Unit 06

BASIC STATISTICS

Frequency Distribution

A frequency distribution is a tabular arrangement for classifying data into different groups and the number of observation falling in each group corresponds to the respective group. In fact a frequency distribution is a method to summarize data.

Grouped Data:

The data presented in the form of frequency distribution is called grouped data.

Types of Frequency Distribution

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

(a) Discrete Frequency Distribution

(b) Continuous Frequency Distribution

(a) Discrete Frequency Table

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

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

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

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

Example 1: Five coins are tossed 20 times and the number of head recorded at each toss are give below:

3, 4, 2, 3, 3, 5, 2, 2, 2, 1, 1, 2, 1, 4, 2, 2, 3,3,4,2.

Make frequency distribution of the number of heads observed.

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

Frequency Distribution of number of heads		
X	**Tally Marks**	**Frequency**
1	|||	3
2	𝍸 |||	8
3	𝍸 |	5
4	|||	3
5	|	1
Total		**20**

(b) Continuous frequency distribution

The making of continuous frequency distribution involves the following steps:

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

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

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

$$h = \frac{\text{Range}}{k}$$

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

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

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

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

Example 2: The following are the marks obtained by 40 students in mathematics of class X. Make a frequency distribution with a class interval of size 10.

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

Solution: Let X = Marks of a student.

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

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

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

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

Classes/Groups	Tally Marks	Frequency
0–9	II	2
10–19	𝍸 𝍸	10
20–29	𝍸	5
30–39	𝍸 IIII	9
40–49	𝍸 I	6
50–59	𝍸 II	7
60–69	I	1
Total		**40**

Note: The Tally bar method is usually adopted to construct a frequency distribution.

Concepts involved in a Continuous frequency table.

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

(a) Class Limits: The minimum and the maximum values defined for a class or group are called class limits. The minimum value is called the lower class limit and the maximum value is called the upper class limit of that class. For example in the group (5–10), 5 is lower class limit and 10 is called upper class limit.

(b) Class Boundaries:

The real class limits of a class are called class boundaries. A class boundary is obtained by adding two successive class limits and dividing the sum by 2. The value so obtained is taken as upper class boundary for the previous class and lower class boundary for the next class.

(c) Midpoint or Class Mark:

For a given class the average of that class obtained by dividing the sum of upper and lower class limits by 2, is called the midpoint or class mark of that class.

For example the midpoint of a class (6–10) is 8. i.e.

$$\text{Class Marks } x = \frac{6+10}{2} = \frac{16}{2} = 8$$

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

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

Solution:

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

Construction of Histograms:

Histogram

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

Equal Intervals Histogram:

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

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

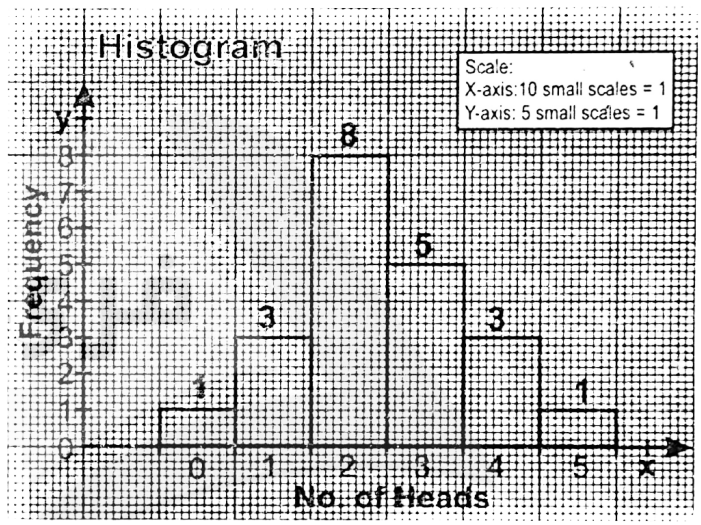
Solution: We proceed as follows:

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

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

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

The resulting Histogram is given below:



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

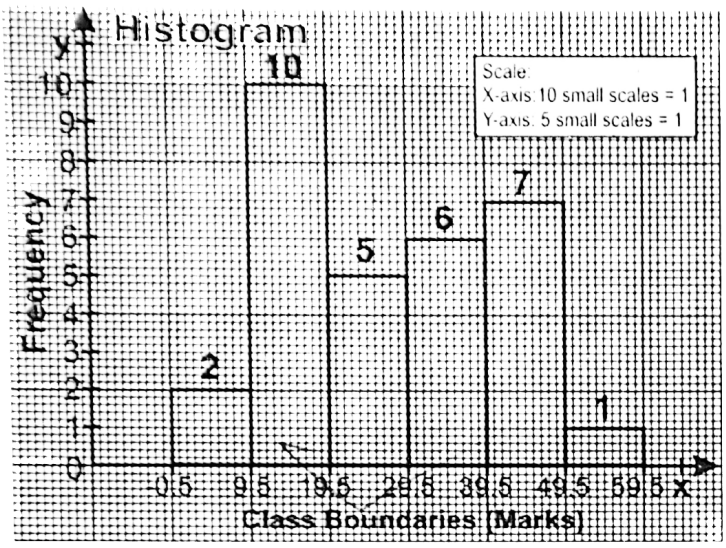
Class Boundaries	Frequency
–0.5–9.5	2
9.5–19–5	10
19.5–29.5	5
29.5–39.5	6
39.5–49.5	7
49.5–59.5	1

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

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

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

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



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

Unequal Intervals Histogram:

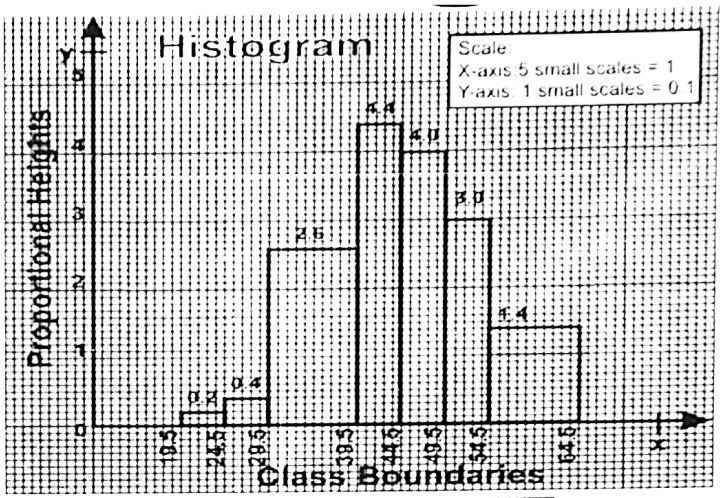
If the class intervals are un–equal, the frequency must be adjusted by dividing each class frequency on its class interval size. If the interval becomes double, then frequency is divided by 2, so that the area of the bar is in proportion to the areas of other bars etc.

Example 3: Draw a histogram illustrating the following data:

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

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

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



Construction of Frequency Polygon

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

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

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

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

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

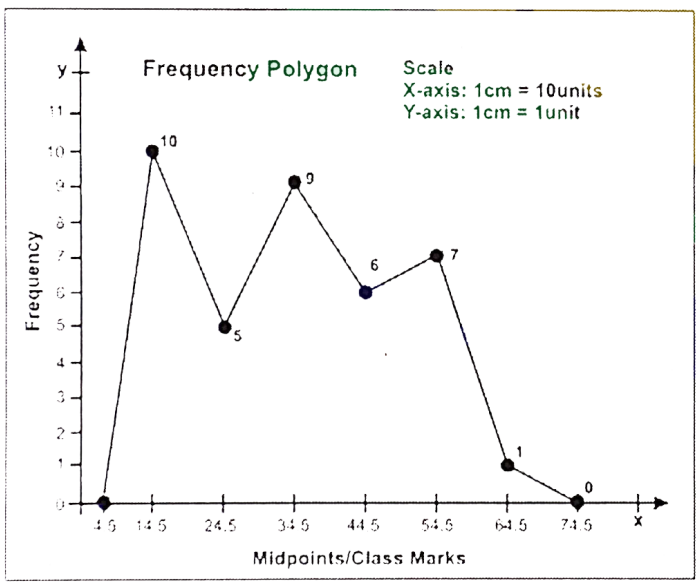
Step 2: Calculate class marks or midpoints for the given distribution.

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

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

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

Step 5: Join all the points by means of line segment.



Cumulative Frequency Distribution:

(i) Construction of Cumulative Frequency Table

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

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

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

Solution: The cumulative frequency distribution is constructed below.

Class Boundaries	f	CF	Class Boundaries	CF
14.5–19.5	0	0	Less than 19.5	0
19.5–24.5	1	0+1=1	Less than 24.5	1
24.5–29.5	2	1+2=3	Less than 29.5	3
29.5–34.5	26	3+26=29	Less than 34.5	29
34.5–39.5	22	29+22=51	Less than 39.5	51
39.5–44.5	20	51+20=71	Less than 44.5	71
44.5–49.5	15	71+15=86	Less than 49.5	86
49.5–54.5	14	86+14=100	Less than 54.5	100

(ii) Drawing of cumulative Frequency Polygon or Ogive

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

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

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

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

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

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

Class Limits	Frequency
4–6	2
7–9	4
10–12	8
13–15	3

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

Class Limits	Class Boundaries	F	CF
1–3	**0.5–3.5**	**0**	**0**
4–6	3.5–6.5	2	0+2=2
7–9	6.5–9.5	4	2+4=6
10–12	9.5–12.5	8	6+8=14
13–15	12.5–15.5	3	14+3=17

Classes	Tally Marks	f	Mid points (x)	Class Boundaries
20–24	𝍸 I	6	22	19.5–24.5
25–29	𝍸 𝍸	10	27	24.5–29.5
30–34	𝍸 𝍸 II	12	32	29.5–34.5
35–39	𝍸 IIII	9	34	34.5–39.5
40–44	III	3	39	39.5–44.5
	$\sum f = 40$			

Less than Cumulative Frequency Distribution

C.B	f	C.F	Class Boundaries	F
14.5–19.5	0	0	Less than 19.5	0
19.5–24.5	6	0+6=6	Less than 24.5	6
24.5–29.5	10	6+10=16	Less than 29.5	16
29.5–34.5	12	16+12=28	Less than 34.5	28
34.5–39.5	9	28+9=37	Less than 39.5	37
39.5–44.5	3	37+3=40	Less than 44.5	40

Q. 3 From the following data representing the salaries of 30 teachers of a school. Make a frequency distribution taking class interval size of Rs. 100, 450, 500, 550, 580, 670, 1200, 1150, 1120, 950, 1130, 1230, 890, 780, 760, 670, 880, 890, 1050, 980, 970, 1020, 1130, 1220, 760, 690, 710, 750, 1120, 760, 1240 06(003)

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

Solution: Let Salaries of teachers are X.

X: 450, 500, 550, 580, 670, 1200, 1150, 1120, 950, 1130, 1230, 890, 780, 760, 670, 880, 890, 1050, 980, 970, 1020, 1130, 1220, 760, 690, 710, 750, 1120, 760, 1240

Minimum value = X_{min} = 450

Maximum value = X_{max} = 1240

Number of observations = n = 30

Size of class intervals = Rs.100

Classes	Tally Marks	Frequency (f)
450–549	II	2
550–649	II	2
650–749	IIII	4
750–849	𝍸	5
850–949	III	3
950–1049	IIII	4
1050–1149	𝍸	5
1150–1249	𝍸	5
		Σf = 30

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

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

(a) Find the most frequent load shedding hours

(b) Find least load shedding intervals

(Hint: Make classes 2—3,4—5,6—7,.....)

Solution:

Let load shedding hours are X.

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

Minimum value = X_{min} = 2

Maximum value = X_{max} = 14

Classes	Tally marks	Frequency (f)
2–3	II	2
4–5	I	1
6–7	𝍸 IIII	9
8–9	𝍸	5
10–11	𝍸 I	6
12–13	𝍸	5
14–15	III	3
Total		**Σf = 31**

(a) The most frequent load shedding hours are (6–7)

(b) The least load shedding interval is (4–5)

Classes	Tally Marks	f	Mid points (x)	Class Boundaries
20–24	𝍸 I	6	22	19.5–24.5
25–29	𝍸 𝍸	10	27	24.5–29.5
30–34	𝍸 𝍸 II	12	32	29.5–34.5
35–39	𝍸 IIII	9	34	34.5–39.5
40–44	III	3	39	39.5–44.5
	$\sum f = 40$			

Less than Cumulative Frequency Distribution

C.B	f	C.F	Class Boundaries	F
14.5–19.5	0	0	Less than 19.5	0
19.5–24.5	6	0+6=6	Less than 24.5	6
24.5–29.5	10	6+10=16	Less than 29.5	16
29.5–34.5	12	16+12=28	Less than 34.5	28
34.5–39.5	9	28+9=37	Less than 39.5	37
39.5–44.5	3	37+3=40	Less than 44 5	40

Q. 3 From the following data representing the salaries of 30 teachers of a school. Make a frequency distribution taking class interval size of Rs. 100, 450, 500, 550, 580, 670, 1200, 1150, 1120, 950, 1130, 1230, 890, 780, 760, 670, 880, 890, 1050, 980, 970, 1020, 1130, 1220, 760, 690, 710, 750, 1120, 760, 1240 06(003)

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

Solution: Let Salaries of teachers are X.

X: 450, 500, 550, 580, 670, 1200, 1150, 1120, 950, 1130, 1230, 890, 780, 760, 670, 880, 890, 1050, 980, 970, 1020, 1130, 1220, 760, 690, 710, 750, 1120, 760, 1240

Minimum value = X_{min} = 450

Maximum value = X_{max}= 1240

Number of observations = n = 30

Size of class intervals = Rs.100

Classes	Tally Marks	Frequency (f)
450–549	II	2
550–649	II	2
650–749	IIII	4
750–849	𝍸	5
850–949	III	3
950–1049	IIII	4
1050–1149	𝍸	5
1150–1249	𝍸	5
		Σf = 30

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

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

(a) Find the most frequent load shedding hours

(b) Find least load shedding intervals

(Hint: Make classes 2—3,4—5,6—7,.....)

Solution:

Let load shedding hours are X.

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

Minimum value = X_{min} = 2

Maximum value = X_{max}= 14

Classes	Tally marks	Frequency (f)
2–3	II	2
4–5	I	1
6–7	𝍸 IIII	9
8–9	𝍸	5
10–11	𝍸 I	6
12–13	𝍸	5
14–15	III	3
Total		**Σf = 31**

(a) The most frequent load shedding hours are (6–7)

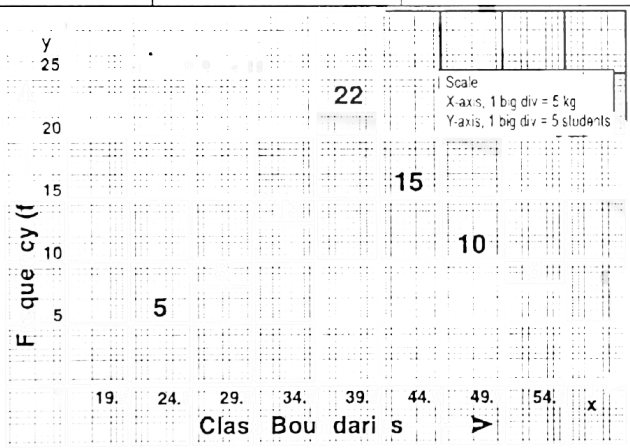
(b) The least load shedding interval is (4–5)

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

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

(i) Solution: Histogram

Weights	Frequency (f)	Class Boundaries
20–24	5	19.5–24.5
25–29	8	24.5–29.5
30–34	13	29.5–34.5
35–39	22	34.5–39.5
40–44	15	39.5–44.5
45–49	10	44.5–49.5
50–54	8	49.5–54.5

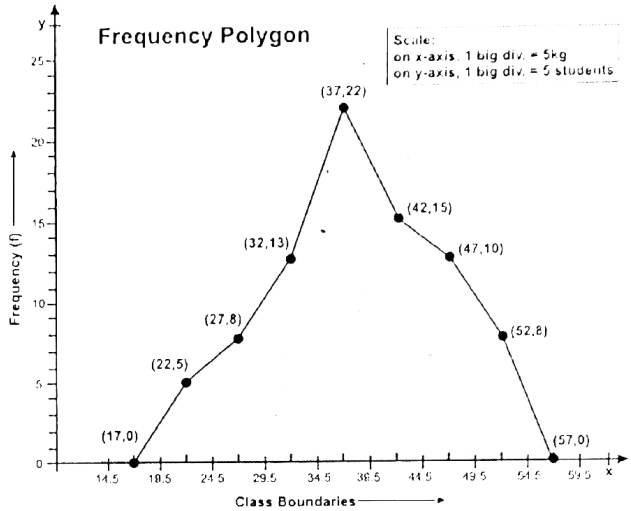


(ii) Frequency Polygon:

Weights (kg)	(f)	Class Boundaries	Midpoint
15–19	0	14.5–19.5	17
20–24	5	19.5–24.5	22
25–29	8	24.5–29.5	27
30–34	13	29.5–34.5	32
35–39	22	34.5–39.5	37
40–44	15	39.5–44.5	42
45–49	10	44.5–49.5	47
50–54	8	49.5–54.5	52
55–59	0	54.5–59.5	57

Note: Two additional groups with same size of class interval are taken. One before the very first group and the second after the very last group. These two groups will have frequency "0".

Frequency Polygon:



Average or Central Value

The representative value of the variables around which the majority of the observations tends to concentrate shows the tendency or behavior of the distribution. This representative value is called average or the central value.

Measures of Central Tendency:

The measures or techniques that are used to determine the central value are called measure of central tendency. Some measures of central tendency are discussed here.

(1) Arithmetic Mean:

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

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

Computation of Arithmetic Mean:

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

Ungrouped Data:

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

a. Direct method

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

b. Indirect method

i. Short Cut method

$$\bar{x}=A+\frac{\sum D}{n}$$

- D = X – A
- X = Observations
- A = Assumed mean from observation

ii. Coding Method

$$\bar{x}=A+\frac{\sum U}{n}\times h$$

- $U=\frac{x-A}{h}$
- X = Observations
- A = Assumed mean from observation
- h = common difference or common divisor

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

Students	1	2	3	4	5	6	7
Marks	45	60	74	58	65	63	49

Solution: Let X = marks of a students.

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

or $$\bar{x}=\frac{45+60+74+58+65+63+49}{7}$$

$$\bar{x}=\frac{414}{7}=59.14 \text{ Marks}$$

Example 2: The salaries of five teaches are as follows. Find the mean salary using direct and indirect methods and compare the results.

11500,12400,15000,14500,14800

Solution:

a. Mean by Direct Method

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

$$=\frac{74000}{5}=Rs.\ 13640$$

Mean by Indirect Method (Short Cut, Coding)

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

X	$D_i=(x_i-13000)$	$u_i=\frac{(x_i-A)}{100}$
11500	–1500	–15
12400	–600	–6
15000	2000	20
14500	1500	15
14800	1800	18
$\sum x_i = 74000$	$\sum D_i = 3200$	$\sum u_i = 32$

(i) Using Short method

$$\bar{x}=A+\frac{\sum D}{n}$$

$$\bar{x}=13000+\frac{3200}{5}=13000+640=Rs.\ 13640$$

(ii) Using Coding Method

$$\bar{x}=A+\frac{\sum U}{n}\times h$$

$$\bar{x}=13000+\frac{32}{5}\times 100$$

$$\bar{x}=13000+\frac{3200}{5}=13000+640=\text{Rs.}13640$$

Group Data:

A data in the form of frequency distribution is called grouped data. To find the arithmetic mean from the group data we define formulae under direct and indirect methods as given below.

(a) Using Direct method

$$\bar{x}=\frac{\sum fx}{\sum f}$$

(b) Using Indirect method

(i) Short Cut method

$$\bar{x}=A+\frac{\sum fD}{\sum f}$$

- D = X – A
- X = Midpoint of a class
- A = Assumed mean

(ii) **Coding method**

$$\bar{X} = A + \frac{\sum fU}{\sum f} \times h$$

$$U = \frac{X - A}{h}$$

- X = Midpoint of a class
- A = Assumed mean
- h = Size of Class Interval

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

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

Solution: we compute mean as follows.

(Number of Heads) X	Frequency (f)	fx
1	3	
2	8	16
3	5	15
4	3	12
5	1	5
Total	$\sum f = 20$	$\sum fx = 51$

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{51}{20} = 2.55 \text{ or 3 heads}$$

(since the variable is discrete)

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

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

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

Classes / Groups	Frequency (f)	Midpoints (x)	fx
0 – 9	2	4.5	9
10 – 19	10	14.5	145
20 – 29	5	24.5	122.5
30 – 39	9	34.5	310.5
40 – 49	6	44.5	267
50 – 59	7	54.5	381.5
60 – 69	1	64.5	64.5
Total	$\sum f = 40$		$\sum fx = 1300$

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

Example 5: Find arithmetic mean using indirect formulae taking x = 34.5 as the provisional mean in example 4.

Solution: We use the following formulae.

(i) $\bar{X} = A + \frac{\sum fD}{\sum f}$ (Short Cut Formula)

(ii) $\bar{X} = A + \frac{\sum fU}{\sum f} \times h$ (Coding Formula)

Given A = 34.5. We may take h = 10 and make the following calculations.

Classes / Groups	f	Mid Points	D= X–34.5	U = (X–A)/ 10	fD	fU
0–9	2	4.5	–30	–3	–60	–6
10–19	10	14.5	–20	–2	–200	–20
20–29	5	24.5	–10	–1	–50	–5
30–39	9	34.5	0	0	0	0
40–49	6	44.5	10	1	60	6
50–59	7	54.5	20	2	140	14
60–69	1	64.5	30	3	30	3
Total	**$\Sigma f = 40$**		**1300**		**$\Sigma fD = -80$**	**$\Sigma fU = -8$**

Substituting the totals in the above formulae we get.

(i) $\bar{X}=A+\frac{\sum fD}{\sum f}$ $\Rightarrow$ $\bar{X} = 34.5 + \left(\frac{-80}{40}\right) = 34.5 - 2 = 32.5$ gm

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

$\bar{X}=34.5+\left(\frac{-8}{40}\right)\times 10$ $\Rightarrow$ $\bar{X}=34.5+\left(\frac{-80}{40}\right)=34.5-2=32.5\,\text{gm}$

(2) Median:

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

Median From Ungrouped Data

Case1. When the number of observations (n) is odd.

$$\tilde{x}=\left(\frac{n+1}{2}\right)^{th} \text{ term in arranged data}$$

Case2. When the number of observations (n) is even.

$$\tilde{x}=\frac{1}{2}\left[\frac{n}{2}\text{th term} + \frac{n+2}{2}\text{th term}\right]$$

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

Solution:

Arranged Data: 79, 82, 86, 92, 93

Since number of observation is odd i.e., n =5

$\tilde{x} = \left(\frac{n+1}{2}\right)^{th}$ term in arranged data

$\tilde{x}=\left[\left(\frac{5+1}{2}\right)^{th} \text{term}\right]$ in arranged data

$\tilde{x} = \frac{6}{2} th\ term$ in arranged data

$\tilde{x} = 3^{rd}$ term in arranged data

$\tilde{x} = 86$

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

Solution:

Arranged Data: 1.9, 2.3, 2.5, 2.7, 2.9, 3.1

As number of observations are even i.e., n = 6

So, $\tilde{x}=\frac{1}{2}\left[\frac{n}{2}\text{th term} + \frac{n+2}{2}\text{th term}\right]$

$\tilde{x}=\frac{1}{2}\left(\frac{6}{2}\text{th term} + \frac{6+2}{2}\text{th term}\right)$

$\tilde{x}=\frac{1}{2}\left(3^{rd}\text{ term} + 4^{th}\text{ term}\right)$

$\tilde{x}=\frac{2.5+2.7}{2}=\frac{5.2}{2}=2.6$ milligram

Grouped Data (Discrete)

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

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

Example 3: Find median for the following frequency distribution.

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

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

X	Frequency	Cumulative frequency
1	3	3
2	8	3 + 8 = 11
3	5	11 + 5 =16
4	3	16 + 3 =19
5	1	19 + 1 = 20

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

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

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

Median = 2

Grouped Data (Continuous)

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

i. Determine class boundaries
ii. Make cumulative frequency column
iii. Determine the median class using cumulative frequency, i.e., the class containing $\left[\frac{n}{2}\right]^{th}$ observation.
iv. Use the formula $\text{Median} = l + \frac{h}{f}(\frac{n}{2} - c)$

Where

l = Lower class boundary of the median class.
h = class interval size of the median class.
f = frequency of the median class.
c = cumulative frequency of the class preceding the median class.

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

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

Solution:

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

Now

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

Median class = class containing $\left[\frac{40}{2}\right]^{th}$ observation

Median class = class containing 20^{th} observation

Median class = 144.5–153.5

Where

l = Lower class boundary of median class = 144.5
h = class interval size of the median class = 9
f = frequency of the median class = 12
c = cumulative frequency of the class preceding the median class =17

So, $\tilde{x} = l + \frac{h}{f}(\frac{n}{2} - c)$

$\tilde{x} = 14.5 + \frac{9}{12}\left(\frac{40}{2} - 17\right)$

$\tilde{x} = 144.5 + \frac{9}{12}(20 - 17)$

$\tilde{x} = 144.5 + \frac{9\times3}{12} = 144.5 + \frac{27}{12}$

$\tilde{x} = 144.5 + 2.25 = 146.75 \Rightarrow \tilde{x} = 146.8$

(b) Mode:

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

(i)Ungrouped Data and Discrete Grouped Data:

Mode = The Most frequent observation

(ii) Grouped Data (Continuous)

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

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

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

Where

l = lower class boundary of the modal class.
h = class interval size of the modal class

f_m = frequency of the modal class
f_1 = frequency of the class preceding the modal class
f_2 = frequency of the class succeeding the modal class.

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

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

Solution:

We note that most occurring observation in the given data is 6. So, Mode = 6

Example 2: Find Mode for the frequency distribution

(number of Heads) x	Frequency
1	3
2	8
3	5
4	3
5	1

Solution:

Since for x = 2 the frequency is maximum, means 2 heads appear the maximum number of times i.e. 8. So, Mode = 2

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

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

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

(a) Determine the class boundaries first

(b) Find the class with maximum frequency

Classes / Groups	Class Boundaries	Frequency
0 – 9	–0.5 – 9.5	2
10 – 19	9.5 – 19.5	10
20 – 29	19.5 – 29.5	5
30 – 39	29.5 – 39.5	9
40 – 49	39.5 – 49.5	6
50 – 59	49.5 – 59.5	7
60 – 69	59.5 – 69.5	1
Total		**40**

From the above table we see that the group (9.5 – 19.5) has maximum frequency (i.e.10) so, Modal Group = (9.5 – 19.5)

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

$f_m = 10,\ l = 9.5,\ h = 10,\ f_1 = 2$ and $f_2 = 5$

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

$$\text{Mode} = 9.5 + \frac{8}{20-7} \times 10$$

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

(4) Geometric Mean:

Geometric mean of a variable x is the n^{th} positive root of the product of the $x_1, x_2, x_3, \ldots\ldots, x_n$ observation. In symbols we write, (Basic Formula)

$$G.M = (X_1 . X_2 . X_3 . \ldots\ldots . X_n)^{1/n}$$

The Geometric mean can also be calculated by using logarithm.

For Ungrouped data

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

For Grouped data

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

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

Solution: (i) Using basic formula

$G.M = (x_1 . x_2 . x_3)^{\frac{1}{3}}$

$G.M = (2 \times 4 \times 8)^{1/3}$

$G.M = (64)^{1/3} = 4$

(ii) Using logarithmic formula.

X	Log X
2	0.3010
4	0.6021
8	0.9031
Total	$\sum \log X = 1.8062$

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

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

G.M = Antilog (0.6021)

G.M = 4.00003 = 4

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

Marks in Percentage	Frequency / (No of students)
33–40	28
41–50	31
51–60	12
61–70	9
71–75	5

Solution: We proceed as follows:

Marks in %	f	X	log X	f log X
33–40	28	36.5	1.562293	43.7442
41–50	31	45.5	1.658011	51.39835
51–60	12	55.5	1.744293	20.93152
61–70	9	65.5	1.816241	16.34617
71–75	5	73	1.863323	9.316614
$\sum f = 85$			$\sum f \log X = 141.7369$	

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

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

G.M = Antilog (1.66749) = 46.50

(5) **Harmonic Mean:**

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

For Ungrouped data.

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

For Grouped data.

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

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

X	12	5	8	4

Solution:

X	1/X
12	0.0833
5	0.2
8	0.125
4	0.25
Total	$\sum \frac{1}{X} = 0.6583$

$$H.M = \frac{n}{\sum \frac{1}{X}} = \frac{4}{0.6583} = 6.076$$

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

Classes	o. of Students
33–40	28
41–50	31
51–60	12
61–70	9
71–75	5

Solution:

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

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

$$H.M = \frac{85}{1.8706} = 45.441$$

Properties of Arithmetic Mean:

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

(b) Mean is affected by change in origin.
(c) Mean is affected by change in scale.
(d) Sum of the deviations of the variable x from its mean is always zero.

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

Solution: Since the variable say X here is taking same observations so by property (i) $\bar{X} = 34$

Verification:

$$\bar{X} = \frac{\sum X}{n} = \frac{34+34+34+34+34+34}{6}$$

$$\bar{X} = \frac{204}{6} \Rightarrow \bar{X} = 34$$

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

Solution: Given the values of X.

X: 4, 5, 8, 6, 2

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

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

From the above table we get.

$$\bar{X} = \frac{\sum x}{n}, \quad \bar{X} = \frac{25}{5} = 5$$

$$\bar{Y} = \frac{\sum y}{n}, \quad \bar{Y} = \frac{50}{5} = 10$$

$$\bar{Z} = \frac{\sum z}{n}, \quad \bar{Z} = \frac{250}{5} = 50$$

We note that

a. $\bar{Y} = 10 = 5+5 = \bar{X}+5$
b. $\bar{Z} = 50 = 10(5) = 10\bar{X}$

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

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

The Weighted Arithmetic Mean

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

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + \ldots + w_nx_n}{w_1 + w_2 + \ldots + w_n} = \frac{\sum wx}{\sum w}$$

Is called the weighed arithmetic mean.

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

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

Solution:

Number of employee are treated as a weight *(w)* and monthly earnings as variable (x).

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

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

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

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

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

Solution

Days	Attendance	3–days moving	
		Total	Average
Sun.	24	–	–
Mon.	55	107	107/3 = 35.67
Tue.	28	128	128/3 = 42.67
Wed.	45	124	124/3 = 41.33
Thu.	51	150	150/3 = 50
Fri.	54	165	165/3 = 55
Sat.	60	–	–

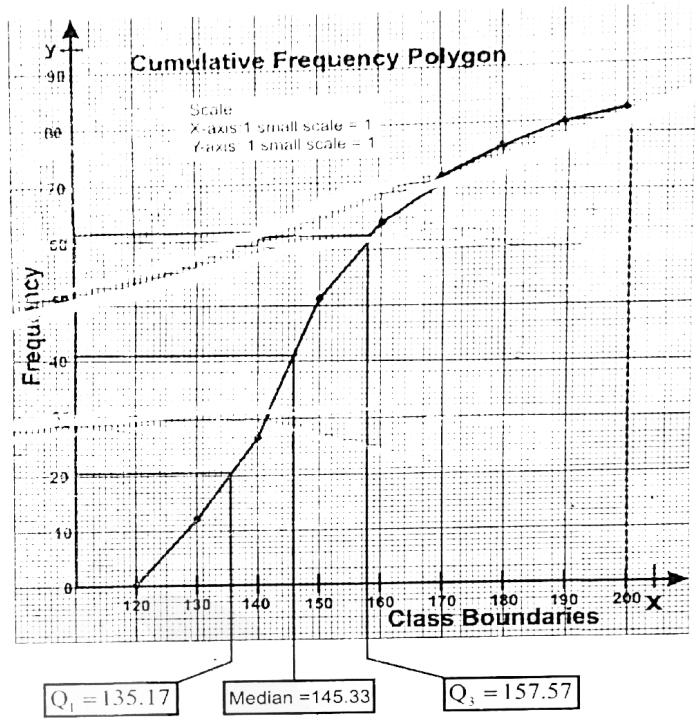
By adding the first three values we get 107 which is placed at the center of these values i.e., Monday and then dropping the first observation i.e., 24 and adding the next 3 values, we get 128 and placed at the middle of these three values and so on. For average values, divide 3 days moving total by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

Graphical Location of Median, Quartiles and Mode.

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

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

Cumulative Frequency Polygon:



Finding

Q_1:

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

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

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

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

Finding

Q_2 or Median

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

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

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

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

Finding Q_3

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

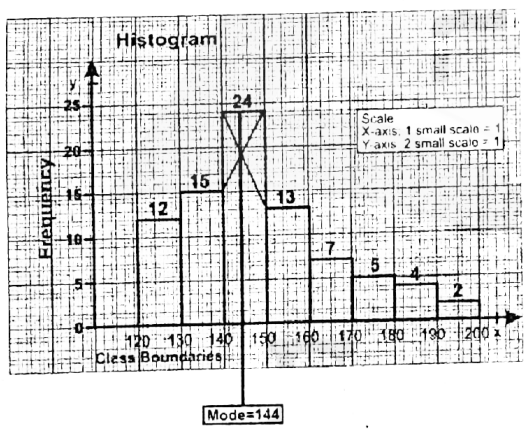
b. On the graph locate 61.5 on y–axis and draw a horizontal line segment parallel to x–axis touching the polygon.

c. Draw a vertical line segment from the point touching x–axis.

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

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

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



Steps:

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

EXERCISE 6.2

Q. 1. What do you mean by measures of central tendency?

Ans. The Specific value of the variable around which the majority of the observations of the data tend to concentrate is called average or central value. The measures or techniques that are used to determine the central value are called measure of central tendency.

Examples.

i. Arithmetic mean
ii. Median
iii. Mode
iv. Geometric mean
v. Harmonic mean

Q.2. Define Arithmetic mean, geometric mean, Harmonic mean, mode and median.

Answer

(i) Arithmetic Mean:

Arithmetic mean is a measure that determines a value of the variable under study by dividing the sum of all values of the variable by their number of observations. We denote Arithmetic mean by $\overline{X}$. In symbols it is defined as.

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

(ii) Mode:

The most frequent occurring observation in the data or the observation that occurs maximum number of times in the given data is called mode

Mode = the most frequent or common value in the ungrouped data.

(iii) Median:

Median is the middle most observation in an arranged data set. It divides the data set into two equal parts. Median is

represented by $\tilde{x}$. For ungrouped data median is calculated by the following formula.

- **When n is odd**

$$\tilde{x} = \left(\frac{n+1}{2}\right) \text{th term in arranged data}$$

- **When n is even**

$$\tilde{x} = \frac{1}{2}\left[\frac{n}{2} \text{th term} + \frac{n+2}{2} \text{ th term}\right]$$

(iv) Geometric Mean:

Geometric mean of a variable x is the n^{th} positive root of the product of the $x_1, x_2, x_3, \ldots\ldots\ldots, x_n$ observation. In symbols we write, (Basic Formula)

$$G.M = (X_1 . X_2 . X_3 . \ldots\ldots . X_n)^{1/n}$$

The Geometric Mean can also be calculated by using logarithm.

- **For Ungrouped data:**

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

- **For Grouped data:**

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

(v) Harmonic Mean:

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

- **For Ungrouped data:**

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

- **For Grouped data:**

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

Q.3. Find arithmetic mean by direct method for the following set of data.

(i) 12, 14, 17, 20, 24, 29, 35, 45
(ii) 200, 225, 350, 375, 270, 320, 290

(i) Solution:

Arithmetic mean by direct formula

X = 12, 14, 17, 20, 24, 29, 35, 45

$$\overline{X} = \frac{\sum X}{n}$$

$$\overline{X} = \frac{12+14+17+20+24+29+35+45}{8}$$

$$\overline{X} = \frac{196}{8}$$

$$\overline{X} = 24.5$$

(ii) Solution:

Arithmetic mean by direct formula

x = 200, 225, 350, 375, 270, 320, 290

$$\overline{X} = \frac{\sum X}{n}$$

$$\overline{X} = \frac{200+225+350+375+270+320+290}{7}$$

$$\overline{X} = \frac{2030}{7}$$

$$\overline{X} = 290$$

Q.4. For each of the data in Q. No. 3. Compute arithmetic mean using indirect method.

(i) Solution:

Arithmetic Mean by indirect (short) formula

X = 12, 14, 17, 20, 24, 29, 35, 45

Let Assumed mean = A = 20

Number of Observations = n = 8

X	D = X – A
12	12 – 20 = –8
14	14 – 20= –6
17	17 – 20 = –3
20	20 – 20 = 0
24	24 – 20 = 4
29	29 – 20 = 9
35	35 – 20 = 15
45	45 – 20 = 25
$\sum D = 36$	

$$\overline{X} = A + \frac{\sum D}{n}$$

$$\overline{X} = 20 + \frac{36}{8}$$

$$\overline{X} = 20 + 4.5$$

$$\overline{X} = 24.5$$

(ii) Solution:

Arithmetic mean by indirect (short) formula

Let assumed mean = A = 225

Number of observations = n = 7

X	D = X – A
200	200 – 225 = –25
225	225 – 225 = 0
270	270 – 225 = 45
290	290 – 225 = 65
320	320 – 225 = 95
350	350 – 225 = 125
375	375 – 225 = 150
$\sum D = 455$	

$$\bar{X} = A + \frac{\sum D}{n}$$

$$\bar{X} = 225 + \frac{455}{7}$$

$$\bar{X} = 225 + 65$$

$$\bar{X} = 290$$

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

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

Solution:

(i) **Arithmetic mean by Direct Method**

Classes / Groups	f	Class Marks or Mid Points (x)	fx
0 – 9	2	4.5	9
10 – 19	10	14.5	145
20 – 29	5	24.5	122.5
30 – 39	9	34.5	310.5
40 – 49	6	44.5	267
50 – 59	7	54.5	381.5
60 – 69	1	64.5	64.5
Total	$\sum f = 40$		$\sum fx = 1300$

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{1300}{40}$$

$$\bar{X} = 32.5$$

$\bar{X}$ = 32.5 Marks

(ii) **Arithmetic mean by indirect (short) method**

Let assumed mean = A= 34.5

Class /Groups	f	Class Mark (x)	D= X–A	fD
0 – 9	2	4.5	4.5 – 34.5 = –30	–60
10 – 19	10	14.5	14.5 – 34.5 = –20	–200
20 – 29	5	24.5	24.5 – 34.5= –10	–50
30 – 39	9	34.5	34.5 – 34.5 = 0	0
40 – 49	6	44.5	44.5 – 34.5 = 10	60
50 – 59	7	54.5	54.5 – 34.5 = 20	140
60 – 69	1	64.5	64.5 – 34.5 = 30	30
Total	$\sum f = 40$			$\sum fD = -80$

$$\bar{X} = A + \frac{\sum fD}{\sum f}$$

$$\bar{X} = 34.5 + \frac{-80}{40}$$

$$\bar{X} = 34.5 + (-2)$$

$$\bar{X} = 34.5 - 2$$

$\bar{X}$ = 32.5 Marks

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

Class Limits	Frequency
4 – 6	10
7 – 9	20
10 – 12	13
13 – 15	7
Total	50

Also compute Geometric mean and Harmonic mean

Solution:

(i) Arithmetic mean by direct formula.

Class Limits	Frequency (f)	Class Marks (x)	fx
4 – 6	10	5	50
7 – 9	20	8	160
10 – 12	13	11	143
13 – 15	7	14	98
Total	$\sum f = 50$		$\sum fx = 451$

$$\overline{X} = \frac{\sum fx}{\sum f} = \frac{451}{50}$$

$$\overline{X} = 9.02$$

(ii) Arithmetic Mean by Indirect (Short) method

Here A = 8 Assumed mean (A) = 8

Class Limits	f	Class Marks (x)	D= x–A	fD
4 – 6	10	5	5 – 8= –3	– 30
7 – 9	20	8	8 – 8= 0	0
10 – 12	13	11	11 – 8=3	39
13 – 15	7	14	14 – 8=6	42
Total	50			$\sum fD - 51$

$$\overline{X} = A + \frac{\sum fD}{\sum f} \Rightarrow \overline{X} = 8 + \frac{51}{50}$$

$$\overline{X} = 8 + 1.02 \Rightarrow \overline{X} = 9.02$$

(iii) Geometric Mean:

Class Limit	f	Marks (x)	$\log x$	$f \log x$
4 – 6	10	5	0.6990	6.990
7 – 9	20	8	0.9031	18.062
10 – 12	13	11	1.0414	13.5382
13 – 15	7	14	1.1461	8.0227
Total	50		$\sum f \log x = 46.6129$	

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

$$G.M = \text{Anti log}\left(\frac{46.6129}{50}\right)$$

$$G.M = \text{Anti log}(0.932258)$$

or $$G.M = \text{Anti log}(0.9323)$$

$$G.M = 8.557$$

(iv) Harmonic Mean:

Class Limits	f	Class Marks (x)	f/x
4–6	10	5	2
7–9	20	8	2.5
10–12	13	11	1.1818
13–15	7	14	0.5
Total	$\sum f = 50$		$\sum f/x = 6.1818$

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

$$H.M = \frac{50}{6.1818} = 8.088$$

or H.M = 8.089

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

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

Solution:

Arranged Data:

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

Mode:

From the arranged data we see that 4 and 9 both are repeated 6 times i.e. the most frequent values are 4 and 9. So,

Mode = 4 and 9

Median:

As n is even (n=38). So,

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

$$\text{Median} = \frac{1}{2}\left[\frac{38}{2}\text{th item} + \frac{38+2}{2}\text{th item}\right]$$

$$\text{Median} = \frac{1}{2}[19\text{th item} + 20\text{th item}]$$

$$\text{Median} = \frac{1}{2}[7+7]$$

$$\text{Median} = \frac{1}{2}[14]$$

$$\text{Median} = 7$$

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

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

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

X	Frequency	Cumulative frequency
1	3	3
2	8	3 + 8 = 11
3	5	11 + 5 = 16
4	3	16 + 3 = 19
5	1	19 + 1 = 20

(i) **Mode:** As most frequent number is 2 i.e., 2 is repeated 8 times. So, Mode = 2 Ans.

(ii) **Median:**

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

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

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

Median = 2

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

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

Solution:

(i) **Arithmetic Mean**

Class Intervals	Frequency (f)	Class marks (x)	fx
1 – 3	2	2	4
4 – 6	3	5	15
7 – 9	5	8	40
10 – 12	4	11	44
13 – 15	6	14	84
16 – 18	2	17	34
19 – 21	1	20	20
	$\Sigma f = 23$		**$\Sigma fx = 241$**

$$\overline{X} = \frac{\Sigma fx}{\Sigma f} = \frac{241}{23} = 10.478kg$$

(ii) **Median:**

Class Intervals	(f)	Class Boundaries	Cumulative Frequency
1 – 3	2	0.5 – 3.5	2
4 – 6	3	3.5 – 6.5	2 + 3 = 5
7 – 9	5	6.5 – 9.5	5 + 5 = 10
10 – 12	**4**	**9.5 – 12.5**	**10 + 4 = 14**
13 – 15	6	12.5 – 15.5	14 + 6 = 20
16 – 18	2	15.5 – 18.5	20 + 2 = 22
19 – 21	1	18.5 – 21.5	22 + 1 = 23

Median = class containing

$$\left(\frac{n}{2}\right)^{th} = \left(\frac{23}{2}\right)^{th} = 11.5^{th} \text{ item}$$

As 11.5^{th} item is present in the group (9.5 – 12.5) so median group is (9.5 – 12.5).

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

$$\text{Median} = 9.5 + \frac{3}{4}\left(\frac{23}{2} - 10\right)$$

$$\text{Median} = 9.5 + \frac{3}{4}(11.5 - 10)$$

$$\text{Median} = 9.5 + \frac{3}{4}(1.5)$$

$$\text{Median} = 9.5 + \frac{4.5}{4}$$

$$\text{Median} = 9.5 + 1.125$$

$$\text{Median} = 10.625 \text{ kg}$$

(iii) **Mode:**

Class Intervals	Frequency	Class Boundaries
1 – 3	2	0.5 – 3.5
4 – 6	3	3.5 – 6.5
7 – 9	5	6.5 – 9.5
10 – 12	$f_1 \rightarrow$ 4	9.5 – 12.5
13 – 15	$f_m \rightarrow$ 6	12.5 – 15.5
16 – 18	$f_2 \rightarrow$ 2	15.5 – 18.5
19 – 21	1	18.5 – 21.5

As the group (12.5–15.5) has maximum frequency (i.e. 6) so modal group is (12.5–15.5).

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

$$\text{Mode} = 12.5 + \frac{(6-4)}{2(6) - 4 - 2} \times 3$$

$$\text{Mode} = 12.5 + \frac{(2 \times 3)}{12 - 6}$$

$$\text{Mode} = 12.5 + \frac{6}{6}$$

$$\text{Mode} = 12.5 + 1$$

$$\text{Mode} = 13.5 \text{ kg}$$

Q.10 A student obtained the following marks at a certain examination: English 73, Urdu 82, Mathematics 80, History 67 and Science 62.

(i) If the weights accorded these marks are 4, 3, 3, 2 and 2, respectively, what is an appropriate average mark?

(ii) What is the average mark if equal weights are used?

Solution:

Subjects	Marks (x)	Weights (w)	Wx
English	73	4	292
Urdu	82	3	246
Maths	80	3	240
History	67	2	134
Science	62	2	124
		$\sum w = 14$	$\sum wx = 1036$

$$\text{Weighted mean } \bar{X}_w = \frac{\sum wx}{\sum w}$$

$$\bar{X}_w = \frac{1036}{14}$$

$$\bar{X}_w = 74 \text{ marks}$$

(ii) Solution: Let equal weight is 2, then

Subjects	Marks	Weight	Wx
English	73	2	146
Urdu	82	2	164
Maths	80	2	160
History	67	2	134
Science	62	2	124
		$\sum w = 10$	$\sum wx = 728$

$$\text{Equal weighted mean} = \frac{\sum wx}{\sum w} = \frac{728}{10}$$

$$= 72.8 \text{ marks}$$

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

Solution:

Liters (w)	Price per liters (x)	Payments (Rs) (wx)
21.3	39.90	849.87
18.7	42.90	802.23
23.5	40.90	961.15
Total liters Σw = 63.5		**Total Payments Σwx = 2613.25**

$$\text{Average price} = \frac{\text{Total payment}}{\text{Total liters}}$$

$$\bar{X}_w = \frac{\sum wx}{\sum w}$$

$$\bar{X}_w = \frac{2613.25}{63.5}$$

$$\bar{X}_w = 41.15 \text{ rupees per liter}$$

Q. 12 Calculate simple moving average of 3 year from the following data.

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

Solution:

Years	Values	3 years Moving	
		Total	Average
2001	102	–	–
2002	108	340	$\frac{340}{3}=113.33$
2003	130	378	$\frac{378}{3}=126$
2004	140	428	$\frac{428}{3}=142.66$
2005	158	478	$\frac{478}{3}=159.33$
2006	180	534	$\frac{534}{3}=178$
2007	196	586	$\frac{586}{3}=195.33$
2008	210	626	$\frac{626}{3}=208.67$
2009	220	660	$\frac{660}{3}=220$
2010	230	–	–

Q.13 Determine graphically for the following data and check your answer by using formulae.

(i) Median and Quartiles using cumulative frequency polygon.

(ii) Mode using Histogram

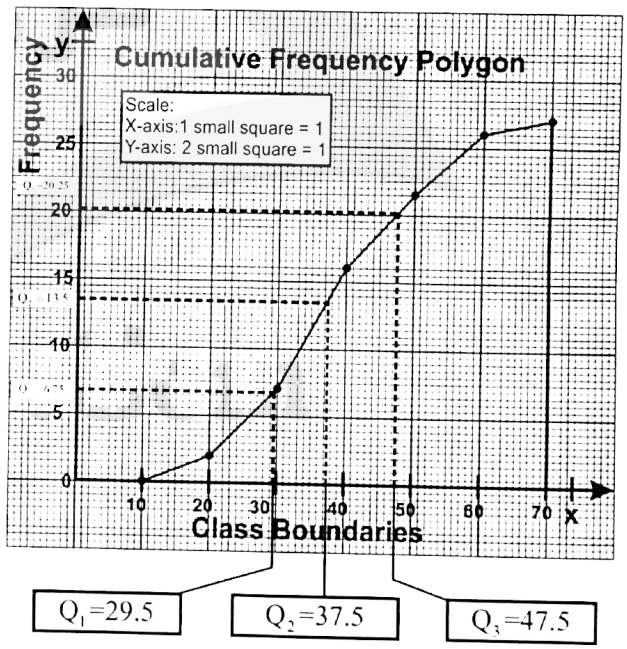
Class Boundaries	Frequency
10 – 20	2
20 – 30	5
30 – 40	9
40 – 50	6
50 – 60	4
60 – 70	1

Solution:

(i) Less than Frequency Table:

Class Boundaries	C. F
Less than 10	0
Less than 20	2
Less than 30	7
Less than 40	16
Less than 50	22
Less than 60	26
Less than 70	27

Cumulative Frequency Polygon:



Finding Q_1: Take the value of Q_1 on the x-axis against the $\frac{27}{4}$th observation (6.75) on the y-axis which is 29.5.

Finding Q_2: Take the value of Q_2 on the x-axis against the $2\left(\frac{27}{4}\right)$th observation (13.5) on the y-axis which is 37.5.

Finding Q_3: Take the value of Q_3 on the x-axis against the $3\left(\frac{27}{4}\right)$th observation (20.25) on the y-axis which is 47.5.

Verification by Formulae:

Class Boundaries	f	C. F
10–20	2	2
20–30	5	2+5=7
30–40	9	7+9=16
40–50	6	16+6=22
50–60	4	22+4=26
60–70	1	26+1=27

Verification of Q_1 by Formula:

Q_1 Class = Class Containing $(\frac{n}{4})$th item

Q_1 Class = Class Containing $(\frac{27}{4})$th item

Q_1 Class = Class Containing (6.75)th item

As 6.75 item is present in the group (20–30) so Q_1 Class is (20–30)

$$\text{Value of } Q_1 = l + \frac{h}{f}(\frac{n}{4} - c)$$

$$= 20 + \frac{10}{5}\left(\frac{27}{4} - 2\right)$$

$$= 20 + \frac{10}{5}(6.75 - 2)$$

$$= 20 + \frac{10}{5}(4.75)$$

$$= 20 + \frac{47.5}{5}$$

$$= 20 + 9.5$$

$$Q_1 = 29.5$$

This result (29.5) is exactly same as obtained from the graph.

Verification of Median (Q_2) by Formula:

Median Class = Class Containing $2\left(\frac{n}{4}\right)$th item

Median Class = Class Containing $(\frac{27}{2})$th item

Median Class = Class Containing (13.5)th item

As 13.5 item is present in the group (30–40) so, Median Group is (30–40)

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

$$= 30 + \frac{10}{9}\left(\frac{27}{2} - 7\right)$$

$$= 30 + \frac{10}{9}(13.5 - 7)$$

$$= 30 + \frac{10}{9}(6.5)$$

$$= 30 + \frac{65}{9}$$

$$= 30 + 7.22 \Rightarrow Q_2 = 37.22$$

The result is very close to the value (37.5) which is obtained from graph.

Verification of Q_3 by Formula:

Q_3 Class = Class Containing $3(\frac{n}{4})$th item

Q_3 Class = Class Containing $3(\frac{27}{4})$th item

Q_3 Class = Class Containing $3(6.75)$th item

Q_3 Class = Class containing 20.25^{th} item

As 20.25^{th} item is present in the group (40–50) so, Q_3 Class is (40–50).

$$\text{Value of } Q_3 = l + \frac{h}{f}(\frac{3n}{4} - c)$$

$$= 40 + \frac{10}{6}\left(\frac{3\times 27}{4} - 16\right)$$

$$= 40 + \frac{10}{6}(20.25\text{-}16)$$

$$= 40 + \frac{10}{6}(4.25)$$

$$= 40 + \frac{42.5}{6}$$

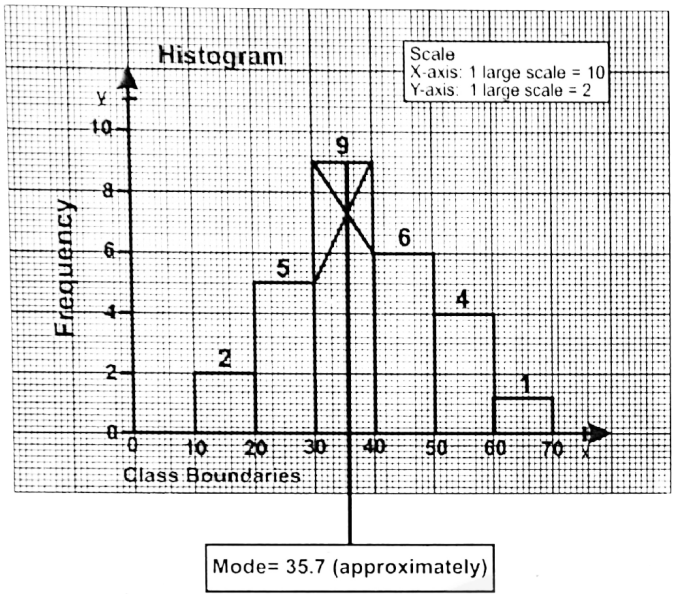
$$= 40 + 7.08 \Rightarrow Q_3 = 47.08$$

The result is very close to the value (47.5) which is obtained from graph.

(ii) Solution:

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

Histogram:



From the Graph

Mode = 35.75

Verification of Mode by Formula

Class Boundaries	Frequency
10–20	2
20–30	$f_1 \rightarrow$ 5
30–40	$f_m \rightarrow$ 9
40–50	$f_2 \rightarrow$ 6
50–60	4
60–70	1

As the group (30–40) has maximum frequency (9).

So modal group is (30–40)

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

$$= 30 + \frac{(9-5)\times 10}{2(9)-5-6}$$

$$= 30 + \frac{4\times 10}{18-11}$$

$$= 30 + \frac{40}{7}$$

$$= 30 + 5.71$$

$$\textbf{Mode} = \mathbf{35.71}$$

This result is very close to the value (35.5) which is obtained from the graph.

Dispersion

Statistically, dispersion means the spread or scatterness of observations in a data set.

The spread or scatterness in a data set can be seen in two ways.

i. The spread between two extreme observations in a data set.

ii. The spread of observations around an average say their arithmetic mean.

Measures of Dispersion

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

(1) Range:

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

Range = $X_{max} - X_{min} = X_m - X_o$ Where,

$X_{max} = X_m$: the maximum observation.

$X_{min} = X_o$: the minimum observation.

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

Range = (upper class boundary of last group)– (lower class boundary of first group)

Example 1: Find Range for the following weights of students.

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

Solution: Given that,

Maximum Value = $X_m = 110$

Minimum Value = $X_o = 49$

$R = X_m - X_o$

$R = 110 - 49 = 61$

Example 2: Find the Range for the following distribution.

Class / groups	f
10–19	10
20–29	5
30–39	9
40–49	6
50–59	7
60–69	1
Total	$\sum f = 38$

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

Class / groups	Class Boundaries	f
10–19	9.5–19.5	10
20–29	19.5–29.5	5
30–39	29.5–39.5	9
40–49	39.5–49.5	6
50–59	49.5–59.5	7
60–69	59.5–69.5	1
Total		$\sum f = 38$

Upper class boundary of last group = 69.5
Lower class boundary of first group = 9.5
Range = 69.5–9.5 = 60

(2) Variance:

Variance is defined as the mean of the squared deviation of X_i (i = 1,2,……., n) observations from their arithmetic mean. Variance is denoted by S^2.
In symbols.

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

(3) Standard Deviation:

Standard deviation is defined as the positive square root of mean of the squared deviations of X_i (i= 1,2,….., n) observations from their arithmetic mean. It is denoted by "S" and it is abbreviated as "S.D".
In symbols Standard Deviation of X is

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

Computation of variance and standard deviation

Ungrouped Data:

The formulae of variance are given by:

- Definitional Formula

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

- Computational Formula

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

The formulae of standard deviation are given by:

- Definitional Formula

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

- Computational Formula

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

Example 3: The marks of six students in Mathematics are as follows. Determine variance and standard deviation. Also interpret the results.

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

Solution: Let X = marks of a student.

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

Mean: $\bar{X} = \frac{\sum X}{n}$, $\bar{X} = \frac{372}{6} = 62$ marks

Variance:

- By Definitional Formula

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

$$S^2 = \frac{2600}{6} = 433.33 \text{ (square marks)}$$

- By Computational Formula

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

$$S^2 = \frac{25664}{6} - \left(\frac{372}{6}\right)^2 = 42773.333 - 3844$$

$$S^2 = 433.333 \text{(square marks)}$$

Standard Deviation:

- By Definitional Formula

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

$$S = \sqrt{\frac{2600}{6}} = \sqrt{433.33} = 20.81 \text{ marks}$$

- By Computational Formula

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

$$S = \sqrt{\frac{25664}{6} - \left(\frac{372}{6}\right)^2} = \sqrt{42773.333 - 3844}$$

$$S = \sqrt{433.333} = 20.81 \text{ marks}$$

Grouped Data:

Following Formulae are used to find variance and standard deviation

Variance:

- Definitional Formula

$$\text{Var}(x) = S^2 = \frac{\sum f(x-\bar{x})^2}{\sum f}$$

- Computational Formula

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

Standard Deviation:

- Definitional Formula

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

- Computational Formula

$$S.D(x) = S = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

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

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

Solution:

X (gm)	f	fx	$x-\bar{x}$	$(x-\bar{x})^2$	$f(x-\bar{x})^2$
4.5	2	9	–28	784	1568
14.5	10	145	–18	324	3240
24.5	5	122.5	–8	64	320
34.5	9	310.5	2	4	36
44.5	6	267	12	144	864
54.5	7	381.5	22	484	3388
64.5	1	64.5	32	1024	1024
$\sum f = 40$		$\sum fx = 1300$		2600	10440

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{1300}{40} = 32.5$$

Variance:

- Using Definitional formula:

$$S^2 = \frac{\sum f(x-\bar{x})^2}{\sum f}$$

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

- Using computational formula:

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

$$S^2 = \frac{52690}{40} - \left(\frac{1300}{40}\right)^2$$

$$S^2 = 1317.25 - (32.5)^2$$

$$S^2 = 1317.25 - 1056.25$$

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

Standard Deviation

- Using Definitional formula:

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

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

- Using computational formula:

$$S.D(x) = S = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

$$S = \sqrt{\frac{10440}{40} - \left(\frac{1300}{40}\right)^2} = \sqrt{1317.25 - (32.5)^2}$$

$$S = \sqrt{1317.25 - 1056.25} = \sqrt{261} = 16.155 \text{ gm}$$

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

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

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

X	Y	$X-\bar{X}$	$(X-\bar{X})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
60	62	–2	4	0	0
70	62	8	64	0	0
30	65	–32	1024	3	9
90	68	28	784	6	36
80	67	18	324	5	25
40	48	–22	484	–14	196
372	**372**	**0**	**2684**	**0**	**266**

Mean for group A = $\bar{x}=\frac{\sum x}{n}=\frac{372}{6}=62$

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

S.D(x) = $\sqrt{\frac{\sum(x-\bar{x})^2}{n}}=\sqrt{\frac{2684}{6}}=\sqrt{447.333}=21.15$ gm

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

Comparison:

From the results we conclude that the variation in Group B is smaller than that of Group A. This implies the marks of students in Group B are closer to their Mean than that of Group A.

EXERCISE 6.3

Q.1 What do you understand by Dispersion?

Dispersion: Statistically, dispersion means the spread or scatterness of observations in a data set.

Q.2 How do you define measures of dispersion?

Measures of Dispersion:

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

For example Range, Variance and Standard Deviation etc.

Q.3 Define range, standard deviation and variance?

Definitions:

Range:

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

Range = $X_{max} - X_{min} = X_m - X_o$ Where,

$X_{max} = X_m$: the maximum, observation.

$X_{min} = X_o$: the minimum observation.

Variance:

Variance is defined as the mean of the squared deviation of X_i,(i=1,2,3,....., n) observations from their arithmetic mean.

Variance from Ungrouped Data

- **Definitional Formula**

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

- **Computational formula:**

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

Variance from Grouped Data

- **Definitional Formula**

$$\text{Var}(x) = S^2 = \frac{\sum f(x - \bar{x})^2}{\sum f}$$

- **Computational Formula**

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

Standard Deviation:

Standard deviation is defined as the positive square root of mean of the squared deviations of X_i, (i= 1,2,3....., n) observations from their arithmetic mean.

Standard Deviation from Un-Grouped Data

- **Definitional Formula**

$$S.D(X) = S = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

- **Computational Formula**

$$S.D(x) = S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

Standard Deviation from Grouped Data

- **Definitional Formula**

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

- **Computational Formula**

$$S.D(x) = S = \sqrt{\frac{\sum fx^2}{\sum f} - \left(\frac{\sum fx}{\sum f}\right)^2}$$

Q.4 The salaries of five teachers in Rupees are as follows. 11500, 12400, 15000, 14500, 14800

Find range and standard deviation

Solution

Maximum value = X_{max} = Rs. 15000

Minimum value = X_{min} = Rs. 11500

Range (R)= ?

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

$R = 15000 - 11500$

Range = Rs. 3500

Standard Deviation:

Let X = 11500, 12400, 15000, 14500, 14800

We know that $\bar{X} = \frac{\sum X}{n} = \frac{68200}{5} = 13640$

X	$x - \bar{x}$	$(x - \bar{x})^2$
11500	11500 – 13640 = –2140	4579600
12400	12400 – 13640 = –1240	1537600
15000	15000 – 13640 = 1360	1849600
14500	14500 – 13640 = 860	739600
14800	14800 – 13640 = 1160	1345600
$\sum x = 68200$	$\sum (x-\bar{x})^2 = 10052000$	

Standard Deviation:

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n}}$$

$$S = \sqrt{\frac{10052000}{5}}$$

$$S = \sqrt{2010400}$$

$$S = 1417.88575$$

Q. 5

a. Find the standard deviation "S" of each set of numbers.

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

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

b. Calculate variance for the data:

10, 8, 9, 7, 5, 12, 8, 6, 8, 2

(i) Solution:

X = 12, 6, 7, 3, 15, 10, 18, 5

Number of observations = n = 8

X	X^2
3	9
5	25
6	36
7	49
10	100
12	144
15	225
18	324
$\sum x = 76$	$\sum x^2 = 912$

$$S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$S = \sqrt{\frac{912}{8} - \left(\frac{76}{8}\right)^2}$$

$$S = \sqrt{114 - 90.25}$$

$$S = \sqrt{23.75}$$

$$S = 4.87$$

(ii) Solution:

X = 9, 3, 8, 8, 9, 8, 9, 18

Number of observations = n = 8

X	X^2
3	9
8	64
8	64
8	64
9	81
9	81
9	81
18	324
$\sum x = 72$	$\sum x^2 = 768$

We know that,

$$S = \sqrt{\frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2}$$

$$S = \sqrt{\frac{768}{8} - \left(\frac{72}{8}\right)^2}$$

$$S = \sqrt{96 - 81}$$

$$S = \sqrt{15} \quad \Rightarrow \quad S = 3.87$$

(b). Calculate variance for the data

10, 8, 9, 7, 5, 12, 8, 6, 8, 2

Solution:

Number of Observations = n = 10

X	X^2
10	100
8	64
9	81
7	49
5	25
12	144
8	64
6	36
8	64
2	4
$\sum x = 75$	$\sum x^2 = 631$

We know that,

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

$$S^2 = \frac{631}{10} - \left(\frac{75}{10}\right)^2$$

$$S^2 = 63.1 - (7.5)^2$$

$$S^2 = 63.1 - 56.25 \quad \Rightarrow S^2 = 6.85$$

Q. 6. The lengths of 32 items are given below. Find the mean length and standard deviation of the distribution.

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

Solution:

Length	(*f*)	Class Marks (x)	*fx*	x^2	fx^2
20–22	3	21	63	441	1323
23–25	6	24	144	576	3456
26–28	12	27	324	729	8748
29–31	9	30	270	900	8100
32–34	2	33	66	1089	2178
$\sum f = 32$		$\sum fx = 867$			$\sum fx^2 = 23805$

(i) Arithmetic Mean

$$\bar{X} = \frac{\sum fx}{\sum f} = \frac{867}{32} = 27.093$$

(ii) Standard Deviation

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

$$S = \sqrt{\frac{23805}{32} - \left(\frac{867}{32}\right)^2}$$

$$S = \sqrt{743.90 - (27.093)^2}$$

$$S = \sqrt{743.90 - 734.03065}$$

$$S = \sqrt{9.86935}$$

$$S = 3.1415$$

Q. 7 For the following distribution of marks calculate range.

Marks in percentage	Frequency / No of students
33 – 40	28
41 – 50	31
51 – 60	12
61 – 70	9
71 – 75	5

Solution:

Marks in percentage	Class Boundaries	Frequency / No of students
33 – 40	32.5 – 40.5	28
41 – 50	40.5 – 50.5	31
51 – 60	50.5 – 60.5	12
61 – 70	60.5 – 70.5	9
71 – 75	70.5 – 75.5	5

Upper class boundary of last class = 75.5
Lower class boundary of 1st class = 32.5
Range = ?
Range (R) = 75.5 – 32.5
R = 43

MISCELLANEOUS EXERCISE – 6

1. A grouped frequency table is also called……
(a) data
(b) frequency distribution
(c) frequency polygon
(d) Histogram

2. Mean of a variable with similar observations say constant k is…
(a) negative (b) k itself
(c) zero (d) one

3. A deviation is defined as a difference of any value of the variable from a……
(a) constant (b) histogram
(c) sum (d) frequency

4. Mean is affected by change in…
(a) value (b) ratio
(c) origin (d) none of these

5. Mean is affected by change in…
(a) place (b) scale
(c) rate (d) none of these

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

7. A frequency polygon is a many sided…
(a) closed figure (b) rectangle
(c) square (d) Circles

8. A data in the form of frequency distribution is called……
(a) Grouped data
(b) Ungrouped data
(c) Histogram
(d) Dispersion

9. A histogram is a set of adjacent…
(a) squares (b) rectangles
(c) circles (d) Dots

10. A cumulative frequency table is also called……
(a) frequency distribution
(b) data
(c) less than cumulative frequency distribution
(d) Histogram

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

12. The measures that are used to determine the degree or extent of variation in a data set are called measures of…….
(a) dispersion
(b) central tendency
(c) average
(d) quartile

13. The positive square root of mean of the squared deviations of x_i (i – 1,2,.... n) observations from their arithmetic mean is called.....
(a) harmonic mean (b) range
(c) S.D (d) variance

14. The value obtained by reciprocating the mean of the reciprocal of $x_1,x_2,x_3..x_n$ observations is called.....
(a) Geometric mean
(b) Median
(c) Harmonic mean
(d) S.D

15. The most frequent occurring observation in a data set is called.....
(a) Mode (b) Median
(c) Harmonic mean
(d) Mean

16. The measure which determines the middlemost observation in a data set is called.....
(a) median (b) mode
(c) mean (d) variance

17. The size of class interval (6–10) is.
(a) 4 (b) 5
(c) 8 (d) 10

18. The arrangement of data is necessary to find the value of.
(a) Mean (b) Median
(c) Mode (d) Range

19. Sum of the deviations of the variable x from its mean is always.....
(a) zero (b) one
(c) same (d) negative

20. The extent of variation between two extreme observations of a data set is measured by.....
(a) average (b) range
(c) quartiles (d) mode

21. The mean of the squared deviations of x_i (i = 1, 2.....,n) observations from their arithmetic mean is called...
(a) variance
(b) standard deviation
(c) range (d) mode

22. The n^{th} positive root of the product of the $x_1,x_2,x_3,.....x_n$ observations is called.....
(a) Mode (b) Mean
(c) Geometric mean
(d) median

23. The observation that divide a data set into four equal part, are called
(a) deciles (b) quartiles
(c) percentiles (d) mode

24. The spread or scatterness of observations in a data set is called .....
(a) average
(b) dispersion
(c) central tendency
(d) quartile

25. The class having maximum frequency is called .....class.
(a) Modal (b) Median
(c) Lower (d) Upper

26. The class containing $\frac{n}{2}$th observation is called.
(a) Modal (b) Median
(c) Boundary of (d) Size of

27. During frequency distribution number of groups should be between....
(a) 5 and 10 (b) 10 and 15
(c) 10 and 20 (d) 5 and 15

28. Direct formula to find mean from ungrouped data.
(a) $\overline{X}=\frac{\sum x}{n}$
(b) $\overline{X}=\frac{\sum fx}{\sum f}$
(c) $\overline{X}=A+\frac{\sum D}{n}$
(d) $\overline{X}=A+\frac{\sum fD}{\sum f}$

29. Formula to find Harmonic mean from grouped data is.....
(a) $\overline{X}=\frac{n}{\sum \frac{1}{x}}$

(b) $\bar{X}=\frac{n}{\sum\frac{f}{x}}$

(c) $\bar{X}=A+\frac{\sum fu}{n}\times h$

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

30. Variance is denoted by.....

(a) V (b) S

(c) S^2 (d) $\bar{X}$

31. Short formula to find mean from grouped data is.....

(a) $\bar{X}=\frac{\sum x}{n}$ (b) $\bar{X}=\frac{\sum fx}{\sum f}$

(c) $\bar{X}=A+\frac{\sum D}{n}$ (d) $\bar{X}=A+\frac{\sum fD}{\sum f}$

32. In class (30–39), lower class limit is.....

(a) 39 (b) 9

(c) 30 (d) 34.5

33. Coding formula to find mean from grouped data is.....

(a) $\bar{X}=\frac{n}{\sum\frac{1}{x}}$

(b) $\bar{X}=\frac{n}{\sum\frac{f}{x}}$

(c) $\bar{X}=A+\frac{\sum u}{n}\times h$

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

34. Formula to find Harmonic mean from ungrouped data is.....

(a) $\bar{X}=\frac{n}{\sum\frac{1}{x}}$

(b) $\bar{X}=\frac{n}{\sum\frac{f}{x}}$

(c) $\bar{X}=A+\frac{\sum fu}{n}\times h$

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

35. Direct formula to find mean from grouped data is.....

(a) $\bar{X}=\frac{\sum x}{n}$ (b) $\bar{X}=\frac{\sum fx}{\sum f}$

(c) $\bar{X}=A+\frac{\sum D}{n}$ (d) $\bar{X}=A+\frac{\sum fD}{\sum f}$

36. The concept of antilogarithm is used to find the value of.....

(a) A.M (b) G. M

(c) H. M (d) Mode

37. Short formula to find mean from ungrouped data is.....

(a) $\bar{X}=\frac{\sum x}{n}$ (b) $\bar{X}=\frac{\sum fx}{\sum f}$

(c) $\bar{X}=A+\frac{\sum D}{n}$ (d) $\bar{X}=A+\frac{\sum fD}{\sum f}$

38. In class (20–29), Midpoint or class mark is.....

(a) 20.5 (b) 24.5

(c) 29 (d) 49

39. Median from the data 82,93,86,92 and 79 is.....

(a) 82 (b) 86

(c) 92 (d) 93

40. Harmonic mean for 12,5,8,4 is...

(a) 6.08 (b) 5.08

(c) 7.08 (d) 4.08

41. Geometric Mean of 2,4,8 is.....

(a) 2 (b) 4

(c) 8 (d) 3

42. Coding formula to find mean from ungrouped data is.....

(a) $\bar{X}=\frac{n}{\sum\frac{1}{x}}$

(b) $\bar{X}=\frac{n}{\sum \frac{f}{x}}$

(c) $\bar{X}=A+\frac{\sum u}{n}\times h$

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

43. Standard deviation is denoted by.....

(a) $\tilde{X}$ (b) S

(c) S^2 (d) $\bar{X}$

44. Types of measures of central tendency are.....

(a) 3 (b) 4

(c) 5 (d) 6

45. Median is denoted by.....

(a) $\bar{X}$ (b) $\tilde{X}$

(c) S (d) S^2

46. If y = 10X then $\bar{y}$=.....

(a) 10 (b) $10\bar{X}$

(c) $\bar{X}$ (d) $10+\bar{X}$

47. If Y = X + 5 then $\bar{Y}$ =

(a) $\bar{X}$ (b) 5

(c) $\bar{X}+5$ (d) $5\bar{X}$

48. In class (10 – 19), upper class limit is.....

(a) 10 (b) 19

(c) 29 (d) 14.5

49. On the basis of types of variable or data, the types of frequency distribution are.

(a) 2 (b) 3

(c) 4 (d) 5

50. Types of dispersion are.....

(a) 4 (b) 5

(c) 6 (d) 3

51. Range for the data 110, 109, 84, 89, 77, 104, 74, 97, 49, 59, 103, 62 is.....

(a) 41 (b) 51

(c) 61 (d) 71

52. If standard deviation is 6 then its variance is.....

(a) $\sqrt{6}$ (b) 36

(c) 3 (d) 6

53. $\sum(X-\bar{X})$ = ...

(a) 0 (b) 1

(c) –1 (d) 2

54. Arithmetic mean of 34,34,34,34,34,34 is

(a) 0 (b) 34

(c) 6 (d) 170

55. Mode from the following data 4,4,5,5,6,6,6,7,7,5,8,8,8,6,5,6,5,7 is.....

(a) 4 (b) 5

(c) 5, 6 (d) 5, 7

56. Median from the data 2.3, 2.7, 2.5, 3.1 and 1.9 is.....

(a) 2.3 (b) 2.5

(c) 2.7 (d) 2.9

57. Which one is formula for weighted Arithmetic mean?

(a) $\frac{\sum w}{\sum wx}$ (b) $\sum wx$

(c) $\sum x$ (d) $\frac{\sum wx}{\sum w}$

58. Range = ...

(a) $X_m + X_o$ (b) $X_m - X_o$

(c) $\frac{X_m}{X_o}$ (d) $\frac{X_o}{X_m}$

ANSWER KEY

1.	b	**2.**	b	**3.**	a	**4.**	c	**5.**	b
6.	a	**7.**	a	**8.**	a	**9.**	b	**10.**	c
11.	b	**12.**	a	**13.**	c	**14.**	c	**15.**	a
16.	a	**17.**	b	**18.**	b	**19.**	a	**20.**	b
21.	a	**22.**	c	**23**	b	**24**	b	**25**	a
26	b	**27**	d	**28**	a	**29**	b	**30**	c
31	d	**32**	c	**33**	d	**34**	a	**35**	b
36	b	**37**	c	**38**	b	**39**	b	**40**	a
41	b	**42**	c	**43**	b	**44**	c	**45**	b
46	b	**47**	c	**48**	b	**49**	a	**50**	b
51	c	**52**	b	**53**	a	**54**	b	**55**	c
56	b	**57**	d	**58**	b				

Q. 2 Write short answer of the following questions.

(i) Define class limits.

Class Limits: The minimum and the maximum values defined for a class or group are called class limits. The minimum value is called the lower class limit and the maximum value is called the upper class limit of that class. For example in the group (5–10), 5 is lower class limit and 10 is called upper class limit.

(ii) Define class mark

Midpoint or Class Mark:

For a given class the average of that class obtained by dividing the sum of upper and lower class limits by 2, is called the midpoint or class mark of that class

(iii) What is cumulative frequency?

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

(iv) Define a frequency distribution.

Frequency Distribution

A frequency distribution is a tabular arrangement for classifying data into different groups and the number of observation falling in each group corresponds to the respective group. In fact a frequency distribution is a method to summarize data.

(v) What is Histogram?

Histogram

A Histogram is a graph of adjacent rectangles constructed on XY–plane. It is a graph of frequency distribution. In practice both discrete and continuous frequency distribution are represented by means of histogram.

(vi) Name two measures of central tendency.

1– Arithmetic Mean 2– Median 3– Mode

(vii) Define Arithmetic mean?

i. Arithmetic Mean:

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

Arithmetic mean of n observations

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

(viii) Write three properties of Arithmetic mean.

Properties of Arithmetic Mean:

Following are the properties of Arithmetic Mean.

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

(b) Mean is affected by change in origin.

(c) Mean is affected by change in scale.

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

(ix) Define Median.

Median:

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

Median From Ungrouped Data

Case 1:

When the number of observations (n) is odd.

$$\tilde{x}=\left(\frac{n+1}{2}\right)\text{th term in arranged data}$$

Case 2:

When the number of observations (n) is even.

$$\tilde{x}=\frac{1}{2}\left[\frac{n}{2}\text{th term}+\frac{n+2}{2}\text{th term}\right]$$

(x) Define mode.

Mode:

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