Exercise 6.1

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

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

Solution:

Frequency distribution of number of family members.

Number of members	**Tally marks**	**Frequency**	**Cumulative Frequency**
2	|	1	1
3	|||	3	1+ 3= 4
4	~~||||~~ |	6	4 + 6 = 10
5	||||	4	10+4=14
6	|||	3	14+3 = 17
7	~~||||~~ |	6	17+6=23
8	~~||||~~	5	23+5 = 28
9	~~||||~~ |	6	28+6=34
10	||	2	34+2 = 36
11	||	2	36+2=38
12	|	1	38+1 = 39
	Total	**39**	

---

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

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

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

Solution:

Frequency Distribution

Class Limits	Tally marks	Frequency	Midpoint	Class Boundaries
20 - 24	~~||||~~ |	6	22	19.5 – 24.5
25 - 29	~~||||~~ ~~||||~~	10	27	24.5 – 29.5
30 - 34	~~||||~~ ~~||||~~ ||	12	32	29.5 – 34.5
35 - 39	~~||||~~ ||||	9	37	34.5 – 39.5
40 - 44	|||	3	42	39.5 – 44.5
		40		

Less than Cumulative Frequency Distribution

Class Boundaries	Frequency f	Cumulative Frequency	Class Boundaries	Cumulative Frequency
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

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

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

Solution:

Class Limits	Tally Marks	Frequency
450 - 549	aa	2
550 - 649	aa	2
650- 749	D	4
750 - 849	E	5
850 - 949	C	3
950 - 1049	D	4
1050 - 1149	E	5
1150 - 1249	E	5
	Total =	30

Q4. The following data shows the daily load shedding duration in hours, in 30 localities of a certain city. Make a frequency distribution of the load shedding duration taking 2 hours as class interval size and answer the following questions.

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

a) Find the most frequent load shedding hours?

b) Find the least load shedding intervals?

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

Solution:

Frequency Distribution Table

Class Limits	Talley marks	Frequency
2 - 3	B	2
4 - 5	|	1
6 - 7	E D	9
8 - 9	E	5
10 - 11	E A	6
12- 13	E	5
14- 15	C	3
	Total =	31

(a) Find the most frequent load shedding hours.

6 - 7

(b) Find the least load shedding intervals.

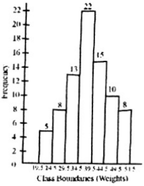
4 – 5

Q5. 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

Solution:

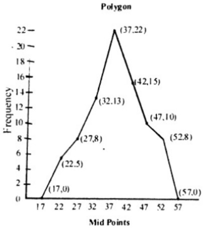
Class Boundaries	Frequency
19.5 - 24.5	5
24.5 - 29.5	8
29.5 - 34.5	13
34.5 - 39.5	22
39.5 - 44.5	15
44.5 - 49.5	10
49.5 - 54.5	8



Class Limits	Mid Points	Frequency
15 - 19	17	0
20 – 24	22	5
25 – 29	27	8
30 – 34	32	13
35 – 39	37	22
40 – 44	42	15
45 – 49	47	10
50 – 54	52	8
55 - 59	57	0

Note:

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



Measures of Central Tendency:

A specific value of the variable around which the majority of the observations tend to concentrate, this representative shows the tendency or behavior of the distribution of the variable under study. This value is called average or the central value. The measures or techniques that are used to determine this central value are called Measures of Central Tendency.

The following measures of central tendency will be discussed in this section:

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

Arithmetic Mean:

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

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

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.

Ungrouped Data:

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

(i) Direct Method (By Definition)

The formula under this method is given by:

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

(ii) Indirect, Short Cut or Coding Methods

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

(i) using an Assumed or Provisional mean

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

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

Deviation from mean of X = $(x_i - \overline{X})$ for i = l, 2, ...n

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

The Formulae used under indirect methods are:

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

Where

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

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

Grouped Data:

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

(a) Using Direct method

$$\overline{X} = \frac{\sum fX}{\sum f}$$

Using Indirect method,

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

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

(b) Median:

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

Ungrouped data

Case-I:

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

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

Case-2:

When the number of observations is even of a set of data arranged in order of magnitude the median is the arithmetic mean of the two middle observations. That is, median is average of

$\frac{n}{2}$ and $\left(\frac{n}{2}+1\right)^{th}$ values.

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

Grouped Data (Discrete)

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

(i) Make cumulative frequency column.

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

Grouped Data (Continuous):

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

(i) Determine class boundaries.

(ii) Make cumulative frequency column.

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

Use the formula:

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

Where

l = lower class boundary of the median class,

h = class interval size of the median class,

f = frequency of the median class,

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

Mode:

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

(i) Ungrouped data and Discrete Grouped data

Mode = the most frequent observation

(ii) Grouped Data (Continuous)

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

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

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

Where

l = lower class boundary of the modal class or group,

h = class interval size of the modal class,

f_m = frequency of the modal class,

f_1 = frequency of the class preceding the modal class

f_2 = frequency of the class succeeding the modal class.

Geometric Mean:

Geometric mean of a variable X is the nth positive root of the product of the $x_1, x_2, x_3,$

……., x_n observations. In symbols we write,

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

The above formula can also be written by using logarithm.

For Ungrouped data

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

For Grouped data

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

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,

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

And for grouped data

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

Properties of Arithmetic Mean:

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

(ii) Mean is affected by change in origin.

(iii) Mean is affected by change in scale.

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

Calculation of Weighted Mean and Moving Averages:

The Weighted Arithmetic Mean:

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

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

is called the weighted arithmetic mean.

Moving Averages:

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

Exercise 6.2

1. What do you understand by measures of central tendency?

Solution:

The specific value of the variable around which the majority of the observations tend to concentrate is called the central tendency.

2. Define Arithmetic mean, Geometric mean, Harmonic mean, mode and median.

Solution:

(i) Arithmetic Means:

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.

$\overline{X} = \frac{\sum x}{n}$ (for ungrouped data) and $\overline{X} = \frac{\sum fx}{\sum f}$ (for grouped data)

(ii) Geometric Means:

Geometric mean of a variable x is the nth positive root of the product of the x_1, x_2, x_3, ... x_n observation.

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

(iii) Harmonic Means:

Harmonic mean refers to the value obtained by reciprocating the mean of the reciprocal of x_1. x_2. x_3. .... x_n observations.

$$\text{H.M.} = \frac{n}{\sum \frac{1}{x}}$$ (for ungrouped data)

and, $$\text{H.M.} = \frac{n}{\sum \frac{f}{x}}$$ (for grouped data)

(iv) Mode:

The most repeated value in an observation is called its mode.

(v) Median:

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

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.

Solution:

(i) A.M = $\overline{X} = \frac{\sum x}{n} = \frac{12+14+17+20+24+29+35+45}{8}$

$= \frac{196}{8} = 24.5$

(ii) A.M = $\overline{X} = \frac{\sum x}{n} = \frac{200+225+350+375+270+320+290}{7}$

$= \frac{2030}{7} = 290$

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

Solution:

(i) Take any constant say 24 and take deviations from it (24).

A = 24

X	D = X - A
12	12 - 24 = -12
14	14 - 24 =-10
17	17 - 24 =-7
20	20 - 24 =-4
24	24 - 24 = 0
29	29 -24 = 5
35	35 - 24 = 11
45	45 - 24 = 21
n = 8	$\sum D$ **= 4**

$$\overline{X} = A + \frac{\sum D}{n} = 24 + \frac{4}{8} = 24 + \frac{1}{2} = 24\frac{1}{2} = 24.5$$

(ii) Take any constant 270 and take deviations from it (270).

A = 270

X	D = X - A
200	200 - 270 =-70
225	225 - 270 =-45
350	350 - 270 = 80
375	375 - 270 = 105
270	270 - 270 = 0
320	320 – 270 = 50
290	290 – 270 = 20
n = 7	$\sum D$ **= 140**

$$\overline{X} = A + \frac{\sum D}{n} = 270 + \frac{140}{7} = 270 + 20 = 290$$

---

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:

Direct Method:

Classes/groups	Mid-points	f	fx
0 - 9	4.5	2	4.5 × 2 = 9.0
10 - 19	14.5	10	14.5 × 10 =145.0
20 -29	24.5	5	24.5 × 5 = 122.5
30 -39	34.5	9	34.5 × 9 ='310.5
40 -49 .	44.5	6	44.5 × 6 = 267.0
50 - 59	54.5	7	54.5 × 7 = 381.5
60 - 69	64.5	1	64.5 × 1 = 64.5
		n = $\sum_f$ = 40	**$\sum_{fx}$=1300**

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

Indirect, short cut method:

Let A =34.5

Classes/ groups	f	Mid-point (x)	D = X - A	U = $\frac{D}{10}$	fD	$f(U)=\frac{f(D)}{10}$
0 - 9	2	4.5	4.5 - 34.5 = -30	-3	-60	-6
10 - 19	10	14.5	14.5 - 34.5 = - 20	-2	-200	-20
20 - 29	5	24.5	24.5 - 34.5 = - 10	-1	-50	-5
30 - 39	9	34.5	34.5 - 34.5 = 0	0	0	0
40 - 49	6	44.5	44.5 - 34.5 = 10	1	60	6
50 - 59	7	54.5	54.5 - 34.5 = 20	2	140	14
60 - 69	1	64.5	64.5 - 34.5 = 30	3	30	3
Total	40				-80	-8

$$\overline{X}=A+\frac{\sum fD}{\sum f}= \qquad or \qquad \overline{X}=A+\frac{\sum f(U)}{\sum f}\times h$$

$$=34.5+\frac{(-80)}{40} \qquad =34.5+\frac{-8}{40}\times h$$

$$=34.5-2 \qquad =34.5+\frac{-8}{40}\times 10$$

$$=32.5 \qquad =34.5-2=32.5$$

6. The following data relates to the ages of children in a school. Compute the mean age by direct and short-cut method taking any provisional mean.

(Hint. Take A = 8)

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

Also Compute Geometric mean and Harmonic mean.

Solution:

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

$$\text{A.M} = \frac{\sum fx}{\sum f} = \frac{451}{50} = 9.02$$

Indirect, short cut method:

Let A = 11

Classes/ groups	f	Mid-point (x)	D = X - A	U = $\frac{D}{3}$	fD	fU = $\frac{fD}{3}$
4 - 6	10	5	5 -11 = - 6	-2	-60	-20
7 - 9	20	8	8 - 11= -3	-1	-60	-20
10 - 12	13	11	11 - 11 = 0	0	0	0
13- 15	7	14	14 - 11 = 3	1	21	7
	50				-99	-33

$$\overline{X} = A + \frac{\sum fD}{\sum f} \qquad or \qquad \overline{X} = A + \frac{\sum f(U)}{\sum f} \times h$$

$$= 11 - \frac{99}{50} \qquad\qquad = 11 - \frac{33}{50} \times 3$$

$$= 11 - 1.98 \qquad\qquad = 11 - \frac{99}{50}$$

$$= 9.02 \qquad\qquad = 11 - 1.98 = 9.02$$

Geometric Mean:

We proceed as follows:

Class limits	f	Mid points x	Log x	f log x
4 - 6	10	5	0.69897	6.9897
7 - 9	20	8	0.90309	18.0618
10 - 12	13	11	1.04139	13.5380
13 - 15	7	14	1.14613	8.02291
$\sum f = 50$				$\sum f$ log x = 46.61248

$$\text{G.M} = \text{Anti log}\left(\frac{\sum f \log X}{\sum f}\right) = \text{Anti log}\left(\frac{46.61248}{50}\right)$$

$$= \text{Anti log}(0.9322496) = 8.553$$

Harmonic Means

Class limits	f	Mid points x	$\frac{f}{x}$
4 - 6	10	5	$\frac{10}{5} = 2.0$
7 - 9	20	8	$\frac{20}{8} = 2.5$
10.- 12	13	11	$\frac{13}{11} = 1.18$
13 - 15	7	14	$\frac{7}{14} = 0.50$
	$\sum f = 50$		$\sum \frac{f}{x} = 6.18$

$$\text{H.M.} = \frac{\sum f}{\sum \frac{f}{x}} = \frac{50}{6.18} = 8.09$$

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:

Writing the observations in ascending order

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:

The most frequent observation = 9, 4

Median:

Number of observations = 38

Therefore, median is the mean of 19th and 20th observation = $\frac{7 + 7}{} = 7$

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

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

Solution:

Mode:

The most frequent observation = 2

Median:

For median, we make cumulative frequency column.

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

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

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

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

$$\text{Median} = 2$$

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

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

Solution:

Class Intervals	f	Mid Points (x)	fx	Class Boundaries	Cumulative frequency
1 -3	2	2	4	0.5 - 3.5	2
4 - 6	3	5	15	3.5 - 6.5	2 + 3 = 5
7 - 9	5	8	40	6.5 - 9.5	5 + 5 = 10
10 -12	4	11	44	9.5- 12.5	10 + 4 = 14
13 -15	6	14	84	12.5 - 15.5	14 + 6 = 20
16 - 18	2	17	34	15.5 - 18.5	20 + 2 = 22
19 - 21	1	20	20	18.5 - 21.5	22 + 1 = 23
Total	23		241		

Mean:

$$\text{Mean} = \overline{X} = \frac{\sum fx}{\sum f} = \frac{241}{23} = 10.478$$

Median:

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

$= $ the class containing $\left(\frac{23}{2}\right)^{th}$ observation

$= $ the class containing $(11.5)^{th}$ observation

Median class is $9.5-12.5$

Here $l=9.5, c=10, f=4, h=3$

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

$$=9.5+\frac{3}{4}\left(\frac{23}{2}-10\right)=9.5+\frac{3}{4}\left(\frac{3}{2}\right)=9.5+\frac{9}{8}=9.5+1.125=10.625$$

Mode:

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

Here $l=12.5, f_m=6, f_1=4, f_2=2, h=3$

$$\therefore \text{Mode} = 12.5+\frac{6-4}{2(6)-4-2}\times 3=12.5+\frac{2}{6}\times 3=12.5+1=13.5$$

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:

Marks (X)	Weight (w)	Xw
73	4	73 × 4 = 292
82	3	82 × 3 = 246
80	3	80 × 3 = 240
67	2	67 × 2 = 134
62	2	62 × 2 = 124
$\sum X = 364$	$\sum w = 14$	$\sum Xw = 1036$

$(i)\ \overline{X_w}=\frac{\sum Xw}{\sum w}=\frac{1036}{14}=74$

$(ii)\ \overline{X}=\frac{\sum X}{\sum n}=\frac{364}{5}=72.8$

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:

X	w	Xw
21.3	39.90	(21.3) (39.90) = 849.87
18.7	42.90	(21.3) (39.90) = 849.87
23.5	40.90	(21.3) (39.90) = 849.87
$\sum X = 63.5$		$\sum Xw = 2613.25$

$$\text{Mean Price} = \frac{\sum Xw}{\sum X} = \frac{2613.25}{63.5} = 41.15 \text{ rupees per liter}$$

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

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

Solution:

Years	Value	3 - years	3 - years moving
2001	102	-	-
2002	108	340	340/3 = 113.33
2003	130	378	378/3 = 126.00
2004	140	428	428/3 = 142.67
2005	158	478	478/3 = 159.33
2006	180	534	534/3 = 178.00
2007	196	586	586/3 = 195.33
2008	210	626	626/3 = 208.67
2009	220	660	660/3 = 220.00
2010	230	-	-

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

(i) Median and Quartiles using cumulative frequency polygon.

(ii) Mode using Histogram

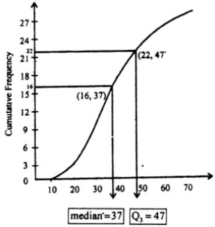
Class Boundaries	Frequency
10-20	2
20-30	5
30-40	9
40-50	6
50-60	4
60-70	1

Part (i)

Solution:

Class Boundaries	f	c. f
Less than 10	0	0
Less than 20	2	2
Less than 30	5	7
Less than 40	9	16
Less than 50	6	22
Less than 60	4	26
Less than 70	1	27
		$\sum f = 27$

Cumulative Frequency Polygon:



Median class Q₃ class

Median class $= \left(\frac{n}{2}\right)^{th}$ observation $= \left(\frac{27}{2}\right)^{th} = (13.5)^{th}$ observation

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

Here $l = 30, c = 7, f = 9, h = 10, n = 27$

Thus median $x = 30 + \frac{10}{9}\left(\frac{27}{2} - 7\right) = 30 + \frac{10}{9}\left(\frac{13}{2}\right) = 30 + 7.22 = 37.22$

To find Q_3

we have to find $3\left(\frac{n}{4}\right)^{th}$ observation

Q_3 Class $= 3\left(\frac{n}{4}\right)^{th}$ observation $= 3\left(\frac{27}{4}\right)^{th}$ observation

$= 3(6.75)^{th}$ observation $= (20.25)^{th}$ observation

Q_3 Class is 40-50.

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

Here $l = 40, c = 16, f = 6, h = 10, n = 27$

$$Q_3 = 40 + \frac{10}{6}\left(\frac{3\times 27}{4} - 16\right) = 40 + \frac{10}{6}(20.25 - 16)$$

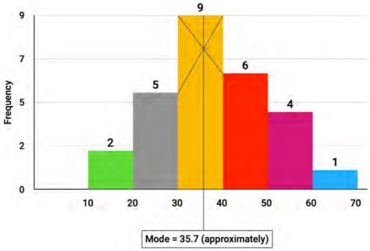
$$= 40 + \frac{10}{6}(4.25) = 40 + 7.08 = 47.08$$

Part (ii)

Solution:

Histogram:

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



From the Graph:

Mode = 35.7

Verification of Mode by Formula:

Class Boundaries	Frequency
10-20	2
20-30	f_1 → 5
30-40	f_m → 9
40-50	f_2 → 6
50-60	4
60-70	1

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

So, the modal group is (30 – 40).

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

Here $l = 30, f_m = 9, f_1 = 5, f_2 = 6, h = 10$

$$\begin{aligned}
\text{Mode} &= 30 + \frac{9-5}{2(9)-5-6} \times 10 \\
&= 30 + \frac{4 \times 10}{18-11} \\
&= 30 + \frac{40}{7} \\
&= 30 + 5.71 \\
\text{Mode} &= 35.71
\end{aligned}$$

The result is very clos to the value (35.7) which is obtained from the graph.

Measures of Dispersion:

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

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

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

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

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

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

(i) Range:

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

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

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

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

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

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

(ii) Variance:

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

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

(iii) Standard Deviation:

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

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

Computation of Variance and Standard Deviation:

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

Ungrouped Data

The formula of Variance is given by:

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

And Standard Deviation is given by:

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

Grouped Data:

The formula of Variance is given by:

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

And Standard Deviation is given by:

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

Exercise 6.3

1. What do you understand by Dispersion?

Solution:

Dispersion:

Dispersion means the spread or scatterness of observations in a data set. By dispersion we mean the extent to which the observations in a sample or in a population are spread out. The main measures of dispersion are range, variance and standard deviation.

2. How do you define measure of dispersion?

Solution:

The measures that are used to determine the degree or extent of variation in a data set are called measure of dispersion.

3. Define Range, Standard deviation and Variance.

Solution:

Range:

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

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

Range = (upper C. B of the last group) - (lower C. B of first group)

Variance:

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

$$\text{Variance} = S^2 = \frac{\sum(\mathrm{X}-\overline{\mathrm{X}})^2}{n}$$

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

Standard Deviation:

The positive square root of the squared deviations of x_i ($i = 1, 2, 3, \ldots, n$) observations from their mean.

$$\text{Standard Deviation} = S = \sqrt{\frac{\sum(\mathrm{X}-\overline{\mathrm{X}})^2}{n}}$$

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

4. The salaries of five teachers in Rupees are as follows.

11500, 12400, 15000, 14500, 14800.

Find Range and standard deviation.

Solution:

X = 11500, 12400, 15000, 14500, 14800

Here, $X_{max} = 15000$, $X_{min} = 11500$

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

= 15000 - 11500 = 3500

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

$$= \frac{11500 + 12400 + 15000 + 14500 + 14800}{5}$$

$$= \frac{68200}{5} = 13640$$

X	$X-\overline{X}$	$(X-\overline{X})^2$
11500	-2140	4579600
12400	-1240	1537600
15000	1360	1849600
14500	860	739600
14800	1160	1345600

$$\sum(X-\overline{X})^2 = 10052000, n = 5$$

$$S.D = S = \sqrt{\frac{\sum\left(X-\overline{X}\right)^2}{n}} = \sqrt{\frac{10052000}{5}}$$

$$= \sqrt{2010400} = 1417.88$$

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.

Solution:

(i)

X	$X-\overline{X}$	$\left(X-\overline{X}\right)^2$
12	2.5	6.25
6	-3.5	12.25
7	-2.5	6.25
3	-6.5	42.25
15	5.5	30.25
10	0.5	0.25
18	8.5	72.25
5	-4.5	20.25

$\sum x = 76$

$\sum\left(X-\overline{X}\right)^2 = 190, n = 8$

$$\overline{X} = \frac{76}{8} = 9.5$$

$$S.D = S = \sqrt{\frac{\sum\left(X-\overline{X}\right)^2}{n}} = \sqrt{\frac{190}{8}}$$

$$= \sqrt{23.75} = 4.87$$

(ii)

X	$X-\overline{X}$	$\left(X-\overline{X}\right)^2$
9	0	0
3	-6	36
8	-1	1
8	-1	1
9	0	0
8	-1	1
9	0	0
18	9	81

$\sum x = 72$

$\sum\left(X-\overline{X}\right)^2 = 120$

n = 8

$$\overline{X} = \frac{\sum x}{n} = \frac{72}{8} = 9$$

$$S.D = S = \sqrt{\frac{\sum\left(X-\overline{X}\right)^2}{n}} = \sqrt{\frac{120}{8}}$$

$$= \sqrt{15} = 3.87$$

b. Calculate variance for the data: 10, 8, 9, 7, 5, 12, 8, 6, 8, 2.

Solution:

(i)

X	$X-\overline{X}$	$\left(X-\overline{X}\right)^2$
10	2.5	6.25
8	0.5	25
9	1.5	2.25
7	- 0.5	.25
5	- 2.5	6.25
12	4.5	20.25
8	0.5	.25
6	-1.5	2.25
8	0.5	.25
2	- 5.5	30.25

$\sum x = 75$ $\sum\left(X-\overline{X}\right)^2 = 68.5$

n = 10

$$\overline{X} = \frac{\sum x}{n} = \frac{75}{10} = 7.5$$

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

$$= \frac{68.5}{10} = 6.85$$

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

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

Solution:

C.I	**f**	**Mid-point (x)**	**fx**	$X-\overline{X}$	$\left(X-\overline{X}\right)^2$	$f\left(X-\overline{X}\right)^2$
20 – 22	3	21	63	−6	36	108
23 – 25	6	24	144	−3	9	54
26 – 28	12	27	324	0	0	0
29 – 31	9	30	270	3	9	81
32 – 34	2	33	66	6	36	72
	$\sum f$=n=32		$\sum fx$ = 867		$\sum\left(X-\overline{X}\right)^2$=90	$\sum f\left(X-\overline{X}\right)^2$=315

$$\overline{X} = \frac{\sum fx}{n} = \frac{667}{32} = 27.093 = 27\,\text{approx}$$

$$S.D = S^2 = \sqrt{\frac{\sum f\left(X-\overline{X}\right)^2}{n}} = \sqrt{\frac{315}{32}}$$

$$= \sqrt{9.84375} = 3.137$$

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:

C.I	**Class Boundaries**	**f**
33 - 40	32.5 – 40.5	28
41 - 50	40.5 – 50.5	32
51 - 60	50.5 – 60.5	12
61 - 70	60.5 – 70.5	9
71 - 75	70.5 – 75.5	5

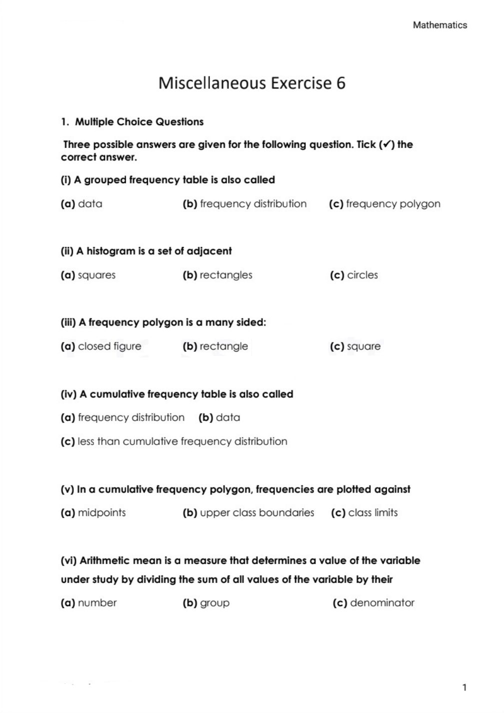
Here, $X_{max} = 75.5$

$X_{min} = 32.5$

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

$= 75.5 - 32.5$

$= 43$



Miscellaneous Exercise 6

1. Multiple Choice Questions

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

(i) A grouped frequency table is also called

(a) data **(b)** frequency distribution **(c)** frequency polygon

(ii) A histogram is a set of adjacent

(a) squares **(b)** rectangles **(c)** circles

(iii) A frequency polygon is a many sided:

(a) closed figure **(b)** rectangle **(c)** square

(iv) A cumulative frequency table is also called

(a) frequency distribution **(b)** data

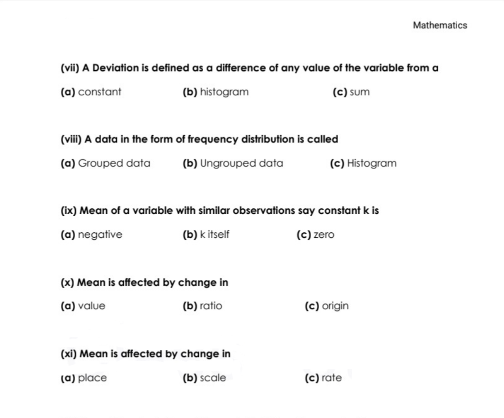
(c) less than cumulative frequency distribution

(v) In a cumulative frequency polygon, frequencies are plotted against

(a) midpoints **(b)** upper class boundaries **(c)** class limits

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

(a) number **(b)** group **(c)** denominator



(vii) A Deviation is defined as a difference of any value of the variable from a

(a) constant **(b)** histogram **(c)** sum

(viii) A data in the form of frequency distribution is called

(a) Grouped data **(b)** Ungrouped data **(c)** Histogram

(ix) Mean of a variable with similar observations say constant k is

(a) negative **(b)** k itself **(c)** zero

(x) Mean is affected by change in

(a) value **(b)** ratio **(c)** origin

(xi) Mean is affected by change in

(a) place **(b)** scale **(c)** rate

(xii) Sum of the deviations of the variable X from its mean is always

(a) zero **(b)** one **(c)** same

(xiii) The nth positive root of the product of the $x_1, x_2, x_3, \ldots, x_n$ observations is called

(a) Mode **(b)** Mean **(c)** Geometric mean

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

(a) Geometric mean **(b)** Median **(c)** Harmonic mean

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

(a) mode **(b)** median **(c)** harmonic mean

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

(a) median **(b)** mode **(c)** mean

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

(a) deciles **(b)** quartiles **(c)** percentiles

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

(a) average **(b)** dispersion **(c)** central tendency

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

(a) dispersion **(b)** central tendency **(c)** average

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

(a) average **(b)** range **(c)** quartiles

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

(a) variance **(b)** standard deviation **(c)** range

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

(a) harmonic mean **(b)** range **(c)** standard deviation

Answers:

(i)	b	**(ii)**	b	**(iii)**	a	**(iv)**	c	**(v)**	b
(vi)	a	**(vii)**	a	**(viii)**	a	**(ix)**	b	**(x)**	c
(xi)	b	**(xii)**	a	**(xiii)**	c	**(xiv)**	c	**(xv)**	a
(xvi)	a	**(xvii)**	b	**(xviii)**	b	**(xix)**	a	**(xx)**	b
(xxi)	a	**(xxii)**	c						

2. Write short answers of the following questions:

(i) Define class limits.

Ans. The minimum and the maximum values defined for a class or group are called class limits.

(ii) Define class mark.

Ans. The mid-point of a class interval is also called class mark. It is obtained by dividing the sum of upper and lower limit by 2.

(iii) What is cumulative frequency?

Ans. The total of frequency up to an upper-class limit or boundary is called the cumulative frequency.

(iv) Define a frequency distribution.

Ans. A table showing frequencies against the class intervals is called a frequency distribution table.

(v) What is a Histogram?

Ans. A histogram is a graph of adjacent rectangles constructed on XY — plane.

(vi) Name two measures of central tendency.

Ans. **(i)** Arithmetic mean.

(ii) Geometric mean

(iii) Harmonic mean

(vii) Define Arithmetic mean.

Ans. 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 the number of values.

(viii) Write three properties of Arithmetic mean.

Ans. (i) mean is affected by change in origin.

(ii) mean is affected by change in scale.

(iii) sum of deviations of the variable x from its mean is always zero.

(ix) Define Median.

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

(x) Define Mode.

Ans. The most frequent occurring observation in the data is mode.

(xi) What do you mean by Harmonic mean?

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

(xii) Define Geometric mean.

Ans. Geometric mean of a variable x is the nth positive root of the product of the $x_1, x_2, x_3, \ldots x_n$, observations.

(xiii) What is Range?

Ans. Range measures the extent of variation between two extreme observations of a data set.

(xiv) Define Standard deviation.

Ans. Standard deviation is the positive square root of mean of the squared deviations of x_i ($i = 1, 2, 3, \ldots, n$) observations from their arithmetic mean.

Summary

✓ Range is the difference between maximum and minimum observation.

✓ The minimum and the maximum values defined for a class or group are called class limits.

✓ The total of frequency up to an upper-class limit or boundary is called the cumulative frequency.

✓ A frequency distribution is a tabular arrangement classifying data into different groups.

✓ A Histogram is a graph of adjacent rectangles constructed on XY-plane, A cumulative frequency polygon or ogive is a graph of less than cumulative frequency distribution.

✓ Arithmetic mean is a measure that determines a value of the variable under study by dividing the sum of all values of the variable by their number.

✓ A Deviation is defined as 'a difference of any value of the variable from any constant. $D_i = x_i - A$.

✓ Geometric mean of a variable X is the nth positive root of the product of the x_1, x_2, x_3,....x_n observations.

✓ Harmonic mean refers to the value obtained by reciprocating the mean of the reciprocal of x_1, x_2, x_3,....x_n observations.

✓ Mode is defined as the most frequent occurring observation of the variable or data.

✓ Median is the measure which determines the middlemost observation in a data set.

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

✓ Range measures the extent of variation between two extreme observations of a data set.

✓ Variance is defined as the mean of the squared deviations of x_i, (i = 1, 2, ... n)

observations from their arithmetic mean.

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