Unit 10 BASIC STATISTICS

After reading this unit, the students will be able to:

- Construct a grouped frequency table.
- Construct histograms with equal and equal class intervals;
- Construct a frequency polygon;
- Construct a cumulative frequency table.
- Draw a cumulative frequency polygon;
- Calculate (for ungrouped and grouped data)
    - arithmetic mean by definition and using deviation from assumed mean,
    - median, mode, geometric mean, harmonic mean.
- Recognize the properties of arithmetic mean;
- Calculate weighted mean and moving averages.
- Estimate median, quartiles and mode graphically
- Measure range, variance and standard deviation.

Introduction

Statistics is that branch of science which deals with the systematic collection of numerical facts or the discipline that includes procedures and techniques used to collect, process and analyze numerical data to make inference and to reach decision in the face of uncertainty.
Statistics has played an important role in almost every field of life like Agriculture, Industries, Business, Information Technology, Banking, Insurance Companies etc.

10.1 Frequency Distribution

The arrangement of data according to their frequencies is called frequency distribution, where frequency is the number of observations of particular values of a particular group.
There are two types of frequency distribution.

1) Ungrouped Frequency Distribution or Discrete Series

If the data is not organized in a proper order (group) is called ungrouped data.
Example: The result of 30 students in a test is given as:
20, 25, 22, 24, 23, 20, 19, 23, 22, 23, 24, 25, 21, 20, 22, 21, 24, 23, 25, 18, 19, 21, 24, 25, 23, 22, 21, 20, 19, 25 is the example of ungrouped data.

2) Grouped Frequency Distribution or Continuous Series
The data organized and summarized in the form of a frequency distribution is called grouped data. (See example 3 and 4)

The table 10.1 is a frequency distribution of the ages (years) of 50 students in a school.

Table 10.1

Age (Years)	Number of Students
9 — 10	10
11 —12	13
13 — 14	15
15 — 16	12

NOT FOR SALE

In the table 10.1, the age of 10 students is between 9 and 10 years, the age of 13 students is between 11 and 12 years, the age of 15 students is between 13 and 14 years and the age of 12 students is between 15 and 16 years. The number of students in each group represents the frequency for the respective group.

Table 10.2 is a frequency distribution of 40 students in a class during a monthly test.

Table 10.2

Classes (Marks)	Frequency (F)
20 — 24	4
25 —29	8
30 — 34	10
35 — 39	12
40 — 44	2
45 — 49	4

Table 10.2, 4 students obtained marks between 20 and 24, 8 students secured marks between 25 and 29, 10 students secured marks between 30 and 34, 12 students secured marks between 35 and 39, 2 students got marks between 40 and 44 and 4 students secured marks between 45 and 49.

10.1.1 Formation of Frequency Distribution

For the formation of Frequency Distribution, the following points must be taken.

i) Range
ii) Number of classes
iii) Class interval
iv) Class Limits
v) Distribution of Data
vi) Class Boundaries

i) Range

Range is the simplest measure of dispersion. It is the difference between the largest value and the smallest value in the data.

If the largest value is denoted by x_l and the smallest value is denoted by x_s then the range denoted by R is given as;

$R = x_l - x_s$

Consider the following examples:

Example 1

Find the range for the following set of data.

3, 5, 9, 7, 13, 10, 6.

Solution

Largest value = $x_l = 13$

Smallest value = $x_s = 13$

$\therefore$ Range $R = x_l - x_s$

$= 13 - 3$

$= 10$

Example 2

Find the range for the following set of data.

7, 3, 10, 9, 2, 15, 4, 16, 11.

Solution

Largest value = $x_l = 16$

Smallest value = $x_s = 2$

$\therefore$ Range $R = x_l - x_s$

$= 16 - 2$

$= 14$

ii) Number of classes

Number of classes must be balanced that is not too large or not too small and the general rule for getting a proper number of classes is to divide the difference of the highest and lowest values of the distribution by the class-interval. i.e

$$\text{Number of classes} = \frac{\text{Highest value} - \text{Lowest value}}{\text{Class interval}}$$

iii) Class interval

The difference of upper and lower class-boundary is called the class interval or width/length. The class-interval is also calculated by using the formula given in the box.

$$\text{Class-interval } (h) = \frac{\text{Largest value} - \text{Smallest value}}{\text{Number of classes}}$$

iv) Class Limits

Every class has two limits, the upper class limit and a lower class limit. Lower limit is represented by symbol ℓ_1 and upper limit by ℓ_2.

For each class the two limits may be fixed such that the mid-point of each class fall on an integer rather than a fraction.

The mid-point of each class is calculating by the formula.

$$\text{Mid-point} = \frac{\text{Lower limit} + \text{Upper limit}}{2}$$

For example if in a data the lowest value is 3 and the highest value is 20, then the better way is to make the classes as 3 — 7, 8 — 12, 13 —17, 18 — 22, because the mid-point of each class is an integer. The mid-points usually denoted by (x_i) are called the class marks.

v) Class Boundaries

For computing class-boundaries, the upper limit of the preceding class is subtracted from the lower limit of the following class and the difference is then divided by '2'. The quantity which is obtained is then subtracted from the lower limit and added to the upper limit of each class.

Example 1

Classes-Limits	Class-Boundaries
1 — 4	0.5 — 4.5
5 — 8	4.5 — 8.5
9 — 12	8.5 — 12.5
13 — 16	12.5 — 16.5
17 — 20	16.5 — 20.5

Formula for calculating class-boundaries (CB)

$$\frac{5-4}{2} = \frac{1}{2} = 0.5$$

So 0.5 is subtracted from the lower limit and added to the upper limit of each class.

Example 2

Classes-Limits	Class-Boundaries
1.5 — 3.4	1.45 — 3.45
3.5 — 5.4	3.45 — 5.45
5.5 — 7.4	5.45 — 7.45
7.5 — 9.4	7.45 — 9.45

Similarly if the class limits are given as above;
Now 0.05 is subtracted from the lower limit and added to the upper limit of each class.

Example 3

Prepare a frequency distribution from the marks of the students in a monthly test: 25, 30, 40, 21, 24, 25, 36, 30, 45, 50, 22, 25, 36, 46, 35, 38, 40, 28, 34, 45, 42, 46, 38, 48, 28, 29, 31, 33, 30, 26. Using 5 as class-interval. Also indicate the class-boundaries.

Class-Limits	Tally Marks	Frequency
21 — 25	~~IIII~~ I	6
26 — 30	~~IIII~~ II	7
31 — 35	IIII	4
36 — 40	~~IIII~~ I	6
41 — 45	III	3
46 — 50	IIII	4

Example 4

Construct a frequency table of the weights (kg) of 30 students from the following data by using 5 as a class-interval.

Find the class-boundaries and class-marks also.

25, 30, 40, 21, 24, 25, 36, 30, 45, 50, 22, 25, 36, 46, 35, 38, 40, 28, 34, 45, 42, 46, 38, 48, 28, 29, 31, 33, 30, 26.

Class-Limits	Tally Marks	Frequency	Class-Boundaries	Class Marks
21 — 25	~~IIII~~ I	6	20.5 — 25.5	23
26 — 30	~~IIII~~ II	7	25.5 — 30.5	28
31 — 35	IIII	4	30.5 — 35.5	33
36 — 40	~~IIII~~ I	6	35.5 — 40.5	38
41 — 45	III	3	40.5 — 45.5	43
46 —50	IIII	4	45.5 — 50.5	48

10.1.2 Graphic Presentation

(i) Histogram

A histogram consists of a set of adjacent rectangles whose basis are marked off by class-boundaries (not class-limits) on the X-axis and whose heights are proportional to the frequencies associated with the respective classes. The area of each rectangle represents the respective class frequencies.

Example 5

Construct a histogram from the following frequency distribution.

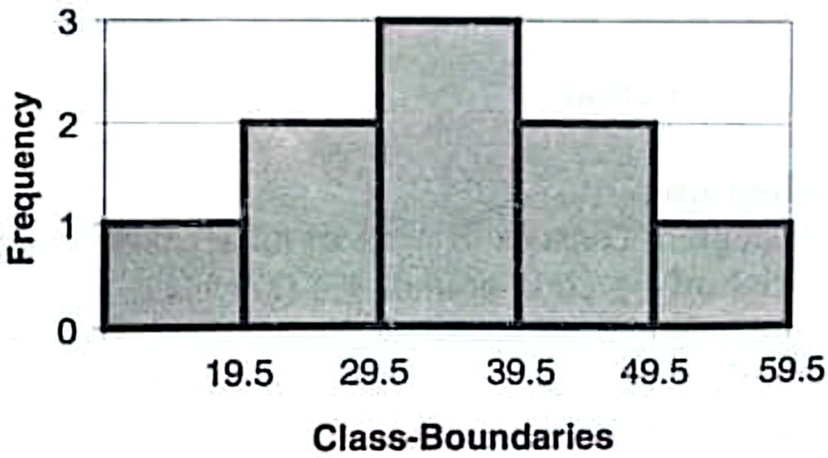
Class-limits	20 — 29	30 — 39	40 — 49	50 — 59	60 — 69
Frequency	1	2	3	2	1

Solution

To draw a histogram class-boundaries are marked along X-axis and frequencies of each class are marked along Y-axis as shown in the Figure.

Class-Limits	Frequency	Class-Boundaries
20 — 29	1	19.5 — 29.5
30 — 39	2	29.5 — 39.5
40 — 49	3	39.5 — 49.5
50 — 59	2	49.5 — 59.5
60 — 69	1	59.5 — 69.9

Histogram from Equal Intervals



Example 6

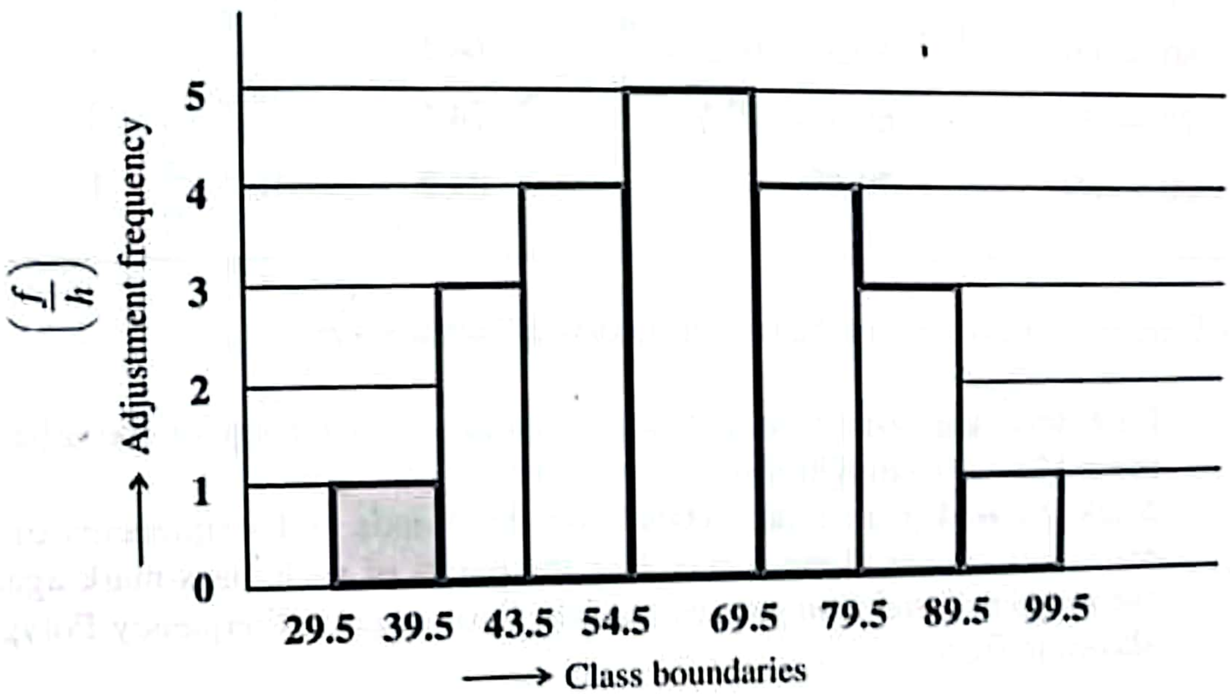
Draw a histogram for the following data.

Class Limits	30 — 39	40 — 43	44 — 54	55 — 69	70 — 79	80 — 89	90 — 99
Frequency	10	12	44	75	40	20	10

NOT FOR SALE

Solution Histogram with unequal intervals

Class-Limits	Class-Boundaries	Class-Intervals (h)	Frequency f	Adjusted Frequency $\frac{f}{h}$
30 — 39	29.5 — 39.5	10	10	1
40 — 43	39.5 — 43.5	4	12	3
44 — 54	43.5 — 54.5	11	44	4
55 — 69	54.5 — 69.5	15	75	5
70 — 79	69.5 — 79.5	10	40	4
80 — 89	79.5 — 89.5	10	20	2
90 — 99	89.5 — 99.5	10	10	1



(ii) **Frequency polygon**

A frequency polygon is the graphic form of a frequency distribution obtained by connecting with straight lines the mid-points (class-marks) of the top of the adjacent rectangles of a histogram.

Example 7

Construct a frequency polygon for the following frequency distribution.

Class Limits	20 — 29	30 — 39	40 — 49	50 — 59	60 — 69	70 — 79	80 — 89
Frequency	1	3	4	5	4	2	1

Solution

Classes	Class-Boundaries	Class Marks (Mid-Points)	Frequency
20 — 29	19.5 — 29.5	24.5	1
30 — 39	29.5 — 39.5	34.5	3
40 — 49	39.5 — 49.5	44.5	4
50 — 59	49.5 — 59.5	54.5	5
60 — 69	59.5 — 69.5	64.5	4
70 — 79	69.5 — 79.5	74.5	2
80 — 89	79.5 — 89.5	84.5	1

The Frequency Polygon can be drawn by two different ways.

(a) First draw the histogram and join the mid-points on top of the adjacent rectangles with straight lines.

(b) Mark the mid-point (class Marks) on the X-axis and frequencies of the respective classes along y-axis. Plot the points of each class-mark against the frequency and then join by straight lines to get a Frequency Polygon, shown in figures.

(ii) Frequency polygon

A frequency polygon is the graphic form of a frequency distribution obtained by connecting with straight lines the mid-points (class-marks) of the top of the adjacent rectangles of a histogram.

Example 7

Construct a frequency polygon for the following frequency distribution.

Class Limits	20 — 29	30 — 39	40 — 49	50 — 59	60 — 69	70 — 79	80 — 89
Frequency	1	3	4	5	4	2	1

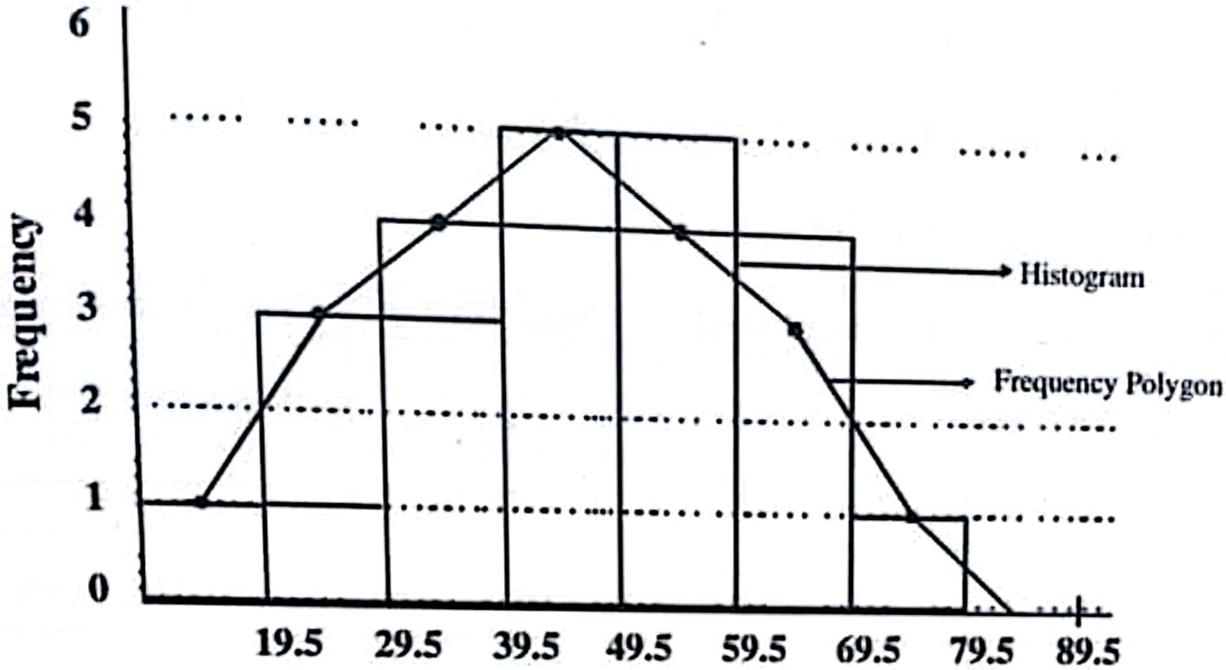
Solution

Classes	Class-Boundaries	Class Marks (Mid-Points)	Frequency
20 — 29	19.5 — 29.5	24.5	1
30 — 39	29.5 — 39.5	34.5	3
40 — 49	39.5 — 49.5	44.5	4
50 — 59	49.5 — 59.5	54.5	5
60 — 69	59.5 — 69.5	64.5	4
70 — 79	69.5 — 79.5	74.5	2
80 — 89	79.5 — 89.5	84.5	1

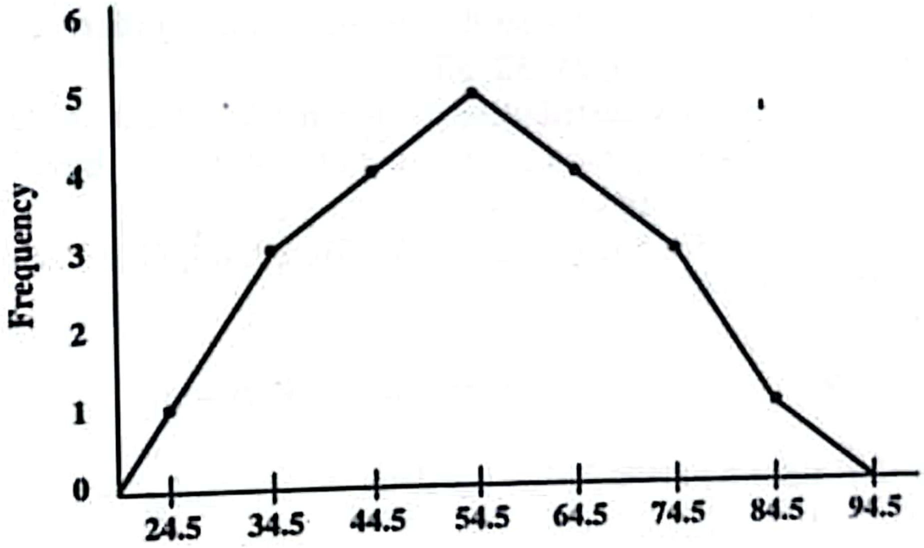
The Frequency Polygon can be drawn by two different ways.

(a) First draw the histogram and join the mid-points on top of the adjacent rectangles with straight lines.

(b) Mark the mid-point (class Marks) on the X-axis and frequencies of the respective classes along y-axis. Plot the points of each class-mark against the frequency and then join by straight lines to get a Frequency Polygon, shown in figures.



Frequency Polygon



NOT FOR SALE

Exercise 10.1

1. Define statistics. Describe at least five applications of statistics in our practical life.
2. What is meant by frequency distribution?
3. Complete the following table.

Class-Limit	Class-Interval	Class-Marks	Class-Boundaries
1 — 3	3	2	0.5 — 3.5
2 — 4			
3.1 — 3.5			
1.10 — 1.20			

4. Write the following data in ascending or descending order.
20, 25, 34, 42, 58, 33, 40, 24, 32, 50.
5. Construct a frequency distribution of the marks of 30 students during a quiz with 100 points by taking 10 as the class-interval. Indicate the class-boundaries and class-marks.
40, 60, 65, 70, 35, 50, 56, 74, 72, 49, 85, 76, 82, 83, 68, 90, 67, 66, 58, 46, 74, 88, 76, 69, 57, 63, 66, 47, 82, 90.
6. Define
i) Histogram ii) Frequency Polygon
7. Draw a Histogram for the following data.

Class-Limit	20 — 24	25 — 29	30 — 34	35 — 39	40 — 44	45 — 49	50 — 54
Frequency	1	3	4	5	4	2	1

NOT FOR SALE

8. The following data gives the weights in (kg) of the students in the 9th class.
25, 30, 32, 29, 24, 40, 36, 37, 28, 27, 41, 42, 35, 39, 31, 32, 34, 42, 40, 43, 36, 26, 22, 23, 42, 39, 35, 41, 39, 29.
(i) Prepare a frequency distribution using a suitable class interval.
(ii) Draw histogram and frequency polygon.

10.2 Cumulative Frequency

If $C_1, C_2, C_3 \ldots C_n$ are n classes given in a data with $f_1, f_2, f_3 \ldots, f_n$, are the frequencies of the respective classes then the sum of frequencies upto each term is called **cumulative frequency**.

Example 8

Find the cumulative frequency of the following data.

x	3	4	5	6	7	8	9	10	11	12
f	1	2	3	4	5	6	7	4	3	8

Solution:

x	f	Method of finding (c.f)	c.f
3	1	1	1
4	2	$1+2=3$	3
5	3	$3+3=6$	6
6	4	$6+4=10$	10
7	5	$10+5=15$	15
8	6	$15+6=21$	21
9	7	$21+7=28$	28
10	4	$28+4=32$	32
11	3	$32+3=35$	35
12	8	$35+8=43$	43

10.2.1 Cumulative Frequency Polygon:

A cumulative frequency polygon is the graphic form of a frequency distribution obtained by plotting the cumulative frequencies of a distribution against the upper or lower class-boundaries depending upon whether the cumulative frequencies is of the "less than" or "more than" type and the points are joined by straight lines.

Example 9

In the following data marks of students are given during first pre-board exam in the subject of maths.

25, 30 27, 28, 35, 36, 40, 41, 42, 45, 50, 44, 29, 26, 36, 31, 43, 46, 52, 53, 51, 42, 37, 27, 33, 46, 44, 34, 51, 54

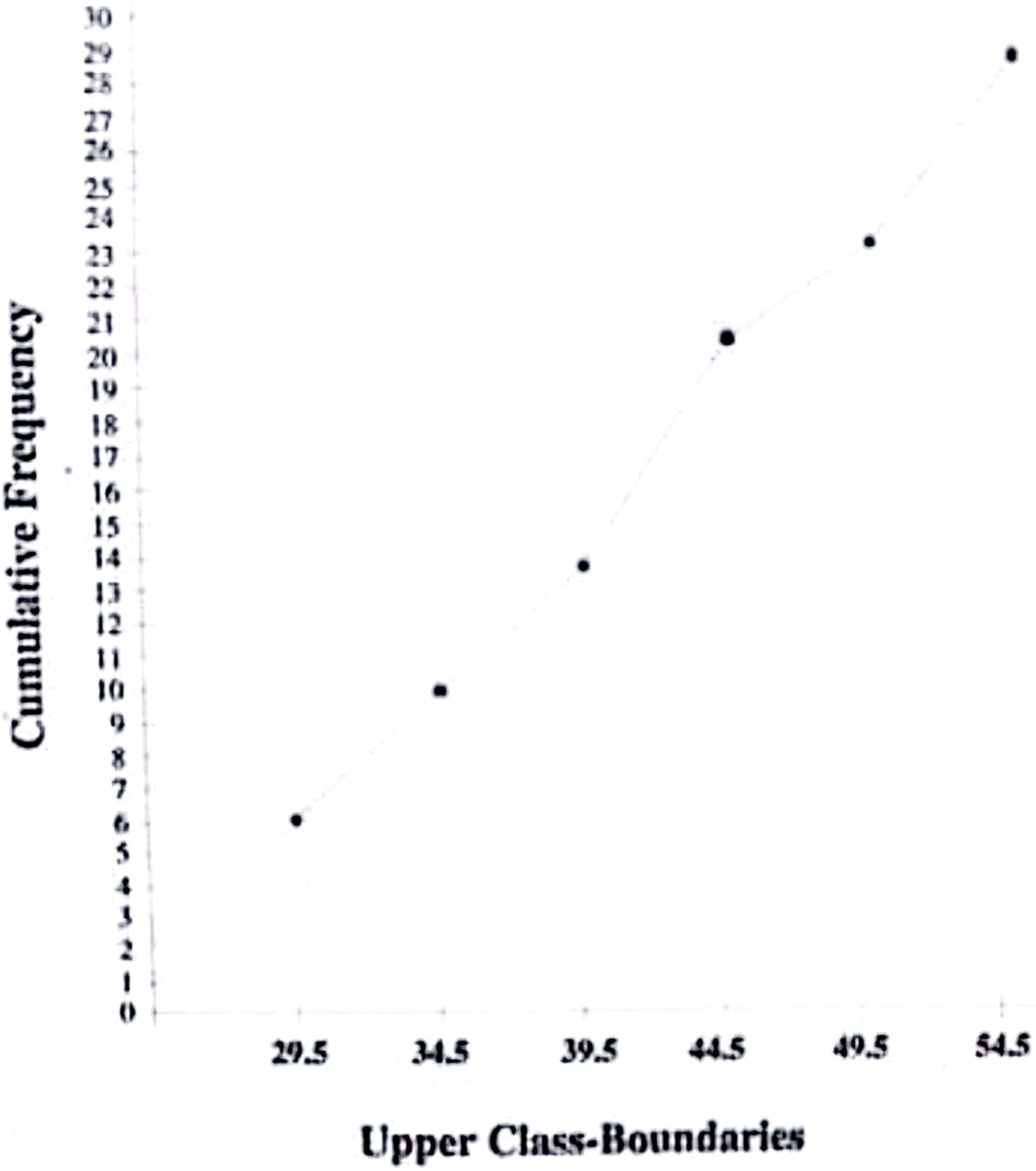
By taking a suitable class-interval, prepare a frequency distribution, find less than cumulative frequency and draw a cumulative frequency polygon.

Solution

To construct the cumulative frequency distribution, we take the class-interval as 5.

Class-limits	Frequency	Class-boundaries	Cumulative Frequency
25 — 29	6	24.5 — 29.5	6
30 — 34	4	29.5 — 34.5	10
35 — 39	4	34.5 — 39.5	14
40 — 44	7	39.5 — 44.5	21
45 — 49	3	44.5 — 49.5	24
50 — 54	6	49.5 — 54.5	30

Cumulative Frequency Polygon



Exercise 10.2

1. The following data give the wages (in Rs) of workers.

   60, 75, 80, 85, 90, 84, 70, 73, 76, 84, 95, 100, 150, 66, 58, 90, 98, 120, 77, 90.

   By taking 10 as the class-interval, prepare.

   i) Frequency distribution

   ii) Cumulative frequency distribution

   iii) Draw a histogram

iv) Frequency polygon

v) Cumulative frequency polygon

2. Draw less than and more than cumulative frequency polygon for the data given.

Marks	Number of Students
40 — 49	1
50 — 59	2
60 — 69	3
70 — 79	4
80 — 89	5
90 — 99	6

3. Determine from the data of Q2, the following.
   i) Number of students obtained marks more than 50.
   ii) Number of students obtained marks less than 70.
   iii) Number of students secured marks between 50 and 70.
   iv) Class interval of all classes
   v) Lower-class boundary of 5th class.

10.3 Measure of Central Tendency or Average

10.3.1 Definition of Central Tendency

Central Tendency of a data is the value that represents the whole data or the stage at which the largest number of items tends to concentrate and so it is called central tendency. Central Tendency or Averages are also sometimes called measures of location, because they locate the centre of a distribution.

10.3.2 Types of Averages

There are two types of Averages

1) Mathematical Averages
   - i) Arithmetic Mean (A.M)
   - ii) Geometric Mean (G.M)
   - iii) Harmonic Mean (H.M)

2) Averages of Positions
   - (i) Median
   - (ii) Mode
   - (iii) Quartiles

1) **Mathematical averages**

i) **Arithmetic Mean**

Arithmetic Mean (for ungrouped data)

a) By direct method

Arithmetic Mean or simply Mean is calculated by adding all values of the data divided by the number of items (values). If $x_1, x_2, x_3, \ldots, x_n$ are n values; then the Arithmetic Mean denoted by ($\bar{x}$) is given as

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

$\bar{x} =$ Arithmetic Mean, $\sum$ = sigma (Notation used for the summation)

$x_i = x_1, x_2, x_3, \ldots x_n$ ($i = 1, 2, 3 \ldots n$)

n = Total number of items in the data.

Example 10

Find the A.M of the values 4, 5, 6, 7, 8, 9, 10, 3, 2

Solution

Let $\bar{x}$ be the A.M of the given items. Then by using formula

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 + x_9 \ldots x_n}{n}$$

$$= \frac{2+3+4+5+6+7+8+9+10}{9} = \frac{54}{9} = 6$$

$\therefore \bar{x} = 6$

b) **By short cut method**

Formula for short cut method is,

$$\bar{x} = a + \frac{\sum D_x}{n} \text{ Where}$$

$\bar{x}$ = Arithmetic Mean

a = Provisional Mean (P.M)

$D_x = (x - a)$ (Deviation from Provisional Mean)

$\sum D_x$ = Sum of Deviations from P.M

n = Total number of values in the data.

Example 11

Calculate the mean by shortcut method of the values 2, 3, 4, 5, 6, 7, 8, 9, 10.

Solution

x	D_x $(x - a)$, a = 6
2	$2 - 6 = -4$
3	$3 - 6 = -3$
4	$4 - 6 = -2$
5	$5 - 6 = -1$
6	$6 - 6 = 0$
7	$7 - 6 = 1$
8	$8 - 6 = 2$
9	$9 - 6 = 3$
10	$10 - 6 = 4$
	$\sum D_x = 0$

By using formula

$$\bar{x} = a + \frac{\sum D_x}{n} = 6 + \frac{0}{9} = 6 + 0$$

$\therefore \bar{x} = 6$

Arithmetic Mean (for grouped data)

a) Direct method

If $x_1, x_2, x_3, \ldots, x_n$ are n values with $f_1, f_2, f_3 \ldots f_n$, as frequencies of the respective values, then the Arithmetic Mean $\bar{x}$ is given as

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \ldots + f_n x_n}{f_1 + f_2 + f_2 + \ldots + f_n} = \frac{\sum fx}{\sum f}$$

Example 12

Calculate the A.M from the following data by Direct Method.

x	20	30	40	50	60
f	1	2	2	3	4

Solution

x	f	fx
20	1	20
30	2	60
40	2	80
50	3	150
60	4	240
	$\sum f = 12$	550

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{550}{12} = 45.83$$

Example 13

Calculate the A.M from the following data by Direct Method.

Classes	0 — 10	10 — 20	20 — 30	30 — 40	40 — 50
Frequency	2	3	4	2	1

Solution

Classes	x (mid-point)	f	fx
0 — 10	5	2	10
10 — 20	15	3	45
20 — 30	25	4	100
30 — 40	35	2	70
40 — 50	45	1	45
		$\sum f = 12$	$\sum fx = 270$

$$A.M = (\bar{x}) = \frac{\sum fx}{\sum f} = \frac{270}{12} = 22.5$$

b) Short cut method

$$(\bar{x}) = a + \frac{\sum fD_x}{\sum f}$$

Example 14

Calculate A.M in Example 13 by shortcut method.

Solution

Classes	x	D_x $(x — 25)$	f	fD_x
0 — 10	5	5 – 25 = – 20	2	– 40
10 — 20	15	15 – 25 = – 10	3	– 30
20 — 30	25	25 – 25 = 0	4	0
30 — 40	35	35 – 25 = 10	2	20
40 — 50	45	45 – 25 = 20	1	20
			$\sum f = 12$	$\sum fD_x = -30$

$$\text{So } (\bar{x}) = a + \frac{\sum fD_x}{\sum f} = 25 + \frac{(-30)}{12} = 25 - 2.5 = 22.5$$

ii) **Geometric Mean (G.M)**

Geometric Mean is the nth positive root of the product of n values.

a) **Geometric Mean from ungrouped data**

If $x_1, x_2, x_3, \ldots, x_n$ are n values in a data, then the Geometric Mean is given as

$$G.M = (x_1 \times x_2 \times x_3 \times \ldots \times x_n)^{\frac{1}{n}}$$

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

Exmaple 15

Find the Geometric Mean (G.M) of the marks obtained by the 9th class students.
60, 65, 70, 80, 85, 90, 75.

Solution

x	**Log** x
60	log 60 = 1.7781
65	log 65 = 1.8129
70	log 70 = 1.8450
75	log 75 = 1.8750
80	log 80 = 1.9030
85	log 85 = 1.9294
90	log 90 = 1.9542
	$\sum \log x = 13.0976$

$$\text{G.M} = \text{Anti-log}\ \frac{\sum \log x}{n}$$

$$= \text{Anti-log}\ \frac{13.0976}{7}$$

$$= \text{Anti-log}\ 1.8710$$

$$= 74.31$$

b) Geometric Mean from Grouped data

Let $x_1, x_2, x_3, \ldots, x_n$ be the mid-points in a frequency distribution and $f_1, f_2, f_3, \ldots f_n$ are their respective frequencies. Then the G.M is calculated by the following formula:

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

Example 16

Calculate the Geometric Mean (G.M) of the following data.

Marks	0 — 20	20 — 40	40 — 60	60 — 80
Number of students	3	4	10	11

Solution

Marks	x	f	$\log x$	$f \log x$
0 — 20	10	3	1.0000	3.0000
20 — 40	30	4	1.4771	5.9084
40 — 60	50	10	1.6989	16.989
60 — 80	70	11	1.8450	20.295
		$\sum f = 28$		$\sum f \log x = 46.1924$

By Formula

$$G.M = \text{Anti-log}\ \frac{\sum f \log x}{\sum f}$$

$$= \text{Anti-log}\ \frac{46.1924}{28}$$

$$= \text{Anti-log}\ 1.6497$$

$$= 44.64$$

iii) Harmonic Mean (H.M)

It is the reciprocal of the Arithmetic Mean of the reciprocal values.

a) Harmonic Mean from ungrouped data

Let there be 'n' values i.e, $x_1, x_2, x_3, \ldots, x_n$ in a data, then the Harmonic Mean (H.M) is calculated by the following formula:

$$H.M = \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+,...+\frac{1}{x_n}} = \frac{n}{\sum\left(\frac{1}{x}\right)}$$

Example 17

Find the Harmonic Mean of the following values. 5, 6, 8, 9 and 10.

Solution

x	$\log \frac{1}{x}$
5	$\frac{1}{5}=0.2$
6	$\frac{1}{6}=0.16$
8	$\frac{1}{8}=0.125$
9	$\frac{1}{9}=0.11$
10	$\frac{1}{10}=0.1$
	$\sum\frac{1}{x}=0.695$

By using formula:

$$H.M = \frac{n}{\sum\left(\frac{1}{x}\right)} = \frac{5}{0.695} = 7.194$$

b) Harmonic Mean from grouped data

Harmonic Mean from grouped data is computed by the following formula:

$$H.M = \frac{\sum f}{\sum\left(\frac{f}{x}\right)}$$

Example 18

Find the Harmonic Mean from the following data.

Classes	0 — 6	6 — 12	12 — 18	18 — 24	24 — 30
Frequency	1	2	5	4	6

Solution

Classes	Frequency (f)	x (Mid-point)	$\frac{f}{x}$
0 — 6	1	3	$\frac{1}{3} = 0.33$
6 — 12	2	9	$\frac{2}{9} = 0.22$
12 — 18	5	15	$\frac{5}{15} = 0.33$
18 — 24	4	21	$\frac{4}{21} = 0.19$
24 — 30	6	27	$\frac{6}{27} = 0.22$
	$\sum f = 18$		$\sum\left(\frac{f}{x}\right) = 1.29$

$$\text{H.M} = \frac{\sum f}{\sum\left(\frac{f}{x}\right)} = \frac{18}{1.29} = 13.95$$

2. Averages of positions

i) Median

If the values (items) of a series arranged in ascending or descending order of magnitude, then the middle term of the series is called its median.

Median is calculated for the following datas:

a) ungrouped data b) discrete data c) continuous data

a) **Median for ungrouped data**

If in a data there are n values, then after arranging the values in ascending or descending order Median is calculated by the following rules:

$$\text{Median} = \text{Size of } \left(\frac{n+1}{2}\right)\text{th item} \qquad (\text{if } n \text{ is odd})$$

or

$$\text{Median} = \frac{1}{2}\left[\text{Size of } \frac{n}{2}\text{th} + \text{Size of } \left(\frac{n+2}{2}\right)\text{th item}\right] \quad (\text{if } n \text{ is even})$$

Example 19

Find the median for the following values.

2, 4, 5, 6, 3

Solution Writing the given data in increasing order

2, 3, 4, 5, 6

S.No	x (value)
1	2
2	3
3	4
4	5
5	6

By using formula

$$\text{Median} = \text{size of } \left(\frac{n+1}{2}\right)\text{th item}$$

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

$$= \text{size of 3rd item}$$

$$= 4$$

So Median $= 4$

Example 20

The following are the daily pocket money in rupees for the children of a family 10, 20, 15, 30. Calculate the median for the data.

Solution

S.No	x
1	10
2	15
3	20
4	30

Since number of items is even, so

$$\text{Median} = \frac{1}{2}\ [\text{Size of } \frac{n}{2}\text{th} + \text{Size of } (\frac{n+2}{2})\text{th}]\text{item}$$

$$= \frac{1}{2}\ [\text{Size of } \frac{4}{2}\text{th} + \text{Size of } (\frac{6}{2})\text{th}]\ \text{item}$$

$$= \frac{1}{2}[\text{size of 2nd} + \text{size of 3rd}]\ \text{item}$$

$$= \frac{1}{2}(15 + 20)$$

$$= \frac{35}{2} = 17.5$$

b) Median for discrete data

In a frequency distribution if $\sum f$ is odd then the value of x opposite to $\left(\frac{\sum f + 1}{2}\right)$th item in the cumulative frequency column is its median. But if $\sum f$ is even, then the value of x opposite to $\frac{\sum f}{2}$ th item in the cumulative frequency column is its median. This can be illustrated with the help of the following examples.

Example 21

The following are the marks obtained by 35 students in a test.

x	10	12	15	20	25	30
f	1	10	5	13	2	4

Solution

x	f	$c.f$
10	1	1
12	10	11
15	5	16
20	13	29
25	2	31
30	4	35

Since $n = \sum f = 35$

So Median = Size of $\left(\frac{n+1}{2}\right)$th item

= Size of $\left(\frac{35+1}{2}\right)$th item

= Size of 18th item

= 20

Example 22

Find the median marks from the following distribution:

(Marks) x	10	20	22	25
Number of students	0	2	4	6

Solution

Marks x	*Frequency*	*c.f*
10	0	0
20	2	2
22	4	6
25	6	12
	$\sum f = 12$	

$$\text{Median} = \text{Size of } \left(\frac{n}{2}\right)\text{th item}$$

$$= \text{Size of } \left(\frac{12}{2}\right)\text{th item}$$

$$= \text{Size of 6th item}$$

$$= 22$$

So, Median = 22

c) **Median from continuous data**

For computing median in continuous data, it is important to make class boundaries and then median is calculated by the following formula:

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

l = lower limit of the median class

h = width (class-interval of the median class)

f = frequency of the median class

$$\frac{n}{2} = \frac{\sum f}{2}$$

c = cumulative frequency of the class preceding the median class.

Example 23

Find the median of the following distribution.

Daily wages (in Rs.)	60—69	70—79	80—89	90—99	100—109
Labour	4	6	8	10	5

Solution

Daily wages	f	**(C.B) Class-Boundaries**	***(C.F) Cumulative Frequency***
60 — 69	4	59.5 — 69.5	4
70 — 79	6	69.5 — 79.5	10
80 — 89	8	79.5 — 89.5	18
90 — 99	10	89.5 — 99.5	28
100 — 109	5	99.5 — 109.5	33
	$\sum f = 33$		

$$\text{Median} = \frac{n}{2}\text{th item} = \frac{33}{2}\text{th item}$$

$$= 16.5\text{ th item}$$

Median lies in the group 79.5 — 89.5

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

$$= 79.5 + \frac{10}{8}(16.5 - 10)$$

$$= 79.5 + \frac{10}{8}(6.5)$$

$$= 79.5 + 8.125$$

$$= 87.625$$

(ii) Mode

The value that appears more times in a data, is called **mode** for the given data or the most frequent value in a data is called **mode**.

Mode is also calculated for the following datas:

a) Ungrouped data b) Discrete data

c) Continuous data

Consider the following examples.

Example 24

From the following sizes of kids trousers, find the model size.

25, 30, 31, 25, 35, 25

Solution

In the above data, 25 is the most frequent value, so the model size is 25.

Example 25

The following data shows the weights of the students. Find the model weight.

Weights (in KG)	40	42	50	51	55
Number of students	10	8	3	2	1

NOT FOR SALE

Solution

Mode = 40, because 40 is the most frequent value in the data.

Example 26

Calculate the mode from the following frequency distribution.

Marks (C.B)	0 — 4	4 — 8	8 — 12	12 — 16	16 — 20
No. of students	3	5	4	6	2

Solution

Marks (C.B)	*No. of students f*
0 — 4	3
4 — 8	5
8 — 12	4
12 — 16	6
16 — 20	2

→ Modal group

The mode lies in the group 12 — 16.
So by using the formula

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

Where

l = lower limit of the modal group

f_m = Frequency of the modal group

f_0 = Frequency of the group preceding the modal group

f_1 = Frequency of the group following the modal group

h = class-interval

So, Mode $= 12 + \left(\frac{6-4}{2(6)-4-2}\right) \times 4$

$$= 12 + \frac{2}{12-6} \times 4$$

$$= 12 + \frac{8}{6}$$

$$= 12 + 1.33 = 13.33$$

(iii) Quartiles

Quartiles are the values which divide an arranged series of data into four equal parts. There are three quartiles Q_1, Q_2, Q_3. Q_1 is called First quartile or lower quartile, Q_3 is called third quartile or upper quartile, while Q_2 is called second quartile or median.

If in a data (ungrouped discrete) there are n values then:

$$Q_1 = \text{Size of } \left(\frac{n+1}{4}\right)\text{th item}$$

$$Q_2 = \text{Size of } 2\left(\frac{n+1}{4}\right)\text{th item}$$

$$Q_3 = \text{Size of } 3\left(\frac{n+1}{4}\right)\text{th item}$$

If n is odd

While

$$Q_1 = \frac{1}{2}\left[\frac{n}{4}\text{th item} + \left(\frac{n}{4}+1\right)\text{ th item}\right]$$

$$Q_2 = \frac{1}{2}\left[\frac{n}{2}\text{th item} + \left(\frac{n}{2}+1\right)\text{ th item}\right]$$

$$Q_3 = \frac{1}{2}\left[\frac{3n}{4}\text{th item} + \left(\frac{3n}{4}+1\right)\text{ th item}\right]$$

If n is even

In continuous data, Q_1 and Q_3 are given as

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

Where l = lower limit of the goup

h = class size or interval

f = frequency of the group in which Q_1 or Q_3 lies.

c = the cumulative frequency of the group above the group in which Q_1 or Q_3 lies.

$\frac{n}{4}$ = location of Q_1 (calculated) and $\frac{3n}{4}$ = location of Q_3.

Example 27

Find Q_1 and Q_3 for the following data.
10, 12, 16, 20, 18, 22, 24

Solution

S. No	x
1	10
2	12
3	16
4	18
5	20
6	22
7	24

Since $n = 7$ which is odd so

Q_1 = size of $(\frac{n+1}{4})$th item

$= \text{size of } (\frac{7+1}{4})\text{th item}$

$= \text{size of } \frac{8}{4} \text{ th item}$

$= \text{size of 2nd item}$

$= 12$

$Q_3 = \text{size of } 3\,(\frac{n+1}{4})\text{th item}$

$= \text{size of } 3\,(\frac{7+1}{4})\text{th item}$

$= \text{size of } 3(\frac{8}{4})\text{th item}$

$= \text{size of 6th item}$

$= 22$

Example 28

Find Q_1 and Q_3 for the following data.
10, 12, 16, 20, 18, 22.

Solution

S. No	*x*
1	10
2	12
3	16
4	20
5	18
6	22
7	24
8	25

Here $n = 8$ which is even. So

$$Q_1 = \frac{1}{2}\left[\frac{n}{4}\text{th item} + \left(\frac{n}{4}+1\right)\text{th item}\right]$$

$$= \frac{1}{2}\left[\frac{8}{4}\text{th item} + \left(\frac{8}{4}+1\right)\text{th item}\right]$$

$$= \frac{1}{2}\left[\text{2nd item} + \text{3rd item}\right]$$

$$= \frac{1}{2}(12 + 16) = \frac{28}{2} = 14$$

$$Q_3 = \frac{1}{2}\left[\frac{3n}{4}\text{th item} + \left(\frac{3n}{4}+1\right)\text{th item}\right]$$

$$= \frac{1}{2}(\text{6th item} + \text{7th item})$$

$$= \frac{1}{2}(22 + 24) = \frac{46}{2} = 23$$

Example 29

Find Q_1 and Q_3 from the following observations.

Classes	13 — 17	18 — 22	23 — 27	28 — 32	33 — 37	38 — 42
Frequency	1	1	2	3	2	1

Solution

Classes	Class-Boundaries	*frequencey*	*Cumulative Frequency*
13 — 17	12.5 — 17.5	1	1
18 — 22	17.5 — 22.5	1	2
23 — 27	22.5 — 27.5	2	4
28 — 32	27.5 — 32.5	3	7
33 — 37	32.5 — 37.5	2	9
38 — 42	37.5 — 42.5	1	10

Location of $Q_1 = \frac{n}{4} th$ item

$$= \frac{10}{4} th \text{ item}$$

$$= 2.50$$

Q_1 lies in the group 22.5 — 27.5. So

$$Q_1 = l + \frac{h}{f}\left(\frac{n}{4} - c\right)$$

$$= 22.5 + \frac{5}{2}(2.50 - 2)$$

$$= 22.5 + \frac{5}{2}(0.50)$$

$$= 22.5 + 1.25 = 23.75$$

Location of $Q_3 = \frac{3n}{4} th$ item

$$= \frac{30}{4} th \text{ item}$$

$$= 7.5$$

Q_3 lies in the group 32.5 — 37.5. So

$$Q_3 = l + \frac{h}{f}\left(\frac{3n}{4} - c\right)$$

$$= 32.5 + \frac{5}{3}(7.5 - 7)$$

$$= 32.5 + \frac{5}{3}(0.5)$$

$$= 32.5 + 0.8 = 33.3$$

10.3 Estimation of Median, Mode and Quartiles graphically

i) **For Median:**

For median the class-boundaries are plotted on the horizontal axis and their cumulative frequencies on the vertical axis. Then points are located above the boundary points against their respective cumulative frequencies. All the points are then joined by straight line to get a curve called 'Ogive'.

Horizontal and vertical lines are drawn from the point where median is located, so the value on the horizontal axis at which the vertical lines intersect the ogive, determines the median.

Example 30

Find median graphically from the following frequency distribution:

Classes	10 — 14	15 — 19	20 — 24	25 — 29	30 — 34	35 — 39
Frequency	1	5	7	2	6	9

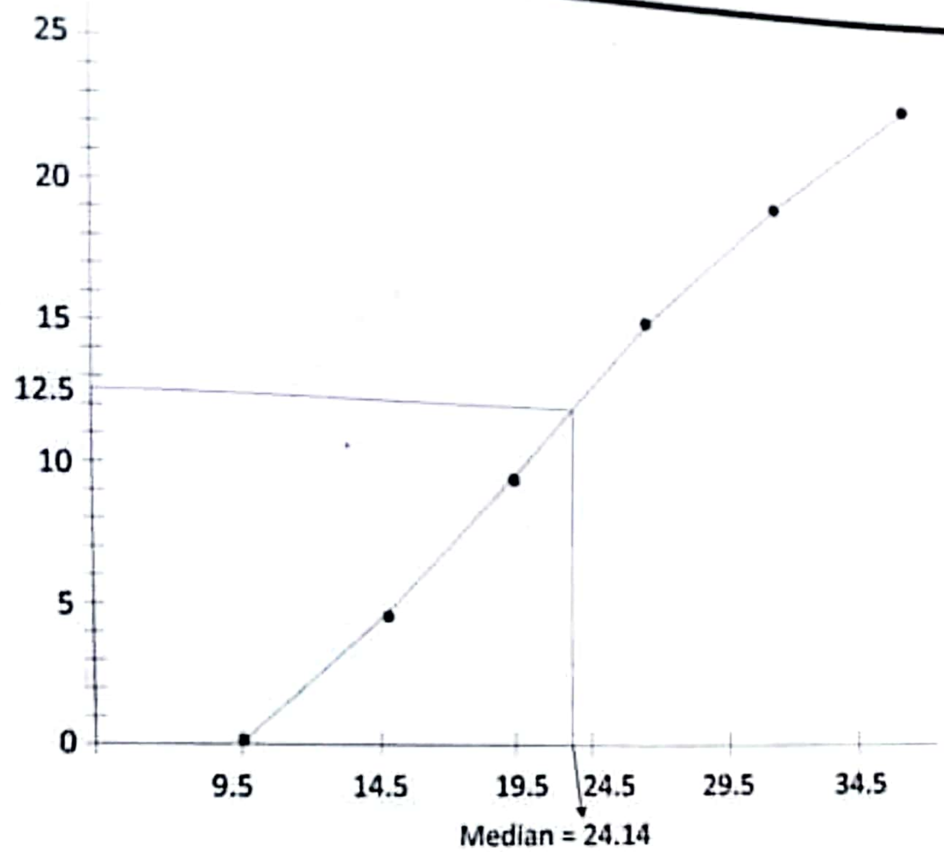
Solution

Classes	***f***	***Cumulative Frequency c.f***	**Class-Boundaries**
10 — 14	1	1	9.5 — 14.5
15 — 19	5	6	14.5 — 19.5
20 — 24	7	13	19.5 — 24.5
25 — 29	2	15	24.5 — 29.5
30 — 34	6	21	29.5 — 34.5
35 — 39	4	25	34.5 — 39.5

Median $= \frac{n}{2}th$ item

$= \frac{25}{2}th$ item

$= 12.5th$ item



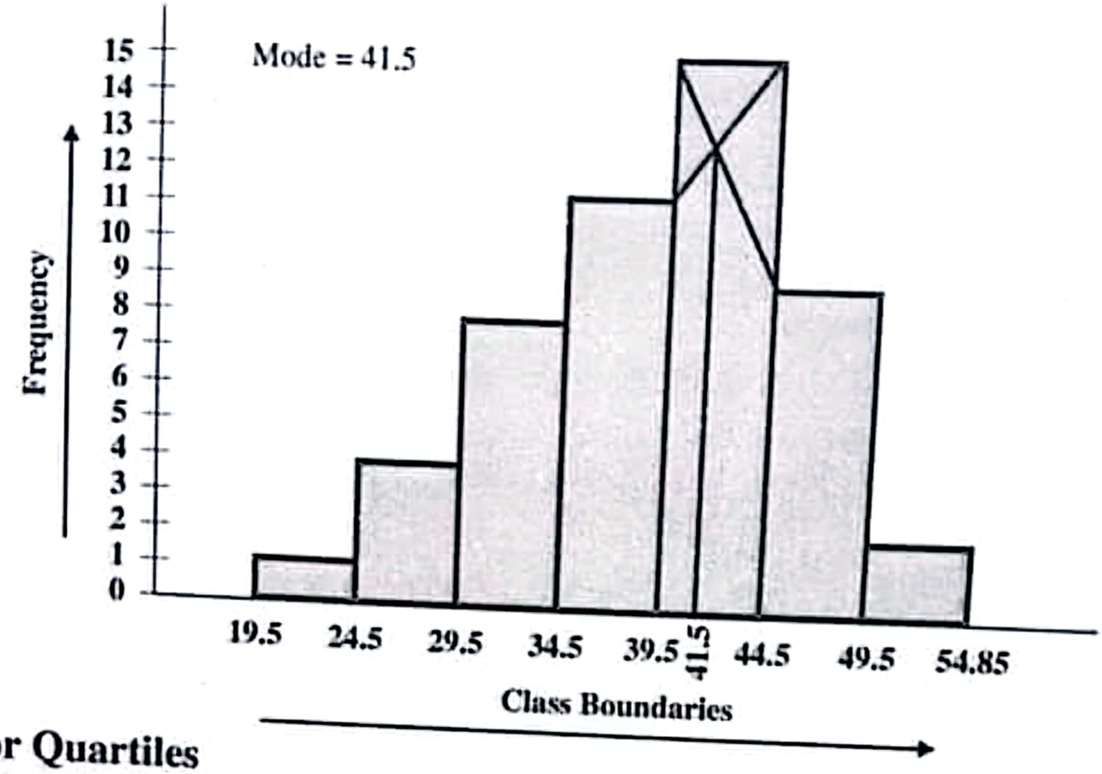
ii) **For mode**
We need histogram to estimate mode graphically which is explained with the help of the following example.

Example 31
Find mode graphically from the following frequency distribution.

Classes	20 — 24	25 — 29	30 — 34	35 — 39	40 — 44	45 — 49	50—54
Frequency	1	4	8	11	15	9	2

Solution

Classes	f	Class-Boundaries
20 — 24	1	19.5 — 24.5
25 — 29	4	24.5 — 29.5
30 — 34	8	29.5 — 34.5
35 — 39	11	34.5 — 39.5
40 — 44	15	39.5 — 44.5
45 — 49	9	44.5 — 49.5
50 – 54	2	49.5 — 54.5



iii) **For Quartiles**

To find the values of Q_1 and Q_3 from graph the cumulative frequency polygon (ogive) is drawn.

Example 32

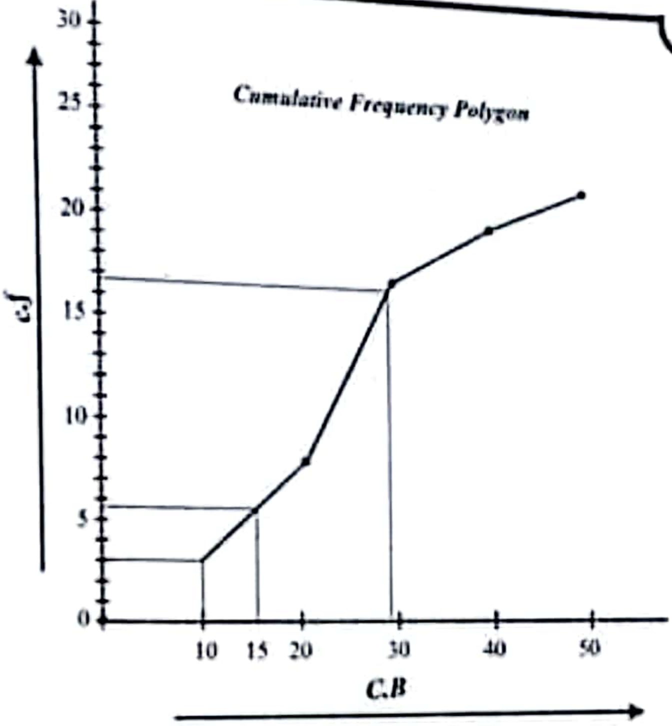
Find Q_1 and Q_3 from the following distribution.

Marks	0 — 10	10 — 20	20 — 30	30 — 40	40 — 50
Frequency	3	5	9	3	2

Solution

Marks	f	$c.f$
0 — 10	3	3
10 — 20	5	8
20 — 30	9	17
30 — 40	3	20
40 — 50	2	22

$$\text{Location of } Q_1 = \frac{n}{4}\text{th item}$$
$$= \frac{22}{4}\text{th item}$$
$$= 5.5\text{th item}$$
$$\text{Location of } Q_3 = \frac{3n}{4}\text{th item}$$
$$= 3\left(\frac{22}{4}\right)\text{th item}$$
$$= 16.5\text{th item}$$



$Q_1 = 15,\ Q_3 = 29.4$

10.4 Properties of Arithmetic Mean

i) The sum of deviations of values measured from their A.M is always equal to zero. i.e $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

ii) The sum of squared deviations of values measured from their A.M is least (minimum).

$\sum_{i=1}^{n}(x_i - \bar{x})^2 \leq \sum_{i=1}^{n}(x_i - a)^2$, where $\bar{x}$ = A.M and a = any constant

iii) If $y_i = ax_i + b$ where a and b are some non-zero constants, than $\bar{y} = a\bar{x} + b$.

iv) If K-subgroups of a data having their respective means $\bar{x}_1, \bar{x}_2 ... \bar{x}_k$, with respective frequencies $n_1, n_2, n_3, \ldots n_k$ then the mean of the combined data is given as:

$$\bar{x}_c = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \ldots + n_k\bar{x}_k}{n_1 + n_2 + n_3 + \ldots + n_k} = \frac{\sum_{i=1}^{n} n_i\bar{x}_i}{\sum_{i=1}^{n} n_i}$$

NOT FOR SALE

10.5 Weighted Average

The numerical values which shows the relative importance of different items is called weights and the average of different items having different weights is called weighted mean. Let $x_1, x_2, x_3, \ldots x_n$ are different values of items having weights $w_1, w_2, w_3 \ldots w_n$ then weighted mean:

$$\bar{x}_w = \frac{x_1w_1 + x_2w_2 + \ldots + x_nw_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum x_iw_i}{\sum w_i}$$

Example 33

The marks obtained by a student in Maths, English, Urdu and Statistics were 70, 60, 80, 65 respectively. Find the average if weights of 2, 1, 3, 1 are assigned to the marks.

x	w	wx
70	2	140
60	1	60
80	3	240
65	1	65
	$\sum w = 7$	505

Now $\bar{x}_w = \frac{\sum w_x}{\sum w} = \frac{505}{7} = 72.14$

Exercise 10.3

1. The following are the marks scored by students of the 9th grade.
   45, 30, 25, 36, 42, 27, 31, 43, 49 and 50.
   Calculate their average score.
2. Find the mean score of the marks given in question 1 by using short-cut method.

3. In a group of 20 boys, one boy has Rs.10, four boys have Rs.20, each ten boys have Rs. 30 each, four boys have Rs. 40 each and one boy has Rs.50.
   i) Construct a table for the given data.
   ii) Calculate the Average amount per student in the group both by Direct method and short-cut method.

4. Calculate the Arithmetic mean by direct method from the following data:

Classes	0 — 10	10 — 20	20 — 30	30 — 40	40 — 50
Frequency	1	4	2	3	5

5. Find the median of the following values:

   3, 4, 6, 8, 11

6. The following are the ages (in years) of some students in a class.

   10, 12, 13, 15, 16, 14.

   Calculate the median of their ages.

7. The distribution given below gives the marks in Mathematics test of 30 students. Find the median of marks.

Marks	1	4	5	7	9	10
No. of students	5	7	10	3	3	2

8. From the following distribution

Daily wages (in Rs)	112—116	117—121	122—126	127—131	132—136
No. of workers	3	20	11	4	5

   i) construct a table.
   ii) find the class-boundary for each group.
   iii) calculate the Median wages.

9. From Q.8, find the following
   i) Mode
   ii) Harmonic Mean
   iii) Geometric Mean

10. Find Median, Q_1, Q_3 and Mode from the following distribution graphically.

Classes	10 — 14	15 — 19	20 — 24	25 — 29	30 —34
Frequency	1	3	7	12	2

10.4 Measure of Dispersion

Dispersion is the scatterdness of values from its central value (Average).

Types of measure of dispersion are:
i) Range
ii) Standard deviation (S.D)
iii) Variance

Formulae of measures of dispersion:

i) Range = Largest value – smallest value

ii) $S.D = \sqrt{\dfrac{\sum f(x-\bar{x})^2}{\sum f}}$

iii) $\text{Variance} = \dfrac{\sum f(x-\bar{x})^2}{\sum f}$

i) Finding Range

Example 34
Find out the range of an employee's monthly earning for six months.
Monthly earning (in thousands) : 4, 6, 10, 9, 10, 12.

Solution
In the given data the largest value is 10 and the smallest value is 4.
i.e $L = 10$ and $S = 4$
So Range $= L - S = 10 - 4 = 6$

Example 35

Calculate the range from the given data.

Classes	5 — 9	10 — 14	15 — 19	20 — 24	25 — 29
Frequency	10	15	12	21	3

Solution

Classes	f	Class-Boundaries
5 — 9	10	4.5 — 9.5
10 — 14	15	9.5 — 14.5
15 — 19	12	14.5 — 19.5
20 — 24	21	19.5 — 24.5
25 — 29	3	24.5 — 29.5

In the given data the lower limit of the first group is 4.5 and the upper limit of the last group is 29.5. So,
Range = 29.5 – 4.5 = 25

ii) Standard Deviation (S.D)

It is the positive square root of the average of squared deviations measured from A.M (arithmetic mean).

i.e $\quad \text{S.D} = \sqrt{\dfrac{\sum (x-\bar{x})^2}{n}}$ for ungrouped data

and $\quad \text{S.D} = \sqrt{\dfrac{\sum f(x-\bar{x})}{\sum f}}$ for frequency distribution

iii) Variance

Variance is the square of standard deviation. Variance is usually denoted by the symbol 'S'. Mathematically it is expressed as:

$$S^2 = \frac{\sum (x-\bar{x})^2}{n} \text{ or } \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2$$ for ungrouped data

and $S^2 = \frac{\sum f(x-\bar{x})^2}{\sum f}$ or $\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$ for discrete and continuous data

Example 36

Find the variance and standard deviation for the following data:

6, 8, 10, 12, 14

Solution

$$\bar{x} = \frac{6+8+10+12+14}{5} = \frac{50}{5} = 10$$

x	$x-\bar{x}$	$(x-\bar{x})^2$
6	$6-10=-4$	16
8	$8-10=-2$	4
10	$10-10=0$	0
12	$12-10=2$	4
14	$14-10=4$	16
		40

$$\text{S.D} = \sqrt{\frac{\sum(x-\bar{x})^2}{n}} = \sqrt{\frac{40}{5}} = \sqrt{8}$$

Variance $(S^2) = (\sqrt{8})^2 = 8$

Example 37

The following is the distribution for the number of rotten eggs found in 60 crates. Find the standard deviation and variance of the rotten eggs.

No. of rotten eggs	0 — 4	4 — 8	8 — 12	12 — 16	16 — 20	20 — 24
No. of crates	5	10	15	20	6	4

Solution

C.B	f	x	fx	x^2	fx^2
0 — 4	5	2	10	4	20
4 — 8	10	6	60	36	360
8 — 12	15	10	150	100	1500
12 — 16	20	14	280	196	3920
16 — 20	6	18	108	324	1944
20 — 24	4	22	88	484	1936
			696		9680

$$\text{Variance} = \frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2$$

$$= \frac{9680}{60} - \left(\frac{696}{60}\right)^2$$

$$= 161.33 - 134.56 = 26.77$$

$$\text{S. D} = \sqrt{Variance} = \sqrt{26.77} = 5.18 \text{ (Approximately)}$$

Exercise 10.4

1. Find the range for the following items:
   11, 13, 15, 21, 19, 23.

2. Calculate the Range, Variance and Standard Deviation for the following discrete data:

x	5	10	11	13	15
f	2	3	4	1	5

3. The following is the distribution for the number of defective bulbs in 30 cottons (Packs). Find Variance and Standard Deviation of defective bulbs.

No. of defective bulbs	0 — 2	2 — 4	4 — 6	6 — 8	8 — 10
No. of packs (f)	1	3	15	10	2

Exercise (Objective Type) (10.5)

1. **True / False Questions.**
   Encircle 'T' for True statement and 'F' for False statement:

i)	Lower limit of the class 10 — 20 is 10.	T	F
ii)	The class-boundary of 3 – 5 is 2.05 — 5.05	T	F
iii)	The Mid-point of 0 — 4 is 3.	T	F
iv)	For Histogram we need the class-limits	T	F
v)	The widths of the rectangles in the Histogram for unequal intervals are equal.	T	F

2. **Fill in the blanks:**

i) The Geometric Mean of 2 and 8 is ________

ii) For cumulative frequency polygon ________ frequency is required.

iii) If n is odd then $\left[\frac{n+1}{2}\right]th$ item is called ________.

iv) There are ________ types of frequency distribution.

v) Frequency polygon is obtained by connecting ________.

3. **Multiple choice questions.**

For each question, four suggested answer are given.

Choose the correct one and rewrite as a, b, c and d in the box.

i) The difference of the largest and smallest value in the data is called

(a) Mean (b) Mode

(c) Range (d) Standard deviation ☐

ii) The formula $\frac{\sum x}{n}$ determine

(a) Arithmetic Mean (b) Median ☐

(c) Mode (d) G.M

iii) To find the class-boundaries for 2 — 3, 4 — 5, then ________ is subtracted from lower limit and added to the upper limit.

(a) 0.5 (b) 0.05 ☐

(c) 0.005 (d) 5

iv) $\frac{\sum f(x-\bar{x})^2}{\sum f}$ is called ________

(a) Range (b) Median ☐

(c) S.D (d) Variance

v) The most frequent value in the data is called its ________

(a) Mean (b) Median ☐

(c) Mode (d) G.M

NOT FOR SALE