7
STATISTICAL METHODS

7.1 INTRODUCTION

In the previous lessons you have learnt about the certain aspects of collection and presentation of statistical data. With the help of systematic presentation we reduce the meaningless mass of statistical data into meaningful tables, bar charts, pie charts and frequency distributions. Once this is achieved the next stage is comparing one set of data with another set of data recorded in the form of table or frequency distribution. Sometimes we may need to compare one table with another table and one frequency distribution with another frequency distribution.

We need statistical tools or methods to make such comparisons. One set of statistical tool is found in ratio, rates and percentages. Another set of statistical tool is found in the averages or the measures of central tendency. In this lesson you will learn about these statistical methods.

7.2 OBJECTIVES

After going through this lesson you will be able to :

- explain the meaning of ratio with example;
- state the various forms of expression of a ratio;
- explain the distinction between rate and ratio;
- explain the concept of percentage;
- explain the meaning of a measure of central tendency;
- explain the concept of an average and arithmetic mean;
- calculate arithmetic mean;
- explain the concept of weighted arithmetic mean.

7.3 RATIOS

In our everyday life we make comparison between quantities. This comparison is often expressed in numerical terms. Suppose you are comparing prices of two brands of pens, one of which is priced at Rs. 6 per pen and the other at Rs. 2 per pen. Let us term these two brands as A and B. We can make comparison between the two brands in two ways : (i) Pen A is priced Rs. 4 higher than pen B or (ii) Pen A is priced 3 times as high as pen B. In the first case the comparison was made by subtraction (Rs.6 minus Rs. 2 = Rs.4). In the second case comparison was made by division (Rs. 6 divided by Rs.2 = Rs. 3). Comparing two numerical values by division is the ratio method of comparison.

(a) Meaning of Ratio :

Ratio is the relation between two quantities. The two quantities are called terms. A ratio is found by dividing the first term by the second term.

In our example of two brands of pens we had two terms, namely pen A priced at Rs. 6 and pen B priced at Rs. 2. Brand A priced at Rs. 6 is the first term because its price is intended to be compared with pen B. Brand B priced at Rs. 2 is the second term because it is in relation to the price of this pen that the brand A pen is intended to be compared. So **'what is compared'** is the first term and **'with which is to be compared'** is the second term.

After the two terms are determined we divide the first term by the second term to find the ratio.

$$\text{Ratio} = \frac{\text{Price of pen A}}{\text{Price of pen B}} = \frac{\text{Rs. 6}}{\text{Rs. 2}} = 3$$

The value of ratio is 3. It shows that pen A is priced 3 times as high as pen B.

(b) Forms of expression of ratio :

A ratio can be expressed in three forms :

(i)	In words	:	The ratio of Rs. 6 to Rs. 2
(ii)	In symbol	:	Rs. 6 : Rs. 2
(iii)	In fraction	:	$\frac{\text{Rs. 6}}{\text{Rs. 2}}$

In all the three forms of expression the value of ratio remains the same. We can state the ratio of pen A to the price of pen B in the three forms simultaneously as follows:

$$\text{Rs.6 to Rs.2} = \text{Rs.6 : Rs.2} = \frac{\text{Rs. 6}}{\text{Rs. 2}} = 3.$$

(c) Significance of the first and second terms :

In our example of pens the price of pen A is compared with price of pen B. The point of interest was that how many times is the price of pen A higher than the price of pen B. Here the price of pen A was the first term, i.e. one which is compared. The price of pen B was the second term, i.e. one with which it is compared.

Suppose our point of interest shifts to pen B. We now want to know that how many times is the price of pen B lower than the price of pen A. The price of pen B is now the first term and the price of pen A is the second term. The value of ratio now is:

$$\text{Rs.}2 : \text{Rs.}6 = \frac{\text{Rs. } 2}{\text{Rs. } 6} = \frac{1}{3}$$

The value of ratio is now 1/3. It shows that pen B is priced 1/3 (one third) of the price of pen A. Thus before attempting to calculate a ratio it is important to determine two things : (i) What is compared? (first term) and (ii) with which it is compared (second term). If we interchange the two terms the value of ratio is reversed. Ratio of price of pen A to the price of pen B is 3 while the ratio of price of pen B to the price of pen A is 1/3 (the reverse of 3).

Sometimes it is said that ratio between two quantities should be calculated only when both are expressed in the same units. It implies that both first term and the second term must be expressed in the same unit of measurement like Rs. : Rs., kilograms : kilograms, meters : meters, etc. For all practical purposes this rule is not observed in statistical calculations.

Whether both the terms are expressed in same units or not, for a ratio to be meaningful it is necessary that both the terms must be related to each other. For example, take the value of the ratio of national income to population in a particular year. Here income is measured in rupees and population in numbers. Although the two terms are measured in different units, the value of the ratio gives us per capita income which is a meaningful relation.

Let us take another example. Suppose we calculate ratio between steel production and food production both expressed in quintals. The ratio tells us that how many times is steel production as compared to food production. The ratio so derived is not directly meaningful even though both the terms are measured in the same units of measurement.

To conclude, in order to calculate a meaningful ratio two things are necessary :

i) Determine 'what is compared' (first term) and 'with which it is to be compared' (second term).

ii) First term and the second term are related to each other.

d) Examples of certain ratios used in Economics :

i) Ratio of national income to population :

$$\frac{\text{National Income}}{\text{Population}} = \text{Per Capita Income}$$

ii) Input - output ratio :

$$\frac{\text{Input}}{\text{Output}} = \text{Input per unit of output}$$

iii) Ratio of consumption to income :

$$\frac{\text{Consumption Expenditure}}{\text{Income}} = \text{Propensity to consume}$$

iv) Ratio of saving to income :

$$\frac{\text{Saving}}{\text{Income}} = \text{Propensity to save}$$

v) Ratio of population to land area :

$$\frac{\text{Population}}{\text{Land Area}} = \text{Density to population}$$

POINTS TO REMEMBER

- Ratio is a relation between two quantities called terms.
- To calculate a ratio it is necessary to determine (a) What is compared (first term) and (b) With which it is to be compared (second term).
- Ratio is calculated by dividing the first term by the second term.
- A ratio can be expressed in words, in symbol and in fraction.
- If the first term and the second term are interchanged, the value of the ratio is reversed.
- For a ratio to be meaningful it is necessary that the two terms must be related to each other.

INTEXT QUESTIONS 7.1

Fill in the blanks with appropriate word from the brackets :

(i) Ratio is calculated by...........(dividing, multiplying) the..........(first, second) term by the.........(first, second) term.

(ii) In a ratio the term what is compared is called .... (first, second) term.

(iii) In a ratio the term with which it is to be compared is called..........(first, second) term.

(iv) When we say 'ratio of the first term to the second term' we describe the ratio in ....(words, symbol,fraction).

(v) When we say 'first term : second term' we describe the ratio in ....(words, symbol, fraction).

(vi) Monthly incomes of A and B are respectively Rs. 2000 and Rs. 1000. Answer the following:

i) Income of A is ....(2, 1/2) times that of B.
ii) Income of B is........(2, 1/2) times that of A.

(vii) Ratio 70 : 80 is..........(the same, not the same) as ratio of 80 : 70.

7.4 RATES

(a) Meaning

In economics we often talk in terms of rates like rate of economic growth, rate of growth of population, birth rate, death rate, agricultural yield rates etc. When we see that how these rates are calculated, we will find that the process of calculation is either the same or nearly the same. Let us take some examples to clarify what we have said. For example, take the rate of yield per hectare of a crop.

$$\text{Rate of yield (in kg.) per hectare of a crop} = \frac{\text{Total production of crop (kgs.)}}{\text{Total area (hectares) under crop}}$$

In the above example we find that process of calculation is the same as that in ratio. Thus rate and ratio are the same in this example. Here yield rate is nothing but the ratio of production to area during a particular year. Rate is thus a ratio between two magnitudes shown over a period of time.

(b) Rate vs Ratio

Although the method of calculation of rate and ratio is the same, the meaning conveyed by rate is somewhat different. Rate is a ratio between two magnitudes shown over a period of time. In our example above, the yield rate per hectare is the ratio of 'production of crop' to 'total area under the crop' during the year.

Rate is different from ratio in another respect. Ratios are generally expressed per unit, while rate can be expressed per unit, per 100 units, even higher. In calculating rates any arbitrary figure can be taken as 'base' depending upon which base is more suitable for comparison. In economic studies we find that 100 is the most common base adopted.

c) Need for arbitrary base in rate

Let us now explain that why is there a need to adopt an arbitrary base different from per unit base. The need for a base like 100 or 1000 or 10,000 arises because 'the value of ratio per unit' sometime is so small that it fails to convey the significance of the rate or ratio. We take an example to explain this point.

Suppose there is a town with population of 100,000. Further suppose that total number of births during the year 1995 in the town is 2340. Let us calculate birth rate per unit of population.

$$\text{Birth rate per unit of population} = \frac{\text{Number of births during the year 1995}}{\text{Population}}$$

$$= \frac{2340}{100{,}000} = .0234$$

The above calculation reveals that there is .0234 birth per unit of population. Here the value of ratio is so small that it may make the comparison difficult. Suppose during the next year, i.e. 1996 total number of births are 2520.

$$\text{Birth rate per unit of population is} = \frac{2520}{100{,}000} = .0252$$

To say that birth rate per unit of population in the year 1996 is .0252 as compared to .0234 in the year 1995. It is very difficult to realise from these figures that how higher is the birth rate in 1996 as compared to 1995. There is, therefore, a need to raise the base.

Suppose we raise the base to 100. The two birth rates in the year 1995 and 1996 would be:

$$\text{Birth rate per 100 in 1995} = \frac{\text{Births during 1995}}{\text{Population}} \times 100$$

$$= \frac{2340}{100{,}000} \times 100 = 2.34$$

$$\text{Birth rate per 100 in 1996} = \frac{2520}{100{,}000} \times 100 = 2.52$$

Comparison between the two birth rates is now more convenient. Still there are some difficulties in comparisons. First the two rates are so small that one may fail to realise the difference and may treat the difference as insignificant. Second, it is better to avoid population figures in fractions. To avoid fractions we have to round them. Rounding a fraction when it is small may significantly affect the rate. So to avoid fractions also it is necessary to raise the base. Let us raise the base to 1000. The two birth rates are now as follows :

$$\text{Birth rate per 1000 in 1995} = \frac{2340}{100{,}000} \times 1000 = 23.4$$

$$\text{Birth rate per 1000 in 1996} = \frac{2520}{100{,}000} \times 1000 = 25.2$$

The difference between the two birth rates is now more clearly visible. Secondly the fractions can be avoided by rounding 23.4 to 23 and 25.2 to 25 without significantly affecting the absolute difference between the two rates. This is why in practical statistics birth rates and death rates are calculated per 1000 of population.

7.5 PERCENTAGES

By now it must be clear to you that percentage is a type of rate or ratio with base 100. Every ratio per unit when multiplied by 100 is converted into percentage. It is calculated as follows:

$$\text{Percentage} = \frac{\text{First Term}}{\text{Second Term}} \times 100$$

Let us come back to the example of pen A and pen B respectively priced Rs. 6 and Rs. 2. If the period of interest is that how much per cent is the price of A in relation to the price of pen B, we calculate the percentage

$$\frac{\text{Price of pen A}}{\text{Price of pen B}} \times 100 = \frac{\text{Rs. } 6}{\text{Rs. } 2} \times 100 = 300\%$$

According to the calculation price of pen A is 300 per cent of the price of pen B. The main step in the calculation of percentage is the same as in ratio plus an additional step. The additional step is to multiply the ratio by 100. The reason behind taking the additional step of raising the base from 'per unit' to 'per 100' is already discussed in the paragraphs on rates.

Some of the important percentages used in economics are : (1) percentage rate of economic growth (2) percentage rate of interest (3) percentage rate of tax (4) percentage rate of capital formation etc.

POINTS TO REMEMBER

- Method of calculation of rate is similar to that of ratio.
- Rate is sometimes distinguished from ratio in two respects :
  - (a) Rate is the ratio between two magnitudes shown over a period of time.
  - (b) Rate can be expressed, besides per unit, per 100, per 1,000, per 10,000, per lakh and even higher.
- Arbitrary higher base for calculation of rate is chosen when
  - (a) value of ratio is very small and
  - (b) there is a need to avoid fractions in comparisons.
- Percentage is a type of ratio or rate with base 100.

INTEXT QUESTIONS 7.2

1. State whether the following statements are true or false :

(i) Rate is calculated in the similar way as a ratio.
(ii) Rate is expressed per unit only.
(iii) Percentage is quite different from rate or ratio.
(iv) Percentage is calculated with base 'per 1000'.

2. Answer the following :

(i) Per capita income of the year 1995 and 1996 is respectively Rs.1000 and Rs.1200 (a) Find out how much per cent is per capita income of the year 1996 as compared to 1995. (b) Find the percentage increase in the year 1996.

(ii) In a city of population with 10 lakhs, the number of births and deaths during the year were 11,500 and 10,500 respectively.

Find out the following :

a)	Birth rate per unit	d)	Death rate per unit
b)	Birth rate per 100	e)	Death rate per hundred
c)	Birth rate per 1000	f)	Death rate per thousand

7.6 MEAN

(a) Meaning of measures of central tendency

You might have observed in the previous lessons that statistical data have a tendency to concentrate around certain values. Such clustering of items or values in the central part of the distribution is an important characteristic of frequency distribution and is known as the central tendency. A measure of central tendency means a value where the concentration of the items or values is found to be the greatest. This is the reason why the value of a central tendency of the given data is regarded to be the most representative of the series or observations.

In a school, a teacher may find the average marks in a particular subject, say, Economics for different sections of a class. This will help him to compare its various sections according to their performance. Further, note that an average is a typical value because it not only sums up the data but also describes it. So an average can be used to describe the complete set of figures and information contained in a series. For example, it may not be possible to remember the income of every person in Delhi. However, their average income can be easily remembered. Similarly incomes of residents of India are difficult to be compared with those of Japan. But their average incomes (known as per capita income) can be compared.

An average can be obtained by using five different measures of central tendency namely : (1) Arithmetic mean (2) Median, (3) Mode, (4) Geometric mean and (5) Harmonic mean.

Different measures of central tendency are found useful in different situations. In this lesson we shall study only about arithmetic mean ($\overline{X}$). Other measures of central tendency are beyond the scope of the syllabus.

(b) The Concept of an Average

The concept of an average is very important in our daily life. In our everyday speech we use this concept so often that its importance need not be overemphasized. Quite often you may have heard remarks like this :

Principal	:	What type of student is Ms. Manju ?
Teacher	:	Sir, she is an average student.
One cricket fan	:	What was the average score of Gavaskar in the test matches?
Other cricket fan	:	Well, I think it was about 49.
Mr.X	:	Does a worker get paid more in city A than in city B?
Mr.Y	:	No, on an average a worker in city A gets Rs.900 per month whereas in city B average wage is Rs.1,000.
Father	:	Is the performance of your school better than that of your sister's school?
Son	:	Yes, Daddy, in 10+2 examination, the average marks of students of our school is 60% as compared to 55% that of my sister's school.

Ramesh	:	What type of a man is your neighbour?
John	:	Oh, I like him he is a typical Indian.

Thus the concept of an average shows something representative, typical of anything which is most commonly found. Your friend is an average student, shows that he is typical of the group he belongs to. The average wage in a particular industrial area might be the most common wage which is being received by most of the workers.

Average score of Gavaskar means that total of runs scored by him in all the innings he had played divided by the total number of innings. Similarly, per capita income is also an average. It can be computed by dividing national income of a country by its population, i.e.

$$\text{Per Capita Income} = \frac{\text{Total National Income}}{\text{Total population}}$$

Besides these, there are various other types of averages which describe the complete figures or information contained in a series. No doubt, therefore, the average is a representative value of a series.

7.7 THE ARITHMETIC MEAN ($\overline{X}$)

What is generally called an 'average' by means of use in statistics, is termed as the arithmetic mean. It is one of the widely used statistical measures. It is found or calculated by taking all the items of a series into account. For example, if you have secured, out of 10 of maximum marks in each subject, 5, 6, 7 and 8 marks in four different subjects, your average marks will be :

$$\frac{5+6+7+8}{4} = \frac{26}{4} = 6.5$$

The above calculations can be expressed in symbols. There are four items 5, 6, 7 and 8. These four items make a series. The symbols used for items are X_1 for the first item, X_2 for the second item, X_3 for the third item and X_4 for the fourth item. Thus 5, 6, 7 and 8 are respectively termed as X_1, X_2, X_3 and X_4. The number of these items is 4 which is termed as 'N'. The sum total of values of these items (i.e. 5+6+7+8 = 26) is termed as ΣX. The mean is termed as $\overline{X}$. In this way the above calculations can be expressed as follows :

$$\overline{X} = \frac{X_1 + X_2 + X_3 + X_4}{N} = \frac{\Sigma X}{N}$$

$$= \frac{5+6+7+8}{4} = \frac{26}{4} = 6.5$$

Of the above symbol $\overline{X}$ is read as 'X-bar' and ΣX as 'sigma x'.

7.8 CALCULATION OF ARITHMETIC MEAN

Data can be organised in many ways like simple array, frequency array and frequency distribution. Since the method of presentation of data is different in each case the method of calculation of arithmetic mean also becomes somewhat different according to the change in the form of presentation. Accordingly we will explain the calculation of arithmetic mean in following methods of presentation.

a) Simple array
b) Frequency array
c) Frequency distribution

(a) Calculation of arithmetic mean in simple array

A simple array is a series of individual items also called individual series. There are two methods of calculation in this case : (i) Direct method and (ii) Indirect method.

(i) Direct method :

The following table gives the marks obtained by 7 students in the subject of statistics :

Table 7.1

Marks
70 (X_1)
70 (X_2)
75 (X_3)
75 (X_4)
85 (X_5)
90 (X_6)
95 (X_7)

$$\text{Arithmetic Mean} = \frac{\text{Sum of all the observations in the series}}{\text{Number of observations}}$$

$$\text{or} \quad \overline{X} = \frac{X_1 + X_2 \ldots\ldots\ldots + X_7}{N}$$

$$= \frac{70 + 70 + 75 + 75 + 85 + 90 + 95}{7} = 80$$

We find that arithmetic mean is 80

POINTS TO REMEMBER

- An average is a value which is typical or representative of a set of data.
- Averages are also called measures of central tendency, since they tend to lie centrally, within a set of data arranged according to magnitude.
- The arithmetic mean or mean of a set of N numbers $X_1, X_2 ... X_n$ is denoted by $\overline{X}$ (read X bar) and is defined as

$$\overline{X} = \frac{X_1 + X_2 + X_3 \ldots\ldots\ldots + X_n}{N} = \frac{\Sigma X}{N}$$

INTEXT QUESTIONS 7.3

1. State whether the following statements are true or false :
   (i) An average is representative of a set of data.
   (ii) The average is used only for calculating average marks of a student.
   (iii) Per capita income is an average.
   (iv) National income of India is an average.

2. Fill in the blanks :
   (i) Arithmetic mean or $\overline{X} = \frac{\text{Sum of all the observations in the series}}{\ldots\ldots\ldots\ldots\ldots}$
   (ii) The sign 'Σ' is read as............in statistics and it means............
   (iii) The marks of a student in 5 subjects were 10, 12, 12, 14, 11. The arithmetic mean of these marks is..............
   (iv) The monthly salaries of 4 men were 800, 500, 1000, 1200. The arithmetic mean of their salaries is..............

(ii) Indirect Method :

If the number of observations is small it is convenient to calculate mean by direct method. If the number of observations is large the direct method becomes inconvenient. In such a case it is better to use the indirect method.

Let us take the same example on marks in the subject of statistics obtained by seven students as in table 7.1. Out of these take any score which seems to you to be closer to the mean. Generally, it is convenient to take some middle value. The value so guessed is called **'assumed mean'**. The symbol for assumed mean is 'A'. The main steps required to be taken in the indirect method are :

(1) Take some observation as assumed mean. For example, we take 75 as assumed mean.

(2) Find deviation of each observation from the assumed mean as is shown in table 7.2. The symbol for such deviation is dx which is equal to (X-A).

(3) Find the sum of these deviations. The symbol is Σdx.

(4) Calculate arithmetic mean from the following method.

$$\overline{X} = A + \frac{\Sigma dx}{N}$$

Table 7.2

Marks (X)	Deviations (X-A) A= 75	= dx
70	70-75	– 5
70	70-75	– 5
75	75-75	0
75	75-75	0
85	85-75	+10
90	90-75	+15
95	95-75	+20
N = 7		$\Sigma dx = +35$

Substituting the value of A, N and Σdx in the above formula, we get :

$$\overline{X} = A + \frac{\Sigma dx}{N} = 75 + \frac{35}{7} = 75 + 5 = 80$$

In this way we find that arithmetic mean is 80.

POINTS TO REMEMBER

- In a set of data where number of observations is large, the assumed mean method is used to simplify or cut sort the calculations.
- The answer obtained by both, the long method and short cut method is the same.
- $\overline{X} = A + \frac{\Sigma dx}{N}$ is the case of ungrouped data when found by the assumed mean method.

INTEXT QUESTIONS 7.4

Fill in the blanks :

1. Taking '10' as assumed mean in a set of numbers 12, 13. 10. 15 and 17 the arithmetic mean is................

2. $\overline{X} = A + \frac{\Sigma dx}{N}$, where

   a) $\overline{X}$ stands for..........

   b) A stands for...........

   c) Σdx stands for.........

   d) N stands for..........

(c) Calculation of Arithmetic Mean in Frequency array :

Like in case of simple array there are two methods of calculation in case of frequency array also.

(i) Direct Method :

The main steps in this method are :

(1) Multiply each value of X by its frequency (f). By doing so we obtain $fx = (X \times f)$.

(2) Take sum of all values of fx. We get Σfx.

(3) Divide Σfx by the number of observations. We get the number of observations by adding all values of 'f'. In other words we get Σf. By dividing Σfx by Σf we get arithmetic mean ($\overline{X}$). So,

$$\overline{X} = \frac{f_1 x_1 + f_2 x_2 \ldots\ldots + f_n x_n}{f_1 + f_2 \ldots\ldots + fn} = \frac{\sum fx}{\sum f}$$

Example :

The table 7.3 gives ungrouped frequency distribution of marks obtained by 20 students in the subject of statistics.

Table 7.3

Marks of students (x)	Number of students (f)	Product (fx)
4	6	24(6×4)
5	3	15(3×5)
6	2	12(2×6)
7	3	21(3×7)
8	4	32(4×8)
9	2	18(2×9)
	$\sum f = 20$	$\sum fx = 122$

Using the formula for finding arithmetic average, we get

$$\overline{X} = \frac{\Sigma fx}{\Sigma f} = \frac{122}{20} = 6.1$$

So the average of marks in statistics in a class of 20 students is 6.1.

(ii) Indirect Method :

Step 1 - Select a suitable assumed average (A).
Step 2 - Find deviations of each observation from A to get dx = (X-A).
Step 3 - Multiply each dx by its respective frequency (f) to get fdx.
Step 4 - Add various fdx to get Σfdx.
Step 5 - Use the formula ;

$$\overline{X} = A + \frac{\Sigma fdx}{\Sigma f}$$

Example :

Use information given in table 7.3 to find arithmetic mean by indirect method.

Solution :

As explained above we prepare the following table :

Table 7.4

Marks in Statistics (X)	Frequency (f)	dx=(X-A) where A=6	fdx (f × dx)
4	6	–2(4-6)	–12
5	3	–1(5-6)	–3
6	2	0(6-6)	0
7	3	1(7-6)	3
8	4	2(8-6)	8
9	2	3(9-6)	6
	$\Sigma f = 20$		$\Sigma fdx = +2$

Using the formula

$$\overline{X} = A + \frac{\Sigma fdx}{\Sigma f} = 6 + \frac{2}{20} = 6 + \frac{1}{10} = 6 + 0.1 = 6.1$$

Note that we get the same value of $\overline{X}$ by indirect method as we got by the direct method.

POINTS TO REMEMBER

- In the case of ungrouped frequency distribution the formula for calculating $\overline{X}$ (arithmetic mean) by direct method

  is $\overline{X} = \frac{\Sigma fx}{\Sigma f}$ (Here $\overline{X}$ stands for arithmetic mean; f stands for frequency of observation; x stands for the variable.)

- In the case of ungrouped series the formula for calculating arithmetic mean ($\overline{X}$) by indirect or short cut method is ;

  $\overline{X} = A + \frac{\Sigma fdx}{\Sigma f}$

INTEXT QUESTIONS 7.5

Fill in the blanks :

1. $\overline{X} = \frac{f_1x_1 + f_2x_2 + f_3x_3 \ldots\ldots\ldots + f_nx_n}{\ldots\ldots\ldots} = \frac{\Sigma fx}{\Sigma f}$

2. $\overline{X} = A + \frac{\Sigma f\,dx}{\ldots\ldots}$

3. If 5, 6, 7 and 10 occur with frequencies 2, 3, 4 and 1 respectively, the arithmetic mean is

   $\overline{X} = \frac{(5)(2) + (6)(3) + (7)(4) + (10)(1)}{2 + 3 + 4 + 1} = \ldots\ldots\ldots\ldots\ldots\ldots$

4. Out of 100 numbers 20 were 5's; 30 were 6's; 40 were 7's and 10 were 8's. The arithmetic mean of numbers

   $\overline{X} = \frac{\Sigma fx}{\Sigma f}$ or $\frac{\Sigma fx}{N} = \frac{(20)(5) + (30)(6) + (40)(7) + (10)(8)}{100} = \ldots\ldots\ldots\ldots$

(c) Calculation of Arithmetic Mean in Frequency Distribution

A frequency distribution classifies the data into groups. In such case also there are two methods : (i) Direct method and (ii) Indirect method.

(i) Direct Method :

There is an additional step in this method in comparison to frequency array case. To get the value of 'fx' we multiply x with f. X, in case of frequency distribution, is a group and not an individual item. It is not possible to multiply the group (say 1-3) with f. So we take first the mid value of the group and then multiply with 'f'. The mid value is obtained by taking simple average of lower (L_1) and upper (L_2) limits of the group. For example :

Mid value of group (1-3) $= \frac{1+3}{2} = \frac{4}{2} = 2$

Now this mid value of group is taken as x. All other steps are the same as in case of frequency array. The main steps are :

(1) Take the mid value of each group as the value of x.
(2) Multiply x with f to obtain fx.
(3) Take the sum of fx to obtain Σfx.
(4) Divide Σfx by Σf to get $\overline{X}$.

Example :

Table 7.5

Marks (out of 10)	Mid value (x)	Students (f)	Product (fx)
1-3	2	5	10
3-5	4	3	12
5-7	6	5	30
7-9	8	7	56
		$\Sigma f=20$	$\Sigma fx=108$

Now let us substitute its values in the formula,

$$\overline{X} = \frac{\Sigma fx}{\Sigma f} = \frac{108}{20} = 5.4$$

The mean marks in the subject of history are 5.4

(ii) Indirect Method :

There are two versions of indirect method. In one version we take assumed mean as an additional step. In another version we take a further additional step in form of step deviation. The two versions are explained below :

(i) Based on assumed mean

The main steps are :

(1) Take mid value (x) of each group.
(2) Take some value of x as assumed mean (A).
(3) Deduct A from x to get dx.
(4) Multiply dx by f to get fdx.
(5) Take sum of all value of fdx to obtain Σfdx.
(6) Divide Σfdx by Σf to get $\overline{X}$.

Example :

Table 7.6

Marks out of 10 in History	Mid value (x)	No. of Students (f)	dx=(X-A) Where A=4	fdx
1-3	2	5	-2	-10
3-5	4	3	0	0
5-7	6	5	2	10
7-9	8	7	4	28
		$\Sigma f=20$		$\Sigma fdx=28$

$$\overline{X} = A + \frac{\Sigma f\,dx}{\Sigma f} = 4 + \frac{28}{20} = 4 + 1.4 = 5.4$$

So mean marks obtained in History are 5.4 (same as obtained by the direct method).

You must have observed that the figures used in the indirect method (short-cut method) are much simpler and so computation is much easier.

(ii) Step Deviation Method :

This is another version of indirect method. In this we take an additional step to make the calculation easy. The step is called **step deviation.** If the value of x is high, value of dx =(X-A) is also likely to be high. To make the calculation simple we first find a common figure by which all the values of dx can be divided. It will reduce the values of dx and make further calculation easy. This common factor by which the values of dx are divided is termed as 'i'. At a latter stage value of dx are again multiplied by this common factor so that final result of arithmetic mean is not affected. The main steps are :

(1) Take mid value of group (X).

(2) Take some assumed mean (A).

(3) Subtract A from X to get dx.

(4) Take a common factor (i) from amongst the values of dx and divide each value of dx by this common factor to get dx$i$ ($= \frac{dx}{i}$).

(5) Multiply dx$i$ by f to get fdx$i$.

(6) Take the sum of fdx$i$ to get $\Sigma f\, dxi$

(7) Apply the following formula to obtain arithmetic mean

$$\overline{X} = A + \frac{\Sigma f\, dxi}{\Sigma f} \times i$$

Example :

Table 7.7

Marks in History out of 10	Mid-value (X)	Number of Students (f)	dx = (X-A) A=4	Step-deviation $dxi = \frac{dx}{i}$ i=2	fdx$i$
1-3	2	5	–2	–1	–5
3-5	4	3	0	0	0
5-7	6	5	2	1	5
7-9	8	7	4	2	14
		$\Sigma f = 20$			$\Sigma fdxi = +14$

Using the formula for step-deviation, we get

$$\overline{X} = A + \frac{\Sigma f\, dxi}{\Sigma f} \times i = 4 + \frac{14}{20} \times 2 = 4 + 1.4 = 5.4$$

Hence $\overline{X}$= 5.4 which is the same value as we obtained in table 7.6.

POINTS TO REMEMBER

- While calculating arithmetic mean ($\overline{X}$) in a grouped frequency distribution, we calculate mid value of each class and treat them as 'X' values.
- By direct method $\overline{X} = \frac{\Sigma fx}{\Sigma f}$
- By short-cut or indirect method $\overline{X} = A + \frac{\Sigma f dx}{\Sigma f}$
- By step-deviation method $\overline{X} = A + \frac{\Sigma f dxi}{\Sigma f} \times i$

INTEXT QUESTIONS 7.6

1. Fill in the blanks :
   The mid-value of limits (i) 10-20 is...........
   (ii) 15-20 is...........
   (iii) 26-30 is...........

2. Complete the following table.

Marks	No. of Students (f)	(mid-point) (x)	fx
5-10	2	....	....
10-15	3	....	....
15-20	5	....	....
		Σf = ...	Σfx=......

3. Give the formula for calculating arithmetic mean and insert the values from table of question No. 2 above.

$$\overline{X} = \frac{\Sigma \ldots\ldots\ldots}{\Sigma \ldots\ldots\ldots} = \ldots\ldots\ldots\ldots = \ldots\ldots\ldots\ldots$$

7.9 WEIGHTED ARITHMETIC MEAN

So far while calculating the arithmetic mean we have given equal emphasis to each item in the series. This equal emphasis may be quite misleading if individual items have different importance, as in the following example :

Example :
Supposing a shopkeeper sells, say, brand A pens for Rs.5, brand B pens for Rs.15 and brand C pens for Rs. 25 each.

Here, N = 3 and $X_1 = 5$, $X_2 = 15$, $X_3 = 25$

$$\overline{X} = \frac{\Sigma x}{N} = \frac{45}{3} = \text{Rs. } 15.00$$

In this example we find that some pens are very cheap and others costly. In addition, the shopkeeper may sell different quantities of different brands. For example, he may sell 100 pens of brand A, 40 pens of brand B and only 20 pens of brand C. As such different brands have different relative importance. In statistics this importance is known as weights (w).

It may be noted that quantities in numbers or kilograms or any other unit, is not the only basis of assigning weights. As you will study in lesson No. 9 we can use other method also namely, value weight. It combines both quantity as well as price. That is price (p) multiplied by quantity (q) gives us value (pq).

In order to find out weighted arithmetic mean the following steps should be taken.

First, multiply each quantity (x) by its weight (w) to obtain different products (wx), i.e.

$$w_1x_1,\ w_2x_2,\ w_3x_3 \ldots\ldots w_nx_n$$

Second, all these products are added to get Σwx, i.e.

$$\Sigma wx = w_1x_1 + w_2x_2 + \ldots\ldots + w_nx_n$$

Third, this sum of products (Σwx) is then divided by the sum of the weights (Σw) to obtain the required weighted arithmetic mean. Thus,

$$\text{Weighted arithmetic mean } \overline{X}w = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots\ldots + w_n x_n}{w_1 + w_2 + w_3 + \ldots\ldots + w_n}$$

$$= \frac{\sum wx}{\sum w}$$

Let us now apply this result to our example of three brands of pens.

Table 7.8

Brand of pens	Price (x)	Quantity sold (w)	wx
A	5	100	500
B	15	40	600
C	25	20	500
		$\sum w = 160$	$\Sigma wx = 1600$

Using the formula for weighted arithmetic mean we get,

$$\overline{X}_w = \frac{\Sigma wx}{\Sigma w} = \frac{1600}{160} = 10$$

POINTS TO REMEMBER

- Sometimes we associate with the number $x_1, x_2 .... x_n$ certain weighting factors or weights $w_1, w_2 ...... w_n$ depending on the significance or importance attached to the numbers. In this case we calculate what is known as weighted mean ($\overline{X}_w$) by the formula,

$$\overline{X}w = \frac{w_1 x_1 + w_2 x_2 + w_3 x_3 + ... + w_n x_n}{w_1 + w_2 + w_3 + ... + w_n}$$

$$= \frac{\sum wx}{\sum w}$$

INTEXT QUESTIONS 7.7

A student's final marks in Mathematics, English, Home Science and Economics are respectively 82, 86, 90 and 70. If respective credits received for these courses are 3, 5, 3 and 1, complete the following :

1. $\overline{X}_w = \frac{\Sigma wx}{\Sigma w} = \frac{(..)(82) + (5)(86) + (3)(90) + (1)(70)}{3 + (...) + (...) + 1}$

2. $\frac{246 + 430 + 270 + 70}{.........}$

3. $\frac{.........}{.........} = 84.66$

7.10 ADVANTAGES AND DISADVANTAGES OF ARITHMETIC MEAN

(a) Advantages :

Arithmetic mean is a most commonly used measure of central tendency. This is because of certain merits possessed by it as compared to other measures. For example,

1. It is easy to understand and calculate, that is, it is the sum of the observations divided by their number.

$$\overline{X} = \frac{\Sigma x}{N}$$

2. It is based on all the observations of the series. It has full coverage. If any part of the series is not known, it cannot be calculated.

3. It is subject to algebraic treatment. That is, various mathematical operations can be carried on it. For example :

a) If arithmetic mean $\overline{X}_1$ and $\overline{X}_2$ and number of observations N_1 and N_2 are known for two series, it is possible to obtain the combined arithmetic mean ($\overline{X}_{1,2}$).

$$\overline{X}_{1,2} = \frac{N_1\overline{X}_1 + N_2\overline{X}_2}{N_1 + N_2}$$

b) If each observation of a series are added, subtracted, multiplied or divided by a common factor, say α, the mean is also added, subtracted, multiplied or divided by the same factor. Let us illustrate one of the four operations – addition. Let the series be 4, 6, 8, 10 and 12. Their mean is :

$$\overline{X} = \frac{4 + 6 + 8 + 10 + 12}{5} = \frac{40}{5} = 8$$

Now increase every observation by 5, we shall find that mean also increases by 5 to 13. New series will be (4+5),(6+5),(8+5),(10+5) and (12+5) or 9,11,13,15 and 17. Their mean is

$$\frac{9 + 11 + 13 + 15 + 17}{5} = \frac{65}{5} = 13 \text{ i.e. } (8+5)$$

(b) Disadvantages :

Although arithmetic mean possesses advantages unmatched by other measures of central tendency yet it suffers from mainly three types of defects.

1. It is greatly affected by the existence of extreme values in the series–extremely large and extremely small. In such cases arithmetic mean is not a typical or representative value. These values artificially pull down or push up the mean value. As for example five students of a class secured 40,45,50,55 and 60 marks out of 100. The average of marks comes to $\frac{40 + 45 + 50 + 55 + 60}{5} = \frac{250}{5} = 50$. Suppose two new students join the class and secure one mark each, you will find that the average of marks of a class of 7 students is pulled down to $\frac{40 + 45 + 50 + 55 + 60 + 1 + 1}{7} = \frac{252}{7} = 36$.

2. Arithmetic mean is also not representative in case,

(a) Classes are unequal as for example, 0-5, 5-10, 10-20, 20-40, 40-50.
(b) Open-end classes such as less than 5, 5-10 ...... 30-35, 35 and above.

POINTS TO REMEMBER

- The arithmetic mean is the most commonly used and easily understood average.
- Its computation is easy and simple (It is easy to calculate).
- It may be treated algebraically (or it is subject to algebraic treatment).
- The arithmetic mean may be greatly affected by extreme values.
- The arithmetic mean may not be truly representative in the case of open-end class intervals.

INTEXT QUESTIONS 7.8

State whether the following statements are true or false :

1. Arithmetic mean is not based on all the observations of the series.

2. It is possible to obtain combined arithmetic mean of two or more series.

3. Arithmetic mean is not affected by extreme values in a series.

4. In the case of an 'open-end' distribution the arithmetic mean may not be truly representative.

WHAT YOU HAVE LEARNT

- Ratios, rates and percentages are simple statistical tools useful in making comparisons.
- Ratio is the relation between two quantities called terms. It is calculated by dividing the first term by the second term.
- Method of calculation of rate and percentage is similar to that of ratio.
- Sometimes rate is distinguished from ratios on two counts; (a) it shows relation between two magnitudes over a period of time and (b) it can be expressed to any arbitrary base like per 100, per 1000, per 10,000 per lakh and even higher.
- Percentage is a type of ratio or rate with base 100 and is most widely used in economic studies.
- An average is a value which is typical or representative of a set of data.
- Averages are also called measures of central tendency, since they tend to be located centrally within a set of data arranged according to magnitude.
- The most common type of measures of 'Central tendency' is the arithmetic mean.
- The arithmetic mean is the sum of all the items divided by their number.
- A weighted arithmetic mean is an average in which the values to be averaged are given varying importance.

TERMINAL EXERCISE

1. Explain the meaning of 'ratio' with the help of an example.
2. Explain the significance of the first and second terms in calculation of ratio.
3. Explain the meaning of 'rate' by giving an example.
4. In what respects 'rate' is sometimes distinguished from 'ratio'?
5. Show with the help of an example the need to adopt a higher base rather than 'per unit' base in calculation of a 'rate'.
6. Show with the help of an example how is percentage calculated? How it is distinguished from rate?
7. What is arithmetic mean ? Explain the method of calculating in case of ungrouped frequency data through both direct and indirect methods.
8. Calculate the arithmetic mean marks of 5 students in Economics who scored 50,52,55,60 and 65 marks by :
   (a) direct method and (b) indirect method.
9. The annual salaries of four men were Rs.5000, Rs.6000, Rs.6500 and Rs. 30,000.
   a) Find the arithmetic mean of their salaries.
   b) Is this average typical of their salaries ?
10. Following are the marks obtained by 60 students of a class. Calculate the arithmetic mean by (a) direct and (b) indirect methods.

Marks in Economics (out of 100)	20, 30, 40, 50, 60, 70
No. of students	8, 12, 20, 10, 6, 4

11. How is arithmetic mean of a grouped frequency distribution calculated ? Explain both direct and indirect methods.
12. Following are the marks of students in History. Calculate the arithmetic mean marks by : (a) Direct method and (b) Indirect method (take 35 as assumed mean).

Marks in History	No.of students
8-10	5
10-20	10
20-30	25
30-40	30
40-50	20
50-60	10

13. Outline the need for a weighted arithmetic mean. How is it calculated ?
14. Calculate the weighted mean of

Number	12,29,14,41
Weights	6, 4, 5, 2

15. State the advantages and disadvantages of arithmetic mean as a measure of central tendency.

ANSWERS

Intext Questions 7.1

(i) dividing, first, second (ii) first (iii) second (iv) words (v) symbol (vi) i) 2 & ii) 1/2 (vii) not the same.

Intext Questions 7.2

(i) True (ii) False (iii) False (iv) False.

2. (i) a) 120 per cent b) 20 per cent
   (ii) a) 0.0115 b) 1.15 c) 11.5 d) 0.0105 e)1.05 f) 10.5

Intext Questions 7.3

1. (i) True (ii) False (iii) True (iv) False
2. (i) No. of observations (ii) Sigma, addition or 'sum of' similar terms (iii) 11.8 (mean marks) (iv)) 875 (mean monthly income).

Intext Questions 7.4

1. 13.4
2. a) Arithmetic mean b) Assumed or guessed mean
   c) Sum of deviations taken from arithmetic mean
   d) Total number of items in the series

Intext Questions 7.5

1. $f_1+f_2+f_3+$______f_n
2. N or Σf
3. 6.6
4. 6.4

Intext Questions 7.6

1. i) 15 ii) 17.5 iii) 28

2.

Marks	No.of students (f)	mid point (X)	fx
5-10	2	7.5	15
10-15	3	12.5	37.5
15-20	5	17.5	87.5
	Σf= 10		Σfx=140.0

3. $\overline{X} = \frac{\Sigma fx}{\Sigma f} = \frac{140}{10} = 14$

Intext Questions 7.7

1. $\frac{(3)(82) + (5)(86) + (3)(90) + (1)(70)}{3 + 5 + 3 + 1}$

2. $\frac{246 + 430 + 270 + 70}{12}$

3. $\frac{1016}{12} = 84.66$

Intext Questions 7.8

1. False 2. True 3. False 4. True

Terminal Exercise

1. Read section 7.3 (a)
2. Read section 7.3 (c)
3. Read section 7.4 (a)
4. Read section 7.4 (b)
5. Read section 7.4 (c)
6. Read section 7.5
7. Read section 7.7 and 7.7 (a)
8. a) 56.4 b) 56.4
9. a) Rs. 11,875
   b) This average score as you can observe is not typical of the data. This is a severe drawback of the arithmetic mean as it is affected by extreme values.
10. a) 41 marks b) 41 marks
11. Read section 7.7 (c)
12. a) 33 marks b) 33 marks
13. Read section 7.9
14. 20
15. Read section 7.10