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.

SOLVED 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	Talley marks	Frequency	Commutative 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, 3-7, 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	Talley marks	Frequency
20 – 24	𝍸 I	6
25 – 29	𝍸 𝍸	10
30 – 34	𝍸 𝍸 II	12
35 – 39	𝍸 IIII	9
40 – 44	III	3
		40

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	13	16 + 13 = 29	Less than 34.5	29
34.5 – 39.5	8	29 + 8 = 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—349, 550—649, ...).

Solution:

Frequency Distribution Table

Class Limits	Talley marks	Frequency
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
	Total =	30

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 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	||	2
4 – 5	|	1
6 – 7	卌 ||||	9
8 – 9	卌	5
10 – 11	卌 |	6
12 – 13	卌	5
14 – 15	|||	3
	Total =	31

(a) Find the most frequent load shedding hours.
6 – 7

(b) Find the least load shedding intervals.
4 – 5

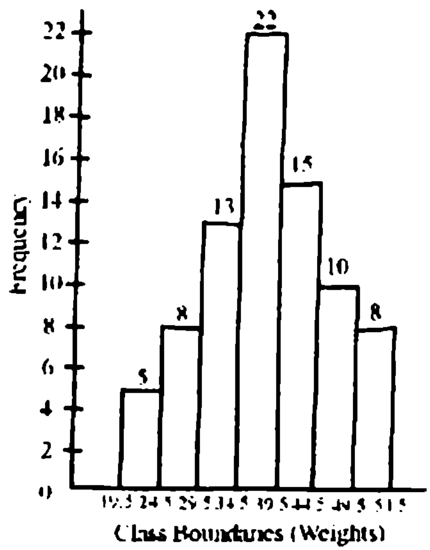
5. Construct a Histogram and frequency Polygon for the following data showing weights of students in kg.

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

Solution:

Class Boundaries	Frequency
19.5 - 24.5	5

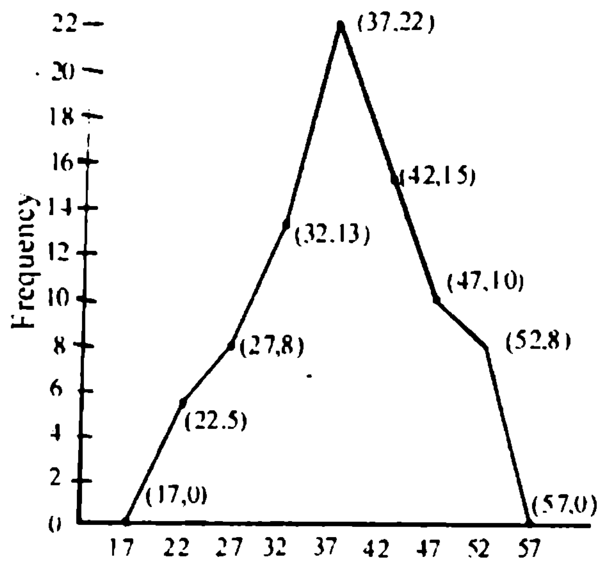
24.5 – 29.5	8
29.5 – 34.5	13
34.5 – 39.5	22
39.5 – 44.5	15
44.5 – 49.5	10
49.5 – 54.5	8



Construction of Frequency polygon

Class Limits	Mid points	Frequency
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

Polygon



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:

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

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:

$$\bar{X} = \frac{\Sigma X}{n} = \frac{\text{Sum of all observation}}{\text{No.of obsrvation}}$$

(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 $= (x_i - \bar{X})$ for i = 1, 2, ...n

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

The Formulae used under indirect methods are:

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

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 a constant multiple of the values of X.

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

$$\bar{X} = \frac{\Sigma fX}{\Sigma f}$$

Using Indirect method,

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

where 'X' 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-1:
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.

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

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

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

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 n^{th} positive root of the product of the x_1, x_2, x_3, ..., x_n observations. In symbols we write,

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

The above formula can also be written by using logarithm.
For Ungrouped data

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

For Grouped data

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

Harmonic Mean:

Harmonic mean refers to the value obtained by reciprocating the mean of the reciprocal of x_1, x_2, x_3, ..., x_n observations. In symbols, for ungrouped data,

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

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.

$$\bar{x}_w = \frac{w_1x_1 + w_2x_2 + ... + x_nx_n}{w_1 + w_2 + ... + w_n} = \frac{\Sigma wx}{\Sigma w}$$

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 days.

SOLVED 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.

$\bar{X} = \frac{\Sigma X}{n}$ (for ungrouped data) and $\bar{X} = \frac{\Sigma fX}{\Sigma f}$ (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. $G.M = (x_1 \times x_2 \times x_3 ....x_n)^{1/n}$

(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.

$H.M = \frac{n}{\Sigma \frac{1}{x}}$ (for ungrouped data) and $H.M = \frac{n}{\Sigma \frac{f}{x}}$ (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.