1–5 is _______.
4. 0–10, 10–20, 20–30,…, are the classes, the lower boundary of the class 20–30 is _______.
5. The mid-value of 20–30 is _______.
6. If 1–5, 6–10, 11–15,…, are the classes of a frequency distribution, then the size of the class is _______.
7. A class interval of a data has 15 as the lower limit and 25 as the size. Then, the class mark is _______.
8. In a histogram, the _______ of all rectangles are equal. (width/length/area)
9. The sum of 12 observations is 600, then their mean is _______.
10. If the lower boundary of the class is 25 and the size of the class is 9, then the upper boundary of the same class _______.
11. If 1–5, 6–10, 11–15, 16–20 ,…, are the classes of a frequency distribution, then the lower boundary of the class 11–15 is _______.
12. Arithmetic mean of first n natural numbers is _______.
13. The width of a rectangle in a histrogram represents frequency of the class. (True/False)
14. If 16 observations are arranged in ascending order, then the median is

$$\frac{(\text{8th observation } + \text{9th observation})}{2}. \text{(True/False)}$$

15. The mean of x, y, z is y, then $x + z = 2y$. (True/False)
16. Range of the scores 25, 33, 44, 26, 17 is _______.
17. Upper quartile of the data 4, 6, 7, 8, 9 is _______.
18. 2(Median – Mean) = Mode – Mean. (True/False)
19. Lower quartile of the data 5, 7, 8, 9, 10 is _______.
20. Consider the data: 2, x, 3, 4, 5, 2, 4, 6, 4, where $x > 2$. The mode of the data is _______.
21. Find the mean and the median of the data 10, 15, 17, 19, 20, and 21.
22. Find the semi-inter quartile range of the data: 32, 33, 38, 39, 36, 37, 40, 41, 47, 34, and 49.
23. Find the mean of first 726 natural numbers.
24. Find the range of the data: 14, 16, 20, 12, 13, 4, 5, 7, 29, 32, and 6.
25. Find the mean of the observations 425, 430, 435, 440, 445, …, 495. (difference between any two consecutive given observation is equal)
26. The mean of 10 observations is 15.5. By an error, one observation is registered as 13 instead of 34. Find the actual mean.
27. Observations of the certain data are $\frac{x}{8}, \frac{x}{4}, \frac{x}{2}, x, \frac{x}{16}$ where $x > 0$. If median of the given data is 8, then find the mean of the given data.
28. The mean of 12 observations is 14. By an error, one observation is registered as 24 instead of −24. Find the actual mean.
29. The mean weight of 20 students is 25 kg and the mean weight of another 10 students is 40 kg. Find the mean weight of the 30 students.
30. Find the variance and standard deviation of the scores 7, 8, 9, 10 and 16.

Short Answer Type Questions

31. Tabulate the given data by taking

    Class intervals: 1–10, 11–20, 21–30, 31–40

    Data: 9, 10, 8, 6, 7, 4, 3, 2, 16, 28, 22, 36, 24, 18, 27, 35, 19, 29, 23, 34.
32. If the mean and the median of a unimodel data are 34.5 and 32.5, then find the mode of the data.
33. The heights of 100 students in primary classes is classified as follows. Find the median.

Height (in cm)	Number of Students
81	22
82	14
83	26
84	23
85	15

34. The weight (in kg) of 25 children of 9th class is given. Find the mean weight of the children.

Weight (in kg)	Number of Children
40	3
41	4
42	6
43	2
44	5
45	5

35. If the mean of the following data is 5.3, then find the missing frequency y of the following distribution.

x	f
4	11
8	2
6	3
7	y

36. The mean of the data is 15. If each observation is divided by 5, and 2 is added to each result, then find the mean of the observations so obtained.

37. Draw the histogram for the following distribution.

Marks	Number of Students
0–10	3
10–20	4
20–30	8
30–40	9
40–50	6

38. Find the mode of the following data.

Class Interval	Frequency
1–5	3
6–10	4
11–15	10
16–20	6
21–25	7

39. A six-faced balanced dice is rolled 20 times, and the frequency distribution of the integers obtained is given below. Find the inter quartile range.

Integer	Frequency
1	3
2	4
3	2
4	5
5	4
6	2

40. Construct a less than cumulative frequency curve and a greater than cumulative frequency curve and answer the following.

Daily Wages (in ₹)	Number of Persons
20–30	5
30–40	12
40–50	17
50–60	36
60–70	20
70–80	10
80–90	8
90–100	2

(i) Find the number of persons who received ₹60 and more than ₹60.

(ii) Find the number of persons who received ₹90 and less than ₹90.

41. Draw the frequency polygon for the following distribution.

Class Interval	Frequency
0–5	8
5–10	12
10–15	20
15–20	16
20–25	4

42. Find the median of the following data.

Class Interval	Frequency
0–20	8
20–40	10
40–60	12
60–80	9
80–100	9

43. Construct a greater than cumulative frequency curve.

Class Interval	Frequency
5–9	1
10–14	5
15–19	10
20–24	19
25–29	25
30–34	21
35–39	15
40–44	3
45–49	1

44. Draw a histogram of the following data on a graph paper and estimate the mode.

Percentage of Marks	Number of Students
0–20	10
20–40	12
40–60	16
60–80	14
80–100	8

45. Find the coefficient of variation of the following dies create series.

Scores	Frequency
1	0
2	4
3	3
4	2
5	1

Essay Type Questions

46. If the mean of the following table is 30, then find the missing frequencies.

Class Interval	Frequency
0–15	10
15–30	a
30–45	b
45–60	8
Total	60

47. Calculate the AM of the following data using short-cut method.

Marks	Number of Students
0–10	3
10–20	4
20–30	6
30–40	8
40–50	9

48. For the following frequency distribution, construct a less than cumulative frequency curve. And also find Q_1, Q_2, Q_3 by using graph.

Class Interval	Frequency
0–9	4
10–19	3
20–29	5
30–39	6
40–49	1
50–59	2
60–69	1

49. Find the standard deviation of the following discrete series.

Scores	Frequency
1	0
2	4
3	3
4	2
5	1

50. Find the variance and SD for the given frequency distribution.

Class Interval	Frequency
1–5	4
6–10	1
11–15	2
16–20	3

CONCEPT APPLICATION

Level 1

1. If the arithmetic mean of the first n natural numbers is 15, then n is _______.

   (a) 15 (b) 30

   (c) 14 (d) 29

2. If the arithmetic mean of 7, 8, x, 11, 14 is x, then x is _______.

   (a) 9 (b) 9.5

   (c) 10 (d) 10.5

3. Find the mode of the data 5, 3, 4, 3, 5, 3, 6, 4, 5.

   (a) 5 (b) 4

   (c) 3 (d) Both (a) and (c)

4. The median of the data 5, 6, 7, 8, 9, 10 is _______.

   (a) 7 (b) 8

   (c) 7.5 (d) 8.5

5. If a mode exceeds a mean by 12, then the mode exceeds the median by _______.

   (a) 4 (b) 8

   (c) 6 (d) 10

6. If the less than cumulative frequency of a class is 50 and that of the previous class is 30, then the frequency of that class is _______.

   (a) 10 (b) 20

   (c) 40 (d) 30

7. If the median of the data, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8 is a, then find the median of the data x_3, x_4, x_5, x_6. (where $x_1 < x_2 < x_3 < x_4 < x_5 < x_6 < x_7 < x_8$)

   (a) a (b) $\frac{a}{2}$

   (c) $\frac{a}{4}$ (d) Cannot be determined

8. The mode of the data 6, 4, 3, 6, 4, 3, 4, 6, 5 and x can be:

   (a) Only 5 (b) Both 4 and 6

   (c) Both 3 and 6 (d) 3, 4 or 6

9. If the greater than cumulative frequency of a class is 60 and that of the next class is 40, then find the frequency of that class.

   (a) 10 (b) 20

   (c) 50 (d) Cannot be determined

10. If the difference between the mode and median is 2, then the difference between the median and mean is _______ (in the given order).

    (a) 2 (b) 4

    (c) 1 (d) 0

11. In a series of observations, SD is 7 and mean is 28. find the coefficient of variation.

    (a) 4 (b) $\frac{1}{4}$

    (c) 25 (d) 12.5

12. If the SD of x_1, x_2, x_3, …, x_n is 5, then find SD of $x_1 + 5$, $x_2 + 5$, $x_3 + 5$, …, $x_n + 5$.

    (a) 0 (b) 10

    (c) 5 (d) Cannot be determined

13. In a series of observations, coefficient of variation is 16 and mean is 25. Find the variance.

    (a) 4 (b) 8

    (c) 12 (d) 16

14. If the SD of y_1, y_2, y_3, …, y_n is 6, then variance of $(y_1 - 3)$, $(y_2 - 3)$, $(y_3 - 3)$, …, $(y_n - 3)$ is _______.

    (a) 6 (b) 36

    (c) 3 (d) 27

15. Lower quartile, upper quartile and interquartile range are Q_1, Q_3 and Q. If the average of Q, Q_1 and Q_3 is 40 and semi-interquartile range is 6, then find the lower quartile.

    (a) 24 (b) 36

    (c) 48 (d) 60

Direction for questions 16 and 17: These questions are based on the following data.

The weights of 20 students in a class are given below.

Weight (In kg)	Number of Students
31	6
32	3
33	5
34	2
35	4

16. Find the median of the above frequency distribution.

(a) 32.5 (b) 33

(c) 33.5 (d) Cannot be determined

17. The interquartile range of the above frequency distribution is _______.

(a) 4 (b) 3

(c) 2 (d) 1

18. If the average of a, b, c and d is the average of b and c, then which of the following is necessarily true?

(a) $(a + d) = (b + c)$

(b) $(a + b) = (c + d)$

(c) $(a - d) = (b - c)$

(d) $\frac{(a+b)}{4} = \frac{(b+c)}{2}$

19. Find the interquartile range of the data 3, 6, 5, 4, 2, 1 and 7.

(a) 4 (b) 3

(c) 2 (d) 1

20. If the mean of the lower quartile and upper quartile is 10 and the semi-interquartile range is 5, then the lower quartile and the upper quartile are _______ and _______.

(a) 2, 12 (b) 3, 13

(c) 4, 14 (d) 5, 15

21. The lower quartile of the data 5, 3, 4, 6, 7, 11, 9 is _______.

(a) 4 (b) 3

(c) 5 (d) 6

22. Find the arithmetic mean of the first 567 natural numbers.

(a) 284 (b) 283.5

(c) 283 (d) None of these

23. If $a < b < c < d$ and a, b, c, d are non-zero integers, the mean of a, b, c, d is 0 and the median of a, b, c, d is 0, then which of the following is correct?

(a) $b = -c$

(b) $a = -d$

(c) Both (a) and (b)

(d) None of these

24. The mean of 16 observations is 16. If one observation 16 is deleted and three observations 5, 5 and 6 are included, then find the mean of the final observations.

(a) 16 (b) 15.5

(c) 13.5 (d) None of these

25. If $L = 44.5$, $N = 50$, $F = 15$, $f = 5$ and $C = 20$, then find the median from of given data.

(a) 84.5 (b) 74.5

(c) 64.5 (d) 54.5

26. If $L = 39.5$, $\Delta_1 = 6$, $\Delta_2 = 9$ and $c = 10$, then find the mode of the data.

(a) 45.5 (b) 43.5

(c) 46.5 (d) 44.5

27. The average weight of 55 students is 55 kg, and the average weight of another 45 students is 45 kg. Find the average weight of all the students.

(a) 48 kg (b) 50 kg

(c) 50.5 kg (d) 52.25 kg

28. If the mean of 26, 19, 15, 24, and x is x, then find the median of the data.

(a) 23 (b) 22

(c) 20 (d) 21

Level 2

29. The mean and median of the data a, b and c are 50 and 35, where $a < b < c$. If $c - a = 55$, then find $(b - a)$.

(a) 8 (b) 7

(c) 3 (d) 5

30. If $a < b < 2a$, and the mean and the median of a, b and $2a$ are 15 and 12, then find a.

(a) 7 (b) 11

(c) 10 (d) 8

31. The variance of $6x_i + 3$ is 30, find the standard deviation of x_i.

(a) $\frac{5}{\sqrt{6}}$ (b) $\sqrt{\frac{5}{6}}$

(c) 30 (d) $\sqrt{30}$

32. The frequency distribution of the marks obtained by 28 students in a test carrying 40 marks is given below.

Marks	Number of Students
0–10	6
10–20	x
20–30	y
30–40	6

If the mean of the above data is 20, then find the difference between x and y.

(a) 3 (b) 2

(c) 1 (d) 0

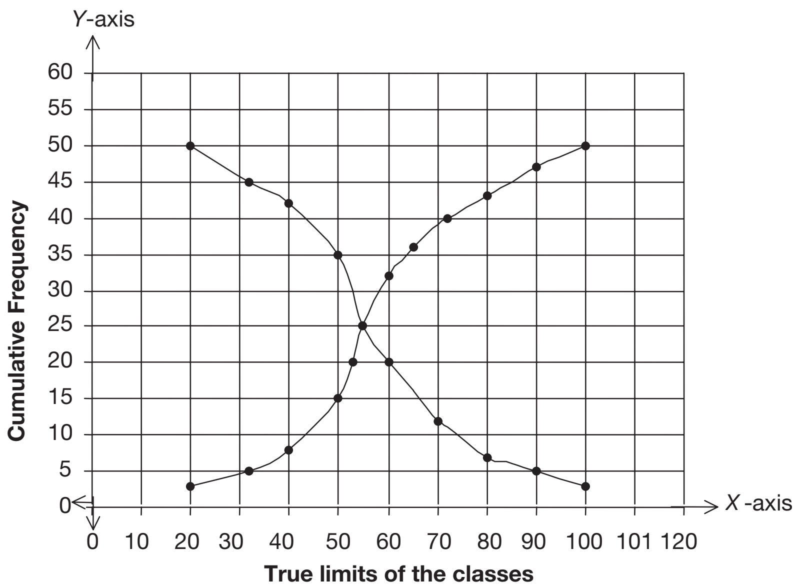


Figure 12.8

Direction for questions 37 and 38: These question are based on the following data above (figure)

The given figure represents the percentage of marks on the X-axis and the number of students on Y-axis.

33. Find the number of students who scored less than or equal to 50% of marks.

(a) 35 (b) 15

(c) 20 (d) 30

34. Find the number of students who scored greater than or equal to 90% of marks.

(a) 47 (b) 45

(c) 5 (d) 10

35. Find the variance of the scores 2, 4, 6, 8 and 10.

(a) 2 (b) 4

(c) 6 (d) 8

PRACTICE QUESTIONS

36. If $A = 55.5$, $N = 100$, $C = 20$, and $\Sigma f_i d_i = 60$, then find the mean from the given data.

(a) 67.5 (b) 57.5

(c) 77.5 (d) 47.5

37. Mode for the following distribution is 17.5 and x is less than 6. Find x.

Class Interval	Frequency
0–5	5
5–10	2
10–15	3
15–20	6
20–25	x

(a) 3 (b) 2

(c) 4 (d) 5

Direction for questions 38 and 39: These questions are based on the following data: consider the following distribution table.

Class Interval	Frequency
0–6	2
6–12	4
12–18	6

38. Find the coefficient of variation for the given distribution.

(a) $\frac{200\sqrt{6}}{11}$ (b) $\frac{200\sqrt{3}}{11}$

(c) $\frac{500}{11}$ (d) $\frac{200\sqrt{5}}{11}$

39. Find the variance for the given distribution:

(a) 24 (b) 12

(c) 20 (d) 25

Directions for questions 40 to 44: Select the correct alternative from the given choices.

40. Find the mean of the quartiles Q_1, Q_2 and Q_3 of the data 5, 9, 8, 12, 7, 13, 10, 14.

(a) 9 (b) 10

(c) 9.5 (d) 11.5

41. Which of the following cannot be determined?

(A) Range of the factors of 64

(B) Range of the first 10 positive integers

(a) A

(b) B

(c) Both (A) and (B)

(d) None of these

42. Find the mean of the following data.

Range of first n natural numbers, range of negative integers from $-n$ to -1 (where $-n < -1$), range of first n positive even integers and range of first n positive odd integers.

(a) $\frac{3}{2}(n-1)$ (b) $\frac{3n-2}{2}$

(c) $\frac{3}{2}(n-2)$ (d) $\frac{4n-3}{2}$

43. The following are the steps involved in finding the mean of the data.

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

Arrange them in sequential order.

(A) $\therefore$ Mean $= \frac{\sum fx}{\sum f} = \frac{110}{25}$

(B) $\Sigma fx = 10 + 24 + 30 + 28 + 18$

$\Sigma f = 1 + 3 + 5 + 7 + 9$

(C) $\therefore$ Mean $= 4.4$

(D) $\Sigma fx = 110$ and $\Sigma f = 25$

(a) ABDC (b) ACBD

(c) BDAC (d) BCAD

44. The mean weight of a group of 9 students is 19 kg. If a boy of weight 29 kg is joined in the group, then find the mean weight of 10 students.

The following are the steps involved in solving the above problem. Arrange them in sequential order.

(A) The mean weight of 10 students $= \dfrac{200}{10}$ kg

(B) The total weight of 9 students = 9 × 19 kg = 171 kg

(C) The total weight of 10 students = (171 + 29) kg = 200 kg

(D) ∴ The mean weight = 20 kg

(a) BCAD (b) BDAC

(c) BDCA (d) BCDA

Level 3

45. The arithmetic mean of the following data is 7. Find $(a + b)$.

x	f
4	a
6	4
7	b
9	5

(a) 4 (b) 2

(c) 3 (d) Cannot be determined

Direction for questions 46 and 47: The questions are based on the following data.

The performance of four students in annual report is given below.

Name of Student	Mean Score $(\bar{x})$	SD (σ)
Dheeraja	75	11.25
Nishitha	65	5.98
Sindhuja	48	8.88
Akshitha	44	5.28

46. Who is more consistent than the others?

(a) Dheeraja (b) Nishitha

(c) Sindhuja (d) Akshita

47. Who is less consistent than the others?

(a) Dheeraja (b) Nishitha

(c) Sindhuja (d) Akshitha

48. If the mean of the squares of first n natural numbers is 105, then find the median of the first n natural numbers.

(a) 8 (b) 9

(c) 10 (d) 11

49. Range of the scores 18, 13, 14, 42, 22, 26 and x is 44 $(x > 0)$. Find the sum of the digits of x.

(a) 16 (b) 14

(c) 12 (d) Cannot be determined

50. Find the arithmetic mean of the series 1, 3, 5, …, $(2n - 1)$.

(a) $\dfrac{2n-1}{n}$ (b) $\dfrac{2n+1}{n}$

(c) n (d) $n + 2$

51. The arithmetic mean of the squares of first n natural numbers is ______.

(a) $\dfrac{(n+1)(2n+1)}{6}$ (b) $\dfrac{n+1}{6}$

(c) $\dfrac{n^2-1}{6}$ (d) None of these

52. If X, M, Z are denoting mean, median and mode of a data and $X : M = 9 : 8$, then find the ratio $M : Z$.

(a) 8 : 9 (b) 4 : 3

(c) 7 : 6 (d) 5 : 4

53. The arithmetic mean of the series 1, 3, 3^2, …, 3^{n-1} is ______.

(a) $\dfrac{3^n}{2n}$ (b) $\dfrac{3^n-1}{2n}$

(c) $\dfrac{3^{n-1}}{n+1}$ (d) None of these

54. The mean of the data x, $x + a$, $x + 2a$, $x + 3a$, …, $(2n + 1$ terms) is ______.

(a) $x + (n - 1)a$ (b) $x + (n + 1)a$

(c) $x + (n + 2)a$ (d) $x + an$

55. The mean height of 25 boys in a class is 150 cm, and the mean height of 35 girls in the same class is 145 cm. The combined mean height of 60 students in the class is ______ (approximately).

(a) 143 cm (b) 146 cm

(c) 147 cm (d) 145 cm

56. The sum of 15 observations of a data is $(434 + x)$. If the mean of the data is x, then find x.

(a) 25 (b) 27

(c) 31 (d) 33

57. The mean weight of 9 students is 25 kg. If one more student is joined in the group the mean is unaltered, then the weight of the 10th student is ______ (in kg.)

(a) 25 (b) 24

(c) 26 (d) 23

58. Observations of some data are $\frac{x}{5}, x, \frac{x}{3}, \frac{2x}{3}, \frac{x}{4}, \frac{2x}{5}$ and $\frac{3x}{4}$ where $x > 0$. If the median of the data is 4, then find the value of 'x'.

(a) 5 (b) 7

(c) 8 (d) 10

TEST YOUR CONCEPTS

Very Short Answer Type Questions

1. 10
2. 5
3. 5.5
4. 20
5. 25
6. 5
7. 27.5
8. width
9. 50
10. 34
11. 10.5
12. $\frac{(n+1)}{2}$
13. False
14. True
15. True
16. 27
17. 8.5
18. False
19. 6
20. 4
21. Mean = 17.
    Median = 18
22. 3.5
23. 363.5
24. 28.
25. 460
26. 17.6
27. 12.4
28. 10
29. 30 kg
30. Variance = 10, SD $= \sqrt{10}$

Short Answer Type Questions

31.

Class Interval	Tally Marks	Frequency
1–10	~~||||~~ |||	8
11–20	|||	3
21–30	~~||||~~ |	6
31–40	|||	3

32. 28.5
33. 83 cm
34. 42.68 kg
35. 4
36. 5
38. 13.5
39. 3
40. (i) 40 (ii) 108
42. 50
44. 53
45. $33\frac{1}{3}$

Essay Type Questions

46. 18, 24
47. 30.33
49. 1
50. Variance = 41
    SD $= \sqrt{41}$

ANSWER KEYS

CONCEPT APPLICATION

Level 1

1. (d)	**2.** (c)	**3.** (d)	**4.** (c)	**5.** (b)	**6.** (b)	**7.** (a)	**8.** (d)	**9.** (b)	**10.** (c)
11. (c)	**12.** (c)	**13.** (d)	**14.** (b)	**15.** (c)	**16.** (b)	**17.** (b)	**18.** (a)	**19.** (a)	**20.** (d)
21. (a)	**22.** (a)	**23.** (c)	**24.** (d)	**25.** (a)	**26.** (b)	**27.** (c)	**28.** (d)		

Level 2

29. (d)	**30.** (b)	**31.** (b)	**32.** (d)	**33.** (b)	**34.** (c)	**35.** (d)	**36.** (a)	**37.** (a)	**38.** (d)
39. (c)	**40.** (c)	**41.** (d)	**42.** (a)	**43.** (c)	**44.** (a)				

Level 3

45. (d)	**46.** (b)	**47.** (c)	**48.** (b)	**49.** (c)	**50.** (c)	**51.** (a)	**52.** (b)	**53.** (b)	**54.** (d)
55. (c)	**56.** (c)	**57.** (a)	**58.** (d)						

CONCEPT APPLICATION

Level 1

1. Arithmatic mean of first n natural numbers is $\frac{n+1}{2}$.

2. Arithmatic mean $= \frac{\text{Sum of the quantities}}{\text{Number of the quantities}}$.

3. An observation which has more frequency in the data is called the mode of the data.

4. If the number of observations is even, then the median of the data is the average of $\left(\frac{n}{2}\right)$th and $\left(\frac{n}{2}+1\right)$th observations.

5. Use the 'empirical formula'.

6. Frequency of a particular class = cumulative frequency of that class − cumulative frequency of the previous class.

7. If the number of observations is even, then the median of the data is the average of $\left(\frac{n}{2}\right)$th and $\left(\frac{n}{2}+1\right)$th observations.

8. An observation which has more frequency in the data is called the mode of the data.

9. Frequency of a particular class = (cumulative frequency of that class) − (cumulative frequency of the next class).

10. Use the empirical formula.

11. Coefficient of variation $= \frac{\text{Standard deviation}}{\text{mean}} \times 100$.

12. SD does not alter when each term is increased by fixed constant.

13. Coefficient of variation $= \frac{\text{Standard deviation}}{\text{Mean}} \times 100$.

14. SD does not alter when each term is decreased by fixed constant.

15. Semi-interquartile range $= \frac{Q_3 - Q_1}{2}$.

16. Find the less than cumulative frequency, then find the median by using formulae.

17. Find the less than cumulative frequency, then find inter-quartile range by using formulae.

18. Average $= \frac{\text{Sum of the quantities}}{\text{Number of the quantities}}$.

19. $Q = Q_3 - Q_1$.

20. Semi-interquartile range $= \frac{Q_3 - Q_1}{2}$.

21. Write the data in the ascending order.

    Lower quartile is $\left(\frac{n+1}{4}\right)$ observation.

22. Arithmetic mean of first 'n' natural numbers is $\frac{n+1}{2}$.

23. Average $= \frac{\text{Sum of the quantities}}{\text{Number of the quantities}}$.

24. Mean $= \frac{\text{Sum of the observations}}{\text{Number of the observations}}$.

25. Median $= L + \left[\frac{\left(\frac{N}{2} - F\right)C}{f}\right]$.

26. Mode $= L_1 + \left[\frac{(f - f_1)C}{2f - (f_1 - f_2)}\right]$.

27. Average $= \frac{\text{Sum of the quantities}}{\text{Number of the quantities}}$.

28. Arithmatic mean $= \frac{\text{Sum of the quantities}}{\text{Number of the quantities}}$.

Level 2

29. (i) Since, a, b, c are in ascending order, b is the median.

    (ii) AM $= \frac{\text{Sum of all observations}}{\text{Total number of observations}}$.

    (iii) Using the above relation, write the relation between c and a, and find the value of a and c. Finally, find the values of b and a.

HINTS AND EXPLANATION

30. (i) Since, $a < b < 2a$, Median $= b = 12$

(ii) Mean $= \dfrac{\text{Sum of all observations}}{\text{Total number of observations}}$.

31. (i) If the variance of $(ax_i + b)$ is k, then SD of $(ax_i + b) = \sqrt{k}$.

(ii) SD of $x_i = \dfrac{\sqrt{k}}{a}$.

32. (i) Find the mid-values (x_i) of class intervals.

(ii) Find $\dfrac{\sum fx}{\sum f}$.

(iii) Equate $\dfrac{\sum fx}{\sum f}$ with mean, and find the values of x and y.

33. In the given graph, one curve represents the less than cumulative frequency, another curve represents the greater than cumulative frequency.

34. Apply the greater than cumulative frequency concept.

35. (i) SD of AP with common difference 'd' is $d\sqrt{\dfrac{n^2-1}{12}}$.

(ii) Variance $= (\text{SD})^2$.

36. (i) Use Arithmetic mean formula for grouped data.

(ii) Mean of grouped data

$= A + \dfrac{\Sigma f_i d_i}{N} \times C.$

38. Find mode of the given data in terms of x and form an equation to find x.

39. (i) First calculate the mean.

(ii) Find deviations about the mean $(x_i - \overline{x})$.

$D = x_i - \overline{x}$, $N =$ Sum of all the frequencies, SD $= \sqrt{\dfrac{\Sigma f D^2}{N}}$.

40. The increasing order of given the observations:

5, 7, 8, 9, 10, 12, 13, 14

$Q_1 = \left(\dfrac{n}{4}\right)$th observation

$Q_1 =$ 2nd observation

$\therefore\ Q_1 = 7$

$Q_2 =$ Mean of $\left(\dfrac{n}{2}\right)$th, $\left(\dfrac{n}{2}+1\right)$th observation

$=$ Mean of 4th, 5th items $= \dfrac{9+10}{2} = 9.5$

$Q_3 = \left(\dfrac{3n}{4}\right)$th $= \left(3 \times \dfrac{8}{4}\right)$th 6th observation

$Q_3 = 12$

Mean of Q_1, Q_2, Q_3 $= \dfrac{7+9.5+12}{3} = \dfrac{28.5}{3} = 9.5.$

41. (a) Range of the factors of 64 is $64 - 1 = 63$.

(b) Range of the first ten-positive integers is 9.

42. Range of the first 'n' natural numbers $= n - 1$

Range of the last n negative integers $= -1 - (-n) = n - 1$.

Range of the first 'n' positive even integers $= 2n - 2$.

Range of first 'n' positive odd integers $= (2n - 1) - (1) = 2n - 2$.

$\therefore$ Mean of the given data is

$\dfrac{(n-1)+(n-1)+(2n-2)+(2n-2)}{4}$

$= \dfrac{6n-6}{4} = \dfrac{3n-3}{2}$

$= \dfrac{3}{2}(n-1).$

43. BDAC is the required sequential order.

44. BCAD is the required sequential order.

Level 3

45. Use the arithmetic mean formulae for descreate series.

46. (i) Coefficient of variation $= \dfrac{\text{SD}}{\text{Mean}} \times 100$

(ii) Using the above, find CV of all the four members.

(iii) The member whose CV is least is more consistent.

47. (i) Find coefficient of variation then decide.

(ii) The one with highest CV is less consistent.

48. The mean of the squares of n natural numbers $\dfrac{(n+1)(2n+1)}{6}$.

49. Range is the maximum value − minimum value.

50. $1 + 3 + 5 + \cdots + (2n - 1) = n^2$

$$\therefore \text{AM} = \frac{n^2}{n} = n.$$

51. $\Sigma n^2 = \dfrac{n(n+1)(2n+1)}{6}$

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

52. Mode = 3 Median − 2 Mean

$Z = 3M - 2X$

Given, $X : M = 9 : 8$

$$\Rightarrow \frac{X}{M} = \frac{9}{8}$$

$$X = \frac{9M}{8}$$

$$\therefore \quad Z = 3M - 2\times\frac{9M}{8} = 3M - \frac{9M}{4}$$

$$Z = \frac{3M}{4} \quad \therefore \quad \frac{M}{Z} = \frac{4}{3}$$

$$\Rightarrow \quad M : Z = 4 : 3.$$

53. $\bar{x} = \dfrac{1+3+3^2+\cdots+3^{n-1}}{n}$

$$= \frac{1\left(\dfrac{3^n-1}{3-1}\right)}{n} = \frac{3^n-1}{2n}$$

54. Sum of the given observations $S_n = x + x + a + x + 2a + \ldots$ for $(2n + 1)$ terms

∴ The total number of terms in the given series is $(2n + 1)$ and first term $= x$ and common difference $= a$

$$= \frac{2n+1}{2}[2\cdot x + (2n+1-1)a]$$

$$= \frac{2n+1}{2}[2x + 2an] = (2n+1)(x+an)$$

$$\text{AM} = \frac{S_n}{2n+1} = \frac{(2n+1)(x+an)}{2n+1}$$

$\text{AM} = x + an.$

55. Given $n_1 = 25$, $n_2 = 35$

$\bar{x}_1 = 150$ and $\bar{x}_2 = 145$

Combined mean $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

$$\bar{x} = \frac{25\times150 + 35\times145}{25+35} = \frac{3750+5075}{60}$$

$$= \frac{8825}{60} = 147 \text{ (approximately)}.$$

56. Given sum of all observations = 434 + x and $n = 15$, $\bar{x} = x$

$$\therefore \quad \frac{434+x}{15} = x$$

$434 + x = 15x$

$$x = \frac{434}{14} \quad \Rightarrow \quad x = 31.$$

57. The sum of the weights of the 9 students = 25 × 9 = 225 kg.

If one more student is joined in the group, then total number of students is 10, and the mean is 25.

The sum of the weights of the 10 students is 25 × 10 = 250 kg.

The weight of the 10th student is 250 − 225 = 25 kg

58. $\dfrac{x}{5}\,\dfrac{x}{4}\,\dfrac{x}{3}\,\dfrac{2x}{5}\,\dfrac{2x}{3}\,\dfrac{2x}{4}\,x \quad \Rightarrow \quad \text{Median} = \dfrac{2x}{5}$

$$\Rightarrow \quad \text{Median} = \frac{2x}{5} = 4 \text{ (given)}$$

$$\Rightarrow \quad x = 10.$$