**Illustration 21 :**

**(i) If $b_{yx} = 0.85$, $u = x - 15$ and $v = y - 20$, find the value of $b_{vu}$.**

**(ii) If $u = \frac{x-5}{3}$, $v = \frac{y-8}{5}$ and $b_{yx} = 0.9$, find the value of $b_{vu}$.**

**(iii) If $u = 10(x - 4.5)$, $v = \frac{y-50}{10}$ and $b_{yx} = 0.25$, find the value of $b_{vu}$.**

**(iv) If $u = 5(x - 40)$, $v = 2(y - 18)$ and $b_{yx} = 1.6$, find the value of $b_{vu}$.**

For the solution of all the questions given above, we shall use the following property of the regression coefficient.

- If $u = x - A$ and $v = y - B$ then $b_{yx} = b_{vu}$
- If $u = \frac{x-A}{c_x}$ and $v = \frac{y-B}{c_y}$ then $b_{yx} = b_{vu} \cdot \frac{c_y}{c_x}$

(i) Since $u = x - 15 = x - A$ and $v = y - 20 = y - B$

$\therefore\ b_{vu} = b_{yx} = 0.85$

(ii) Since $u = \frac{x-5}{3} = \frac{x-A}{c_x}$ and $v = \frac{y-8}{5} = \frac{y-B}{c_y}$

$$b_{yx} = b_{vu} \cdot \frac{c_y}{c_x} \quad \therefore\ b_{vu} = b_{yx} \cdot \frac{c_x}{c_y} = 0.9 \times \frac{3}{5} = 0.54$$

(iii) Since $u = 10(x - 4.5) = \frac{x-4.5}{\frac{1}{10}} = \frac{x-A}{c_x}$ and $v = \frac{y-50}{10} = \frac{y-B}{c_y}$

$$b_{yx} = b_{vu} \cdot \frac{c_y}{c_x} \quad \therefore\ b_{vu} = b_{yx} \cdot \frac{c_x}{c_y} = 0.25 \times \frac{\left(\frac{1}{10}\right)}{10} = 0.25 \times \frac{1}{100} = 0.0025$$

(iv) Since $u = 5(x - 40) = \frac{x-40}{\frac{1}{5}} = \frac{x-A}{c_x}$ and $v = 2(y - 18) = \frac{y-18}{\frac{1}{2}} = \frac{y-B}{c_y}$

$$b_{yx} = b_{vu} \cdot \frac{c_y}{c_x} \quad \therefore\ b_{vu} = b_{yx} \cdot \frac{c_x}{c_y} = 1.6 \times \frac{\left(\frac{1}{5}\right)}{\left(\frac{1}{2}\right)} = 1.6 \times \frac{2}{5} = 0.64.$$

## 3.8 Precautions while using Regression

We know that regression is a functional relation between two correlated variables and hence we can estimate the value of dependent variable from it. The regression analysis is very useful in decision making in the practical fields like economics, trade, industry, education, psychology, sociology, medicine, planning etc. Inspite of vast application of regression analysis, some precautions are necessary while using it.

(1) The reliability of the estimate can be verified by the coefficient of determination $\left(R^2\right)$. So, we should use the estimate only after ascertaining the linearity of regression by the coefficient of determination.

(2) Another point which is necessary to keep in mind while using the regression analysis is, the regression relation obtained by the scatter diagram or by the method of least squares should not be used for the values which are very far from the given values of the independent variable.

e.g. If for some data, there is a high degree of correlation between the rainfall and the yield of wheat, we can say that as rainfall increases, yield of wheat also increases. Now, using the regression relation obtained from the given data, if for some value of rainfall, corresponding yield of wheat is to be estimated then the estimate of yield of wheat can be proper only when the value of rainfall is around the given values of rainfall. If there is heavy rain then the crop may get damaged and the yield of wheat may also decrease. In such a case, the dependent variable (yield) estimated from the above mentioned regression relation may be wrong.

**Summary**

- The concept of regression is studied under the assumption that two variables under study have cause-effect relationship.
- Regression : Functional relation between two related variables.
- Linear Regression : Functional relation between two related variables in which the change in the values of the variables are approximately in constant proportion and this relationship can be determined by a straight line.
- The value of the dependent variable can be estimated for some known value of the independent variable by using regression.
- Regression coefficient : The approximate change in the value of dependent variable for a unit change in the value of independent variable. It is also known as slope of the regression line.
- Error : Mistake occurring in estimating the value of the dependent variable.
- Coefficient of Determination : It is the square of the correlation coefficient between the observed value of dependent variable $Y$ and its estimated values. In case of two variables, it is same as square of the coefficient of correlation between independent variable $X$ and dependent variable $Y$.
- Using the coefficient determination, how much variation in the dependent variable $Y$ is explained by the regression line can be known and the reliability of the regression model can also be known.
- The regression relation should not be used for the values which are very far from the given values of the independent variable.

**List of Formulae :**

Equation of Regression Line

$\hat{y} = a + bx$

Where, $b = b_{yx} =$ Regression Coefficient

(1) $b = \dfrac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2}$

(2) $b = \dfrac{n\Sigma xy-(\Sigma x)(\Sigma y)}{n\Sigma x^2-(\Sigma x)^2}$

(3) $b = \dfrac{n\Sigma uv-(\Sigma u)(\Sigma v)}{n\Sigma u^2-(\Sigma u)^2}$ Here, $u = x - A$ and $v = y - B$

(4) $b = \dfrac{n\Sigma uv-(\Sigma u)(\Sigma v)}{n\Sigma u^2-(\Sigma u)^2} \times \dfrac{c_y}{c_x}$ Here, $u = \dfrac{x-A}{c_x}$ and $v = \dfrac{y-B}{c_y}$

(5) $b = r \cdot \dfrac{s_y}{s_x}$

(6) $b = \dfrac{Cov(x, y)}{s_x^2}$

(7) $a = \bar{y} - b\bar{x}$

(8) Coefficient of Determination $R^2 = \left[r(y, \hat{y})\right]^2 = \left[r(x, y)\right]^2 = r^2$

## Exercise 3

### Section A

**Find the correct option for the following multiple choice questions :**

**1.** Which of the following indicates the functional relation between the two variables ?

(a) Correlation (b) Regression (c) Mean (d) Variance

**2.** The best fitted line of regression can be obtained by which method ?

(a) Least Square Method (b) Karl Pearson's Method

(c) Maximum Square Method (d) Bowley's Method

3. In usual notation, what is $b_{yx}$ ?

(a) Intercept (b) Dependent Variable

(c) The approximate change in the value of $Y$ for a unit change in the value of $X$.

(d) The approximate change in the value of $X$ for a unit change in the value of $Y$.

4. Which of the following is correct ?

(a) $b_{yx} = r \cdot \frac{s_x}{s_y}$ (b) $b_{yx} = r \cdot \frac{s_y^2}{s_x^2}$ (c) $b_{yx} = \frac{Cov(x, y)}{s_y^2}$ (d) $b_{yx} = r \cdot \frac{s_y}{s_x}$

5. The regression line always passes through which point ?

(a) $(\bar{x}, \bar{y})$ (b) $(0, \bar{y})$ (c) $(\bar{x}, 0)$ (d) $(0, 0)$

6. What is error $e$ in estimation in case of line of regression of $Y$ on $X$ ?

(a) $y - \hat{y}$ (b) $\hat{x} - \hat{y}$ (c) $x - \hat{x}$ (d) $\hat{y} - \hat{x}$

7. Which regression line is used if the sale of a commodity depends on its advertisement cost ?

(a) Regression line of advertisement cost on sale

(b) Regression line of advertisement cost on advertisement cost

(c) Regression line of sales on advertisement cost

(d) Regression line of sales on sales

8. Which of the following is a regression line of $Y$ on $X$ ?

(a) $\hat{y} = a + bx + cx^2$ (b) $\hat{x} = c + by$ (c) $\hat{y} = a + bx$ (d) $\hat{y} = a + bx^2$

9. For which value of the correlation coefficient $(r)$, the regression coefficient becomes zero ?

(a) 1 (b) –1 (c) $\frac{1}{2}$ (d) 0

10. What is coefficient of determination in the study of regression for two variables ?

(a) Product of two standard deviations (b) Square of correlation coefficient

(c) Square of covariance (d) Product of two variances

11. If the regression line is $\hat{y} = 10 + 3x$, what is the estimate of $Y$ for $X = 20$ ?

(a) 13 (b) 60 (c) 70 (d) 203

12. What is the value of $b_{yx}$ if the regression line is $2x + 3y - 50 = 0$ ?

(a) $\frac{3}{2}$ (b) $-\frac{3}{2}$ (c) $-\frac{2}{3}$ (d) 2

13. The regression line of $Y$ on $X$ is $\hat{y} = 30 - 1.5x$. What is the value of $\bar{y}$ if $\bar{x} = 10$ ?

(a) 28.5 (b) 20 (c) 15 (d) 45

14. If $u = \frac{x-15}{10}$ and $v = \frac{y-50}{2}$ and $b_{yx} = 7.5$, What is the value of $b_{vu}$ ?

(a) 7.5 (b) 1.5 (c) 37.5 (d) 150

15. If $r = 0.8$, how much part of the total variation in the dependent variable can be explained by the regression model ?

(a) 80 % (b) 64 % (c) 36 % (d) 20 %

## Section B

**Answer the following questions in one sentence :**

1. Define : Linear Regression
2. Define : Regression Coefficient
3. State the Linear Regression model.
4. What is an error in context with a regression line ?
5. Give the name of a method to obtain the best fitted regression line.
6. The regression coefficient is independent of which transformation ?
7. The regression coefficient is not independent of which transformation ?
8. What is the value of error if a sample point is on the fitted line ?
9. Will the regression coefficient change if the values of both the variables are doubled with the help of transformation of scale ?
10. If $r = 0.5, s_x = 2, s_y = 4$, what is the value of $b_{yx}$ ?
11. If a regression line is $\hat{y} = 31.5 + 1.85x$, estimate $Y$ for $X = 10$.
12. If $Y$ and $X$ have the relation $y = a + bx$, where $b > 0$ then what is the value of $r$?
13. If $y = 5 - 3x$ is the relation between $Y$ and $X$ then what is the value of $r$ ?

## Section C

**Answer the following questions :**

1. What are the constants $a$ and $b$ in the regression line $\hat{y} = a + bx$ ?
2. The fitted regression line of $Y$ on $X$ is $\hat{y} = 23.2 - 1.2x$ and one of the observations used in fitting of the line is (6, 17). Find the error in estimating $Y$ for $X = 6$.
3. If $\bar{x} = 30, \bar{y} = 20$ and $b = 0.6$, find the intercept of the regression line of $Y$ on $X$ and write equation of the line.
4. Interpret $b_{yx} = 5$.
5. If $b = 1.5, r = 0.8$ and standard deviation of $X$ is 1.6, find the standard deviation of $Y$.
6. If the regression coefficient of the regression line of $Y$ on $X$ is 0.6 and the standard deviations of $X$ and $Y$ are 5 and 3 respectively, find the coefficient of determination.
7. If the regression line of $Y$ on $X$ is $\hat{y} = 35 + 2x$ and $Cov(x, y) = 50$, find the standard deviation of $X$.
8. For the regression line given in the previous question (7), if the value of $Y$ is to be increased by 10 units, how many units should be increased in the value of $X$ ?
9. If $\bar{x} = 10, \bar{y} = 25, \Sigma(x - 10)(y - 25) = 120$ and $\Sigma(x - 10)^2 = 100$, find the values of $a$ and $b$ for the regression line of $Y$ on $X$.
10. If $b_{yx} = 0.75, u = 6(x - 20)$ and $v = 2(y - 15)$ for the data in the study of a regression line then find the value of $b_{vu}$.

## Section D

**Answer the following questions :**

1. Explain the statement, "There is a cause and effect relationship between two variables" by giving a suitable example. Also define independent variable and dependent variable.
2. Explain the method of scatter diagram for fitting a line of regression and state its limitation.
3. Explain the method of least square for fitting a regression line.
4. State the utility of regression.
5. State properties of regression coefficient. Also state the point through which a regression line always passes.
6. Explain : coefficient of determination
7. State precautions which are necessary while using the regression.
8. For two related variables $X$ and $Y$, $\Sigma(x-\bar{x})^2 = 80, \Sigma(x-\bar{x})(y-\bar{y}) = 60, \bar{x} = 8, \bar{y} = 10$. Obtain the regression line of $Y$ on $X$.
9. If $\bar{x} = 30, \bar{y} = 50, r = 0.8$ and the standard deviations of $X$ and $Y$ are 2 and 5 respectively, obtain the regression line of $Y$ on $X$.
10. If the regression line of $Y$ on $X$ is $\hat{y} = 11 + 3x$ and $s_x : s_y = 3:10$, find the coefficient of determination.
11. In usual notations, $n = 7, \Sigma u = 2, \Sigma v = 25, \Sigma u^2 = 160$ and $\Sigma uv = 409$. Obtain the regression coefficient of a regression line of $Y$ on $X$ and interpret it.
12. If $b_{yx} = 0.8$ then find the value of $b_{vu}$ for the following $u$ and $v$.

    (i) $u = x - 105$ and $v = y - 90$

    (ii) $u = \frac{x-1400}{100}$ and $v = \frac{y-750}{50}$

    (iii) $u = 10(x - 4.6)$ and $v = y - 75$
13. The following results are obtained for a bivariate data.

| **Particulars** | ***x*** | ***y*** |
|---|---|---|
| **No. of observations** | 8 | |
| **Mean** | 100 | 100 |
| **The sum of squares of deviations taken from mean** | 130 | 145 |
| **The sum of product of deviations taken from mean** | 115 | |

Obtain the regression line of $Y$ on $X$.

## Section E

**Solve the following :**

**1.** A manager of the an I.T. company has collected the following information regarding the years of job and monthly income of seven marketing executives.

| **Years of job** | 10 | 6 | 8 | 5 | 9 | 7 | 11 |
|---|---|---|---|---|---|---|---|
| **Monthly income (ten thousand ₹)** | 11 | 7 | 9 | 5 | 6 | 8 | 10 |

Obtain the regression line of the monthly income on the years of job of the marketing executives.

**2.** The information collected regarding price (in ₹ ) of a commodity and its supply (in hundred units) is as follows.

| **Price (₹)** | 59 | 60 | 61 | 62 | 64 | 57 | 58 | 59 |
|---|---|---|---|---|---|---|---|---|
| **Supply (hundred units)** | 78 | 82 | 82 | 79 | 81 | 77 | 78 | 75 |

Obtain the regression line of the supply on the price.

**3.** The following information is obtained for monthly advertisement cost and the sales of the last year for a company providing online shopping.

| **Particulars** | **Advertisement cost (ten thousand ₹)** | **Sales (lakh ₹)** |
|---|---|---|
| **Mean** | 10 | 90 |
| **Standard Deviation** | 3 | 12 |
| $r = 0.8$ | | |

Obtain the regression line of the sales on the advertisement cost.

**4.** The following results are obtained from the information of average rain and yield of a crop per acre in the last ten years of an arid region.

| **Particulars** | **Rainfall (cm)** | **Yield of crop (kg)** |
|---|---|---|
| **Mean** | 18 | 970 |
| **Standard Deviation** | 2 | 38 |
| **Correlation Coefficient** = 0.6 | | |

Estimate the yield of the crop if it rains 20 cms.

5. The information of investment (in lakh ₹) and its market price (in lakh ₹) after six months in share market in the last seven years for a Mutual Fund Company is obtained as follows.

| Particulars | Investment (lakh ₹) $x$ | Market price after six months (lakh ₹) $y$ |
|---|---|---|
| **Mean** | 40 | 50 |
| **Variance** | 100 | 256 |
| **Covariance** = 80 | | |

Obtain the regression line of $Y$ on $X$ and estimate the market price in the share market after six months if there is an investment of ₹ 45 lakh in a year.

## Section F

**Solve the following :**

1. Obtain the regression line of the demand on the price using the following information collected for the demand and the price of a commodity. Estimate the demand of the commodity if price is ₹ 40.

| **Price (₹)** | 38 | 36 | 37 | 37 | 36 | 38 | 39 | 36 | 38 |
|---|---|---|---|---|---|---|---|---|---|
| **Demand (hundred units)** | 12 | 18 | 15 | 12 | 17 | 13 | 13 | 15 | 12 |

2. The information regarding the experience (in years) of eight workers on a machine and their performance ratings based on the nondefective units they manufactured in every 100 units is as follows.

| **Experience of worker (years)** | 5 | 12 | 15 | 8 | 20 | 18 | 22 | 25 |
|---|---|---|---|---|---|---|---|---|
| **Performance rating** | 80 | 82 | 85 | 81 | 90 | 90 | 95 | 97 |

Obtain the regression line of the performance rating on the experience and estimate the performance rating if a worker has an experience of 17 years.

3. The information regarding daily income (in ₹) and expenditure (in ₹) of five labour families earning by daily work.

| **Daily income (₹)** | 200 | 300 | 400 | 600 | 900 |
|---|---|---|---|---|---|
| **Expenditure (₹)** | 180 | 270 | 320 | 480 | 700 |

Obtain the regression line of the expenditure on the daily income. Estimate the expenditure of a family having daily income of ₹ 500.

4. The following information is collected by a firm to know the effect of an advertisement campaign.

| **Year** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
|---|---|---|---|---|---|---|---|---|
| **Advertisement cost (ten thousand ₹)** | 12 | 15 | 15 | 23 | 24 | 38 | 42 | 48 |
| **Sales (crore ₹)** | 5 | 5.6 | 5.8 | 7 | 7.2 | 8.8 | 9.2 | 9.5 |

Obtain the regression line of sales on the advertisement cost. Estimate the sales when the advertisement cost is ₹ 5,00,000.

5. The information of eight construction companies regarding the number of contracts received in a year and the annual profit is as follows.

| No. of contracts | 2 | 5 | 9 | 12 | 6 | 4 | 8 | 10 |
|---|---|---|---|---|---|---|---|---|
| Annual profit (lakh ₹ ) | 100 | 300 | 700 | 1000 | 350 | 250 | 700 | 750 |

Obtain the regression line of the annual profit on the number of contracts. Verify the reliability of the regression model.

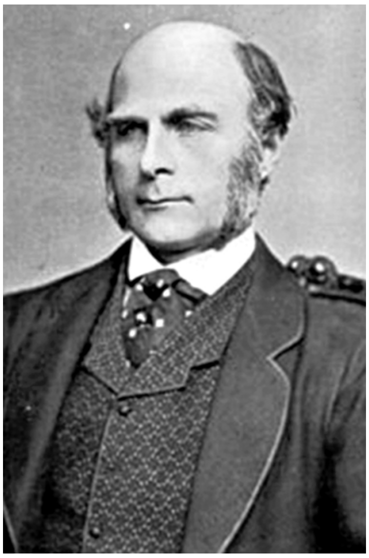
6. Obtain the regression line of $Y$ on $X$ from the following data and estimate $Y$ for $X = 30$.

$n = 10, \Sigma x = 250, \ \Sigma y = 300, \Sigma xy = 7900, \Sigma x^2 = 6500$

7. The following results are obtained for a data.

$n = 12, \Sigma x = 30, \ \Sigma y = 5, \Sigma x^2 = 670, \Sigma xy = 344$

Later on, it was known that one pair (10, 14) was wrongly taken as (11, 4). By correcting the above measures, obtain the regression line of $Y$ on $X$. Estimate $Y$ for $X = 5$.



**Sir Francis Galton**
**(1822 –1911)**

**Sir Francis Galton** was an English Victorian statistician, progressive, polymath, sociologist, psychologist, anthropologist, eugenicist, tropical explorer, geographer, inventor, meteorologist, protogeneticist and psychometrician. He was knighted in 1909.

Galton produced over 340 papers and books. He also created the statistical concept of correlation and widely promoted regression towards the mean. He was the first to apply statistical methods to the study of human differences and inheritance of intelligence, and introduced the use of questionnaires and surveys for collecting data on human communities, which he needed for genealogical and biographical works and for his anthropometric studies.

He was a pioneer in eugenics, coining the term itself and the phrase “nature versus nature”. His book Hereditary Genius (1869) was the first social scientific attempt to study genius and greatness.

As an investigator of the human mind, he founded psychometrics (the science of measuring mental faculties) and differential psychology and the lexical hypothesis of personality. He devised a method for classifying fingerprints that proved useful in forensic science.

●

*"Imperfect prediction, despite being imperfect can be valuable for decision making process."*

**– Michael Kattan**

# 4

# (Time Series)

**Contents :**

## 4.1 Time Series

### Introduction

Two related variables are studied by different methods in Statistics. A special method is applied to study the dependent variable among these related variables by taking time as an independent variable. The data related to values of the variable changing with time are studied in Economics, Sociology and Business Statistics. For example, population of a country, agricultural production, wholesale price index, unemployment statistics, import-export information, annual production of a certain factory, data from share market, bank interest rates, the hourly temperature measured in a city, etc. are presented with respect to time. These data are called time series as they are dependent on time.

### Meaning of Time Series

The statistical data collected at specific intervals of time and arranged in a chronological order is called Time Series. Time series consists of values of a variable associated with time. The estimated value of this variable for the future can be obtained if the values of such a variable are studied over a long period of time. These forecasts are very useful for future planning. For example, the direction, proportion and pattern of variations in the population of a certain region can be known by studying its time series. The necessary infrastructure, medical facilities, employment opportunities, education can be planned for the people of this region in future. The fluctuations in the prices of shares can be known by studying the time series of share prices of different companies and the investors can decide about buying or selling shares. The temperature measured at different places and time as well as the data of rainfall indicate the global changes in the weather which is useful to form policies for conservation of environment. In recent times, time series is extensively used in different methods of business analytics.

The data regarding changes in a variable at specific time interval are shown in the time series. The unit of time is dependent on the variable under study. For example, population data are obtained every ten years, data about total sales tax collected are available annually, quarterly interests are calculated in banks, monthly profits of shops are given, the time for bacterial growth is in hours, etc.

We will see the following illustrations of time series :

(1)

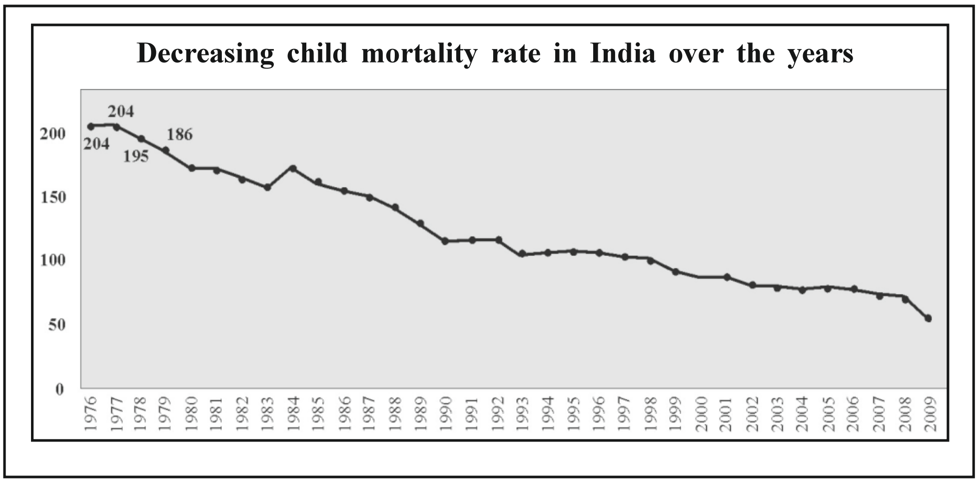
**MACRO FUNDAMENTALS** (in %)

| Year | GDP Growth | Investment Growth | Average WPI Inflation | CAD (As % of GDP) |
|---|---|---|---|---|
| 2002-03 | 4.0 | –0.4 | 3.4 | 0 |
| 2003-04 | 8.1 | 10.6 | 5.5 | 0 |
| 2004-05 | 7.0 | 24.0 | 6.5 | 0.4 |
| 2005-06 | 9.5 | 16.2 | 4.5 | 1.2 |
| 2006-07 | 9.6 | 13.8 | 6.6 | 1.0 |
| 2007-08 | 9.3 | 16.2 | 4.7 | 1.3 |
| 2008-09 | 6.7 | 3.5 | 8.1 | 2.3 |
| 2009-10 | 8.6 | 7.7 | 3.8 | 2.8 |
| 2010-11 | 9.3 | 14.0 | 9.6 | 2.8 |
| 2011-12 | 6.2 | 4.4 | 8.9 | 4.2 |
| 2012-13 | 5.4* | 2.3* | 7.6** | 4.7* |

* April-September ** April-December

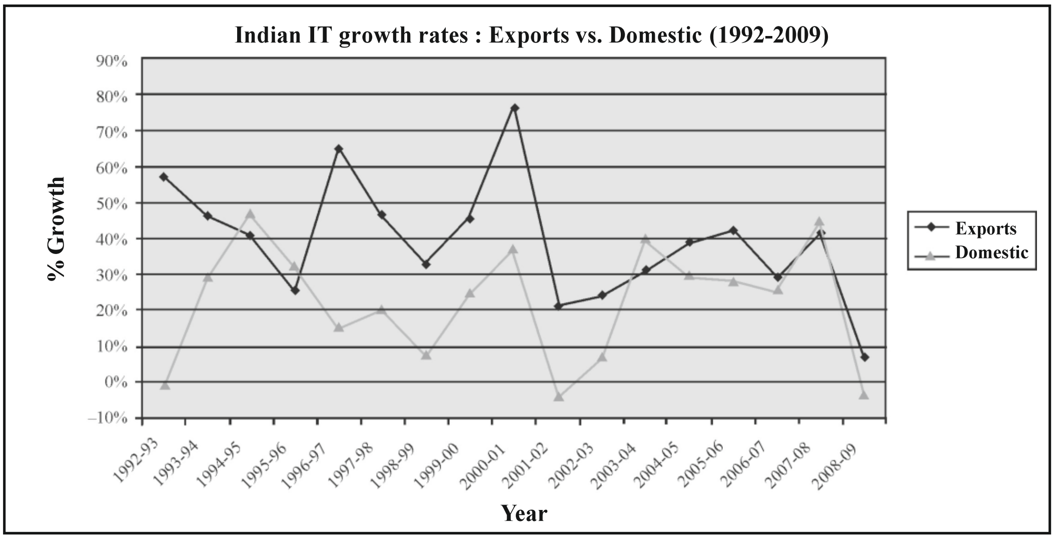
This time series gives information about macro fundamentals of different years which includes growth in Gross Domestic Product (GDP) along with percentage investment growth, Wholesale Price Index (WPI) and Current Account Deficit (CAD).

(2)



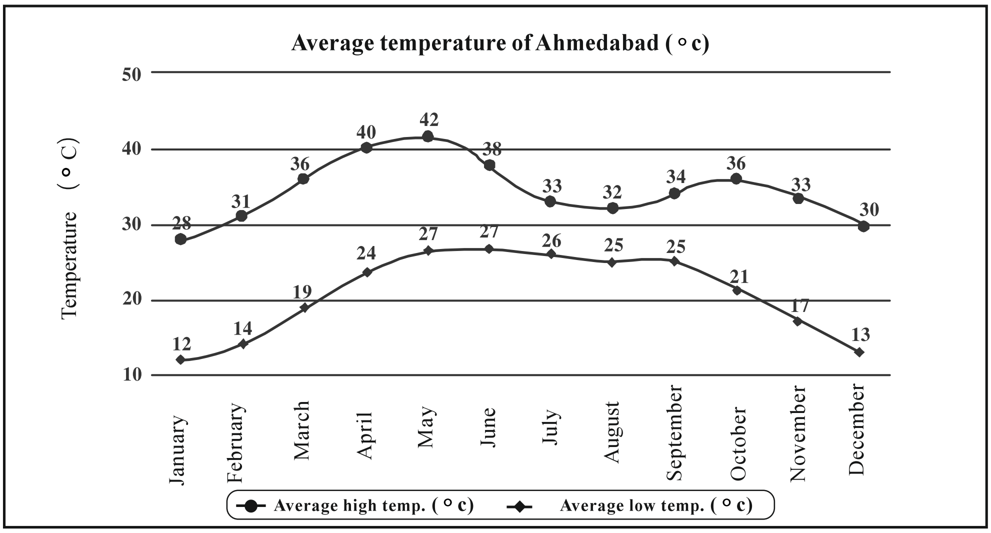

Time Series showing death rate among new-born children in India over different years.

(3)




A comparative study of two time series showing growth in IT sector.

(4)




Time Series showing average monthly maximum and minimum temperatures in Ahmedabad.

**Importance of Time Series**

The information collected in the form of time series is extremely essential in modern era due to increasing uncertainties in trade and business activities. The study of time series gains importance due to following reasons.

(1) The direction and pattern of variations in the values of the series can be known from the past data.

(2) The variations in the future can be estimated from the extent of variation in the values of the series.

(3) Important decisions can be taken from the estimated values of the future and industrial as well as government policies can be easily framed.

(4) Two or more industrialists or government institutions can make a comparative study from the data of the time series obtained by them.

**Definition of Time Series**

Time series is defined as follows :

'A time series is a set of observations taken at specified time periods.'

Usually these observations are taken at equal intervals of time.

The time is taken as an independent variable in the time series which will be denoted by $t$ and the dependent variable associated with it will be denoted by $y_t$. Thus, we shall represent the time series for different units of time as follows :

| **Time $t$** | 1 | 2 | 3 | ... | $n$ |
|---|---|---|---|---|---|
| **Variable $y_t$** | $y_1$ | $y_2$ | $y_3$ | ... | $y_n$ |

**Uses of Time Series**

The variations in the variable of a time series which changes with time are not caused by one specific reason. The variable of a time series is influenced by various factors and all these factors have an effect on the given variable. For example, the price of wheat in the wholesale market changes with time due to various different reasons such as the production of wheat at that time, demand of wheat, the cost of transporting the production to the market, etc. Each of these factors is dependent on other forces. For instance, the production of wheat is affected by various factors like the weather at that time, irrigation facility, the quality of seeds. It is necessary to study the various ways in which these factors affect the variable of the time series. Such a detailed study conducted for the time series is called as the analysis of time series which is done in the following two stages :

(1) To identify the various factors affecting the variable of the time series.

(2) To segregate these factors and determine the extent of effect of each factor on the given variable.

The analysis of time series done in this way is useful in trade, science, social and political fields as follows :

(1) It is possible to know the past situation and use it to obtain the type and measure of the variation.

(2) It is possible to estimate the value of the variable in future using statistical methods.

(3) Proper decisions can be taken for the future using the estimated values and activities can be planned accordingly.

(4) A comparative study can be carried out for the variations in the given variable at different places or time intervals.

(5) The estimates obtained from the past data can be compared with the present values and the reasons for the discrepancies between them can be investigated.

**Activity**

Collect the information with the help of your teachers about the percentage of students passing 12th standard from your school in the last 10 years and present it in the form of a time series.

## 4.2 Components of Time Series

We saw that there is a composite effect of many factors on the variable of the time series which brings fluctuations in the values of the variable. After observing different time series, it is known that the variations in the variable exhibit a specific pattern. The time series can be decomposed in the following components based on this pattern :

**(1) Long-term Component or Trend :** The variation seen in the variable of a time series over a very long period of time is the effect of long-term component or trend. The variable of a time series is generally found to have a continuous increase or decrease. This phenomenon is due to trend. For example, decreasing value of rupee in the international market, increasing usage of mobile phones, rising population of the country, decreasing death rates, etc. The intermittent short-term variations are ignored as the long-term variations in the time series are studied in the trend. The overall changes taking place in the variable of the series are considered here. The factors responsible for these changes produce a very slow pace of variation. For example, the number of literates are increasing in India but this change is taking place slowly in the last 60-70 years. Generally, the causes of such variations are the changing customs in the society, changes in tastes and choices of the people, technological changes in the industry, etc.

The trend of a time series is experienced after a very long time where 'long time' is a completely relative term. An interval of 10-15 years is required to know the trend in agricultural yield or industrial production whereas it may be clear within 4-5 years in the sale of electronic goods. The trend in the series having almost continuous increase or decrease is called a linear trend, which is generally observed in most of the time series. The data in economics, commerce and trade have series in which the rate of increase or decrease of the values does not remain constant. The rate of increase in the values of such series is very slow initially which goes on increasing slowly. The values stabilize after a certain interval of time and then start decreasing gradually. The trend of the series in such a situation is said to be Non-linear or Curvi-linear.

We will denote the component showing trend in the variable $y_t$ of the time series as '$T_t$'.

**(2) Seasonal Component :** The variation occurring in the time series variable almost regularly over a very short period of time is the effect of seasonal componont. The period of oscillation of such variations is usually less than a year. It is necessary to record the short-tem values in the time series to study these variations. It is not possible to get the information about the seasonal component if the yearly values of the given variable are available. The seasonal component affects the time series as follows :

**(i) Effect of natural factors :** The variations in the values of the time series occur in association with the seasons or weather fluctuations. Such variations occur at almost regular intervals. For example, the demand of fans, coolers or A.C. increases during summer whereas the demand of woollens increases in winter, the market prices reduce when the new crop is ready, etc.

**(ii) Effect of man-made factors :** The variations occurring regularly in less than one year period are caused by the festivals, customs in the society, habits of people, etc. For example, the purchase of ornaments increases during marriage season, kites are in demand during Uttarayan, the number of customers increases at the restaurants or theatres during weekends, increase in the purchase of clothing and gift articles during festivals, etc.

As the period of oscillation of these types of variations is almost certain, they are called regular variations. If the time and measure of these variations are known then the traders, producers are benefitted as higher profit can be earned by a control over their inventory.

We shall denote this short-term component of the time series by '$S_t$'.

**(3) Cyclical Component :** The variations occurring in the time series variable at approximately regular intervals of more than one year are the effects of cyclical component. The variations occurring due to this component are less regular as compared to seasonal component. The period of oscillation of these variations can be 2 to 10 years and in specific circumstances it can also be 10-15 years. The cyclical component is also considered as a short-term component as the time interval of variations due to this component is less than that of the time for the whole series which can be 40-50 years or even more. The cycles of boom and recession are examples of these variations. These cycles pass through the four stages namely, depression, recovery, boom, recession. These variations are found in the time series of trade and financial matters such as production, price of an item, prices of shares in share market, investment, etc. The traders can plan suitably with the help of estimate of time and measure of these variations.

This component of the time series is denoted by '$C_t$'.

**(4) Random or Irregular Component :** The effect of irregular or random component is also seen in the variations of the variable of time series in addition to the approximately regular short-term components like seasonal and cyclical which also give short-term effect. If the values in the series change due to sudden and unpredictable causes, the changes are called as random variations. The time-interval and effect of this variation is not certain. The variation which cannot be attributed to any one of the trend, seasonal component, cyclical component is the effect of random component. These fluctuations appear completely unexpected and are irregular. It cannot be predicted, it does repeat regularly and cannot be controlled. This variation occurs due to natural disasters like earthquake, floods or due to man-made predicaments like war, strike, political upheaval. The estimates can contain error due to this component.

This component is denoted by '$R_t$'.

The value of the variable of the time series $y_t$ based on time $t$ is determined with the combined effect of trend ($T_t$), seasonal component ($S_t$), cyclical component ($C_t$) and random component ($R_t$). This relationship is shown as follows in the additive model of the time series :

$$y_t = T_t + S_t + C_t + R_t$$

The seasonal component ($S_t$) does not appear if the yearly values of the variable are given for the time series. To find the effect of each component, trend ($T_t$) is found first using the given values $y_t$. After substracting it from $y_t$, the residual variation shows short-term components $(S_t, C_t, R_t)$. Then seasonal component (if available) and cyclical components are found. Random component is found in the end as $R_t = y_t - (T_t + S_t + C_t)$. The future estimate of the variable $(\hat{y}_t)$ is found by estimating the trend value and then adding the effect of each component at the given time ($t$) as mentioned above.

**Activity**

Prepare a time series of units of electricity consumption for the past one year from the electricity bills of your house. Identify the component of time series showing its effect on the variation in the variable of this series.

## 4.3 Methods for Determining Trend

Trend is an important component of time series. We will study the following methods to estimate it :

### 4.3.1 Graphical Method

This is the easiest method to find trend. The points are plotted on the graph paper by taking the independent variable, time ($t$), on $X$-axis and the dependent variable $y_t$ on $Y$-axis. These points are joined in their order by line segments. This shows the variation in the values of the variable. A smooth curve is then drawn through the middle of the points by personal judgement. This curve shows the trend by ignoring the short-term fluctuations in the series. The future estimates are obtained by extending the curve thus drawn.

The merits and limitations of graphical method are as follows.

**Merits :**

(1) This method is easy to understand and use.

(2) The trend can be found without any mathematical formula or calculations.

(3) This method can be used even if the trend is not linear.

(4) The judgement about the type of curve to be fitted for obtaining trend can be given by this method.

**Limitations :**

(1) It is possible that different people draw different curves. Hence, the uniformity is not maintained in the trend and its estimates.

(2) The estimates cannot be accurate as this is not a mathematical method and it is not possible to know the reliability of the estimates.

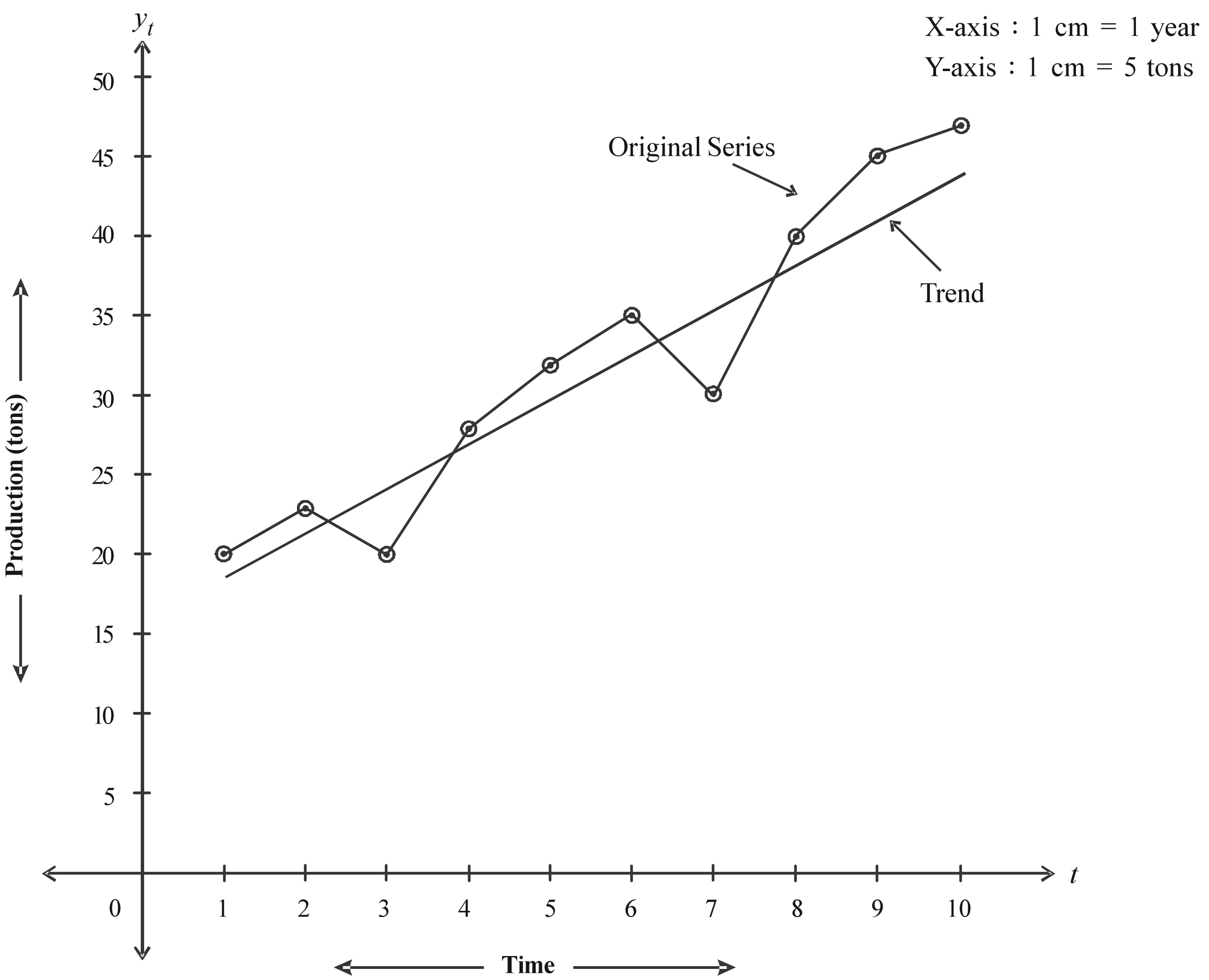
**Illustration 1 : The yearly production (in tons) of a factory is as follows. Obtain the trend using graphical method.**

| **Year** | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Production (tons)** | 20 | 23 | 20 | 28 | 32 | 35 | 30 | 40 | 45 | 47 |

We will represent these data as the following time series :

| **Time $t$** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Production (ton) $y_t$** | 20 | 23 | 20 | 28 | 32 | 35 | 30 | 40 | 45 | 47 |

We will plot these points on a graph by taking $t$ on $X$-axis and production $y_t$ on $Y$-axis. The pattern of points indicates that linear trend is more suitable.




The line passing through the middle of the points shows trend.

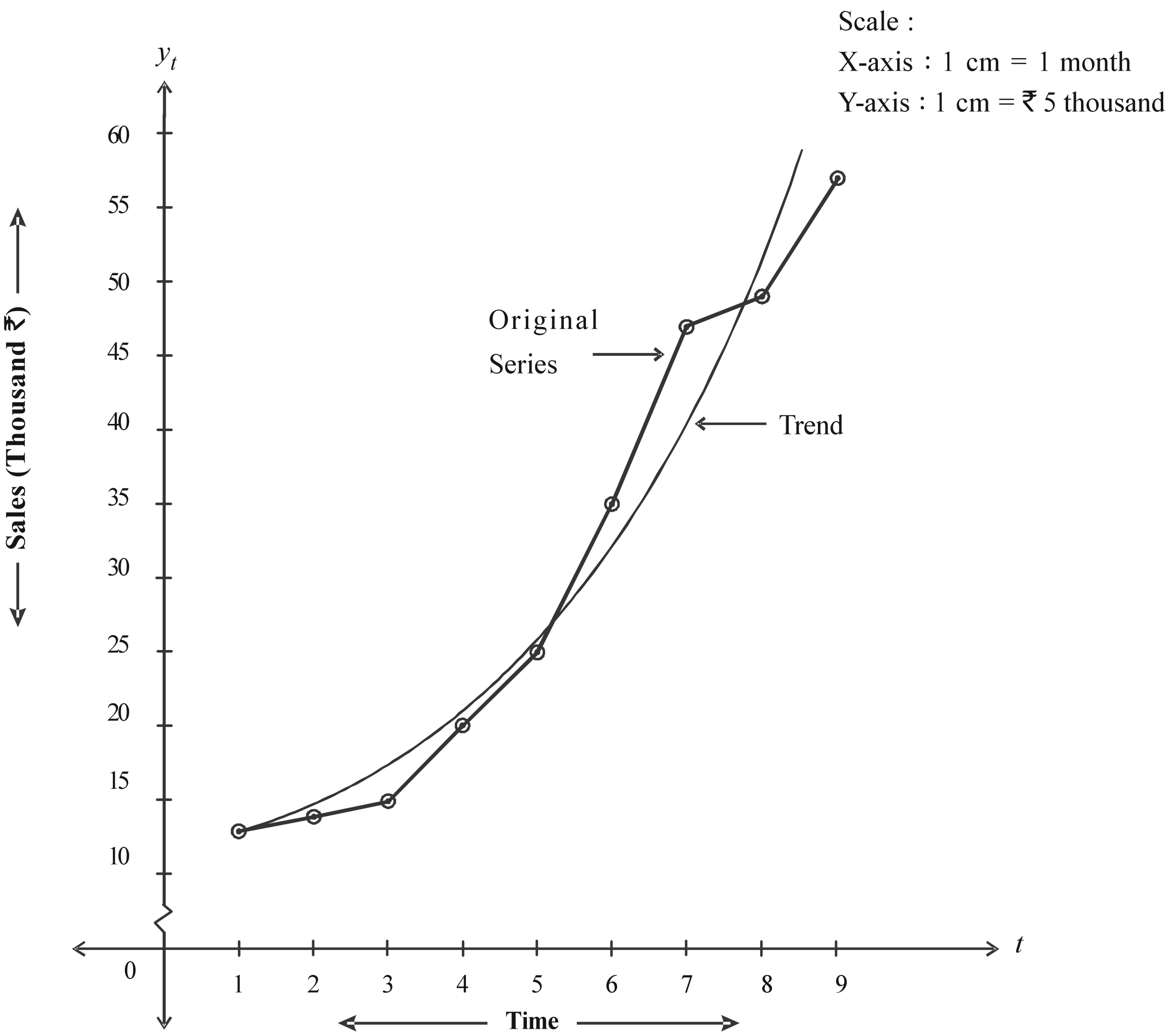
**Illustration 2 : The data about monthly sales (in thousand ₹) of a company are given in the following table. Obtain trend using grapical method.**

| **Month** | Jan. | Feb. | March | April | May | June | July | August | Sept. |
|---|---|---|---|---|---|---|---|---|---|
| **Sales (thousand ₹)** | 13 | 14 | 15 | 20 | 25 | 35 | 47 | 49 | 57 |

We will take the following time series for these data :

| **Time $t$** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| **Sale (thousand ₹) $y_t$** | 13 | 14 | 15 | 20 | 25 | 35 | 47 | 49 | 57 |

We will plot these points on a graph by taking $t$ on $X$-axis and sales $y_t$ on $Y$-axis. It indicates that the non-linear trend is more suitable for these data.




The curve passing through the middle of the points on the graph shows trend.

**EXERCISE 4.1**

**1.** The information about the capacity (in lakh tons) to load ships at a port each year is given below. Find the linear trend using graphical method.

| **Year** | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|
| **Capacity (lakh tons)** | 90 | 97 | 108 | 111 | 127 | 148 | 169 | 200 |

**2.** The number of tourists (in thousand) visiting a certain tourist place is as follows. Find the trend using a suitable graph.

| **Year** | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|---|
| **No. of tourists (thousand)** | 5 | 7 | 10 | 14 | 30 | 41 | 50 |

**3.** The data regarding number of girls ($y_t$) per 1000 boys in the age group 0-6 years of a state are given in the following table. Obtain the linear trend using graphical method.

| **Year** | 1961 | 1971 | 1981 | 1991 | 2001 | 2011 |
|---|---|---|---|---|---|---|
| $y_t$ | 956 | 948 | 947 | 928 | 883 | 890 |

**4.** The data about the closing prices of shares of a company for 10 days are given in the following table. Obtain the trend using graphical method.

| Day | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Price of share (₹) | 297 | 300 | 304 | 299 | 324 | 320 | 318 | 324 | 329 | 328 |

*

### 4.3.2 Method of Least Squares

As a limitation of the graphical method we saw that if a mathematical technique is not used then the trend and the estimates obtained from it change from person to person and its reliability cannot be known. If we want to find the linear trend of the time series by a mathematical method, we will require a specific linear equation which represents trend. We have studied the method of least squares in the chapter of regression to fit a linear equation to the given data which will be used to find the linear trend in a time series.

Suppose the values of the variable $y_t$ in the time series are available based on time $t$. We shall use the linear model $y_t = \alpha + \beta t + u_t$ (where $u_t$ is disturbance variable) to represent the relation between them. The estimated values $\hat{y}_t$ of $y_t$ can be found by fitting this model using the method of least squares. We will use the equation $\hat{y}_t = a + bt$ for this as shown in chapter 3.

We will ignore the suffix $t$ in $y_t$ for simplicity and consider $\hat{y} = a + bt$. The dependent variable is $y$ for the independent variable $t$.

The constants $a$ and $b$ are obtained by the method of least squares as follows :

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2} \quad \text{and} \quad a = \bar{y} - b\bar{t},$$

where $n$ = no. of observations

The linear equation thus obtained is the best linear equation for the given data.

The estimate of trend for the future is obtained using this linear equation.

**Note :** Other equations like polynomial, exponential equations can also be fitted besides the linear equation to find trend.

The merits and limitations of the method of least squares are as follows.

**Merits :**

(1) This method is absolutely mathematical and hence the future estimates do not change subjectively with the person.

(2) The trend estimates can be obtained by this method for each of the given values of $t$.

(3) The trend estimates can also be obtained for intermediate periods as the trend values are obtained using an equation. For example, the trend estimate for the period in the centre of the second and third year can be found by taking $t = 2.5$.

**Limitations :**

(1) This method requires extensive calculations to find trend.

(2) The reliability of the estimated values obtained by this method is less if an appropriate type of trend curve and its suitable equation is not fitted.

**Illustration 3 : The profit earned (in lakh ₹) by a company making computers is as follows. Find the linear equation for the trend from these data by least square method and estimate the profit for the year 2017.**

| **Year** | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|
| **Profit (Lakh ₹)** | 31 | 35 | 39 | 41 | 44 |

The values of profit are given for $n = 5$ years. We will thus denote the given years as $t = 1, 2, \ldots, 5$ respectively.

Calculation for fitting linear trend

| **Year** | **Profit** $y$ | $t$ | $t^2$ | $ty$ |
|---|---|---|---|---|
| 2011 | 31 | 1 | 1 | 31 |
| 2012 | 35 | 2 | 4 | 70 |
| 2013 | 39 | 3 | 9 | 117 |
| 2014 | 41 | 4 | 16 | 164 |
| 2015 | 44 | 5 | 25 | 220 |
| **Total** | **190** | **15** | **55** | **602** |

$\bar{t} = \frac{\Sigma t}{n} = \frac{15}{5} = 3, \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{190}{5} = 38$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{5 \times 602 - 15 \times 190}{5 \times 55 - (15)^2}$$

$$= \frac{3010 - 2850}{275 - 225}$$

$$= \frac{160}{50}$$

$$= 3.2$$

$$a = \bar{y} - b\bar{t}$$

$$= 38 - 3.2 \times 3$$

$$= 38 - 9.6$$

$$= 28.4$$

Equation for trend $\hat{y} = a + bt$

$\therefore \ \hat{y} = 28.4 + 3.2\,t$

We take $t = 7$ for the year 2017.

$\therefore \quad \hat{y} = 28.4 + 3.2 \times 7$

$= 28.4 + 22.4$

$= 50.8$

$\therefore \quad \hat{y} =$ ₹ 50.8 lakh

Thus, the estimated trend value of profit for the year 2017 is ₹ 50.8 lakh.

**Illustration 4 : The dropout rate of students of standard 1 to 5 from primary schools of a district is as follows :**

| **Year** | 2009 – 10 | 2010 – 11 | 2011 – 12 | 2012 – 13 | 2013 – 14 | 2014 – 15 | 2015 – 16 |
|---|---|---|---|---|---|---|---|
| **Dropout rate** | 3.24 | 2.98 | 2.29 | 2.20 | 2.09 | 2.07 | 2.04 |

**Estimate the dropout rate for students from standard 1 to 5 for the year 2016-17 and 2017-18 by fitting a linear equation for trend.**

The data are given for $n = 7$ years. We will thus denote the given years as $t = 1, 2, \ldots, 7$ respectively.

Calculations for fitting linear trend

| **Year** | **Dropout rate $y$** | $t$ | $t^2$ | $ty$ |
|---|---|---|---|---|
| 2009-10 | 3.24 | 1 | 1 | 3.24 |
| 2010-11 | 2.98 | 2 | 4 | 5.96 |
| 2011-12 | 2.29 | 3 | 9 | 6.87 |
| 2012-13 | 2.20 | 4 | 16 | 8.80 |
| 2013-14 | 2.09 | 5 | 25 | 10.45 |
| 2014-15 | 2.07 | 6 | 36 | 12.42 |
| 2015-16 | 2.04 | 7 | 49 | 14.28 |
| **Total** | **16.91** | **28** | **140** | **62.02** |

$$\bar{t} = \frac{\Sigma t}{n} = \frac{28}{7} = 4 , \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{16.91}{7} = 2.4157 \simeq 2.42$$

$$b = \frac{n \Sigma ty - (\Sigma t)(\Sigma y)}{n \Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{7 \times 62.02 - 28 \times 16.91}{7 \times 140 - (28)^2}$$

$$= \frac{434.14 - 473.48}{980 - 784}$$

$$= \frac{-39.34}{196}$$

$$= -0.2007$$

$$\simeq -0.2$$

$$a = \overline{y} - b\overline{t}$$

$$= 2.42 - (-0.2) \times 4$$

$$= 2.42 + 0.8$$

$$= 3.22$$

Equation for trend $\hat{y} = a + bt$

$$\therefore \hat{y} = 3.22 + (-0.2)\, t$$

$$= 3.22 - 0.2\, t$$

We take $t = 8$ for the year 2016-17.

$$\therefore \hat{y} = 3.22 - 0.2 \times 8$$

$$= 3.22 - 1.6$$

$$= 1.62$$

We take $t = 9$ for the year 2017-18.

$$\therefore \hat{y} = 3.22 - 0.2 \times 9$$

$$= 3.22 - 1.8$$

$$= 1.42$$

Thus, the estimates of dropout rates for the students of standard 1 to 5 in this district for the years 2016-17 and 2017-18 are 1.62 and 1.42 respectively.

**Illustration 5 : The data of population (in lakh) of a taluka are given in the following table. Fit a linear equation for the data and find the trend value for each year. Also find the trend estimate for the population in the year 2021.**

| **Year** | 1951 | 1961 | 1971 | 1981 | 1991 | 2001 | 2011 |
|---|---|---|---|---|---|---|---|
| **Population (lakh)** | 15.1 | 16.9 | 18.7 | 20.1 | 21.6 | 25.7 | 27.1 |

The data about population are given which are associated with each decade. We will take $t = 1, 2, \ldots, 7$ respectively for the given years. Hence, we get $n = 7$.

Calculation for fitting linear trend

| Year | Population (lakh) $y$ | $t$ | $t^2$ | $ty$ | Trend values $\hat{y} = 12.66 + 2.02\, t$ |
|---|---|---|---|---|---|
| 1951 | 15.1 | 1 | 1 | 15.1 | 14.68 |
| 1961 | 16.9 | 2 | 4 | 33.8 | 16.7 |
| 1971 | 18.7 | 3 | 9 | 56.1 | 18.72 |
| 1981 | 20.1 | 4 | 16 | 80.4 | 20.74 |
| 1991 | 21.6 | 5 | 25 | 108 | 22.76 |
| 2001 | 25.7 | 6 | 36 | 154.2 | 24.78 |
| 2011 | 27.1 | 7 | 49 | 189.7 | 26.8 |
| **Total** | **145.2** | **28** | **140** | **637.3** | |

$\overline{t} = \frac{\Sigma t}{n} = \frac{28}{7} = 4$ , $\overline{y} = \frac{\Sigma y}{n} = \frac{145.2}{7} = 20.7429 \simeq 20.74$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{7 \times 637.3 - 28 \times 145.2}{7 \times 140 - (28)^2}$$

$$= \frac{4461.1 - 4065.6}{980 - 784}$$

$$= \frac{395.5}{196}$$

$$= 2.0179$$

$$\simeq 2.02$$

$$a = \overline{y} - b\overline{t}$$

$$= 20.74 - 2.02 \times 4$$

$$= 20.74 - 8.08$$

$$= 12.66$$

Equation for trend $\hat{y} = a + bt$

$\therefore \hat{y} = 12.66 + 2.02\, t$

We take $t = 1, 2, .... , 7$ for each of the given year respectively to find the values of trend. Taking $t = 1$,

$$\hat{y} = 12.66 + 2.02 \times 1$$
$$= 12.66 + 2.02$$
$$= 14.68$$

$\therefore \quad \hat{y} = 14.68$ lakh

Similarly we will take $t = 2, 3, .... , 7$ to find the remaining values of trend and show them in the table.

It can be seen here that the values of $\hat{y}$ increase successively by 2.02.

We take $t = 8$ for the year 2021

$$\hat{y} = 12.66 + 2.02 \times 8$$
$$= 12.66 + 16.16$$
$$= 28.82$$

$\therefore \quad \hat{y} = 28.82$ lakh

Thus, the estimate for trend value for the population for the year 2021 is 28.82 lakh.

**Illustration 6 : The data about monthly sales (in thousand ₹) of a company are given in the following table. Fit a linear trend and show it grapically. Estimate the sale for the month of August using the equation obtained.**

| **Month** | January | February | March | April | May | June |
|---|---|---|---|---|---|---|
| **Sale (thousand ₹)** | 80 | 85 | 90 | 76 | 82 | 88 |

The data for $n = 6$ months are given here. Hence, we will take $t = 1, 2, .... , 6$ for the given months respectively.

Calculation for fitting linear Trend

| **Month** | **Sales $y$ (thousand ₹)** | $t$ | $t^2$ | $ty$ | $\hat{y} = 81.79 + 0.49\ t$ |
|---|---|---|---|---|---|
| January | 80 | 1 | 1 | 80 | 82.28 |
| February | 85 | 2 | 4 | 170 | 82.77 |
| March | 90 | 3 | 9 | 270 | 83.26 |
| April | 76 | 4 | 16 | 304 | 83.75 |
| May | 82 | 5 | 25 | 410 | 84.24 |
| June | 88 | 6 | 36 | 528 | 84.73 |
| **Total** | **501** | **21** | **91** | **1762** | |

$$\bar{t} = \frac{\Sigma t}{n} = \frac{21}{6} = 3.5, \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{501}{6} = 83.5$$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{6 \times 1762 - 21 \times 501}{6 \times 91 - (21)^2}$$

$$= \frac{10572 - 10521}{546 - 441}$$

$$= \frac{51}{105}$$

$$= 0.4857$$

$$\simeq 0.49$$

$$a = \bar{y} - b\bar{t}$$

$$= 83.5 - 0.49 \times 3.5$$

$$= 83.5 - 1.715$$

$$= 81.785$$

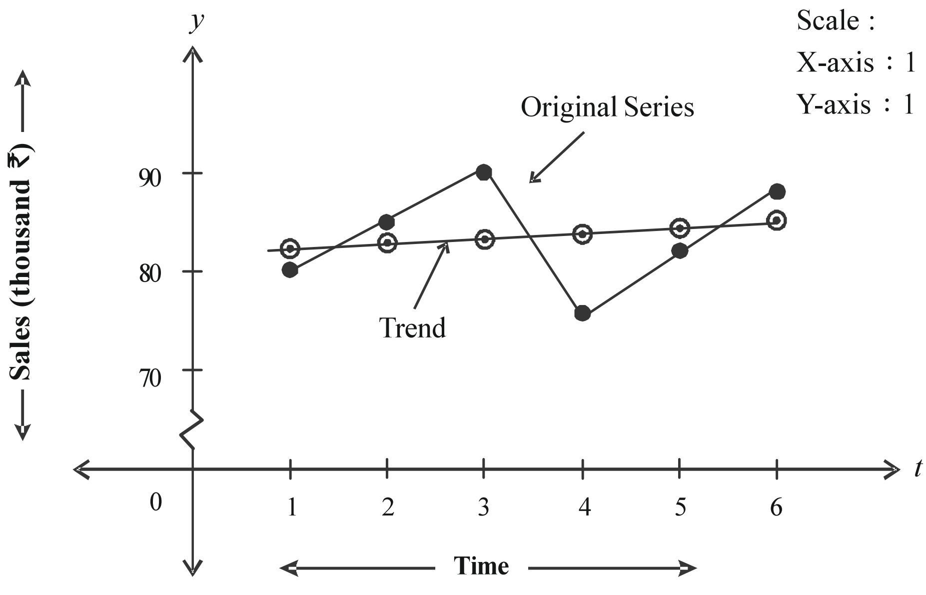
$$\simeq 81.79$$

Equation for trend $\hat{y} = a + bt$

$\therefore \hat{y} = 81.79 + 0.49\,t$

By substituting $t = 1, 2, .... , 6$ successively, we get the corresponding values of $\hat{y}$ which are shown in the table.

The trend values and the values in the given series can be shown in the graph as follows :




Scale :
X-axis : 1 cm = 1 month
Y-axis : 1 cm = ₹ 10 thousand

Now we take $t = 8$ for the month of August

$$\hat{y} = 81.79 + 0.49 \times 8$$

$$= 81.79 + 3.92$$

$$= 85.71$$

$\therefore \hat{y} =$ ₹ 85.71 thousand

Thus, the trend estimate for sales of this company in the month of August is ₹ 85.71 thousand.

**Note :** It is not necessary to take all the values of $\hat{y}$ to draw the line representing the linear equation. The equation of trend can be shown in the graph by joining the values of $\hat{y}$ corresponding to any two values among $t = 1, 2, \ldots, 6$.

**Illustration 7 : Obtain the linear equation for trend for a time series with $n = 8$, $\Sigma y = 344$, $\Sigma ty = 1342$**

Since $n = 8$, we take $t = 1, 2, \ldots, 8$ Hence, we get $\Sigma t = 1 + 2 + \ldots + 8 = 36$ and

$\Sigma t^2 = 1^2 + 2^2 + \ldots + 8^2 = 1 + 4 + \ldots + 64 = 204$.

$$\bar{t} = \frac{\Sigma t}{n} = \frac{36}{8} = 4.5, \qquad \bar{y} = \frac{\Sigma y}{n} = \frac{344}{8} = 43$$

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}$$

$$= \frac{8 \times 1342 - 36 \times 344}{8 \times 204 - (36)^2}$$

$$= \frac{10736 - 12384}{1632 - 1296}$$

$$= \frac{-1648}{336}$$

$$= -4.9048$$

$$\simeq -4.9$$

$$a = \bar{y} - b\bar{t}$$

$$= 43 - (-4.9) \times 4.5$$

$$= 43 + 22.05$$

$$= 65.05$$

Equation for trend $\hat{y} = a + bt$

$$= 65.05 + (-4.9)\,t$$

$$= 65.05 - 4.9\,t$$

## EXERCISE 4.2

1. The information about death rate of a state in different years is given in the following table. Fit a linear equation to find trend and hence estimate the death rate for the year 2017.

| **Year** | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|
| **Death rate** | 7.6 | 6.9 | 7.1 | 7.3 | 7.2 | 6.9 | 6.9 |

2. The data about Cost Inflation Index (CII) declared by the central government are as follows. The year 1981-82 is the base for this index. Find the estimate of this index for the year 2015-16 by fitting the linear equation to these data.

| **Year** | 2007 – 08 | 2008 – 09 | 2009 – 10 | 2010 – 11 | 2011 – 12 | 2012 – 13 | 2013 – 14 | 2014 – 15 |
|---|---|---|---|---|---|---|---|---|
| ***CII*** | 551 | 582 | 632 | 711 | 785 | 852 | 939 | 1024 |

**3.** The number of two wheelers registered (in thousand) in a city in different years is as follows. Use the method of fitting linear equation to these data to obtain the estimates for the number of vehicles registered in the year 2016 and 2017. Also find the trend values for each year.

| **Year** | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|
| **No. of vehicles (thousand)** | 69 | 75 | 82 | 91 | 101 | 115 |

**4.** The average age of women (in years) at the time of marriage obtained from the data of different census surveys in India are given in the following table. Fit an equation for a linear trend from the data and show it on a graph. Find the estimate for the value of the given variable for the year 2021 using the linear equation.

| **Year of census survey** | 1971 | 1981 | 1991 | 2001 | 2011 |
|---|---|---|---|---|---|
| **Average age of women at marriage (years)** | 17.7 | 18.7 | 19.3 | 20.2 | 22.2 |

*

### 4.3.3 Method of Moving Averages

The method of moving averages is very useful to find trend by eliminating the effect of short-term variations. The short-term variations are usually regular and have repetitions. The period of repetition of these variations can be found by past experience or other techniques and the average is found for the number of observations corresponding to this period. Since the average value lies in the center, we get the values that are free from the short-term fluctuations which show the trend.

Suppose, the values of the dependent variable in the given time series are $y_1, y_2, ...., y_n$ for time $t = 1, 2, .... , n$ respectively and the interval for short-term cyclical fluctuations is 3 years. The mean of first three observations $y_1, y_2, y_3$ is found as $\frac{y_1 + y_2 + y_3}{3}$ and it is written against the center of these three values which is $y_2$. Further, the mean of successive three values $y_2, y_3, y_4$ is found as $\frac{y_2 + y_3 + y_4}{3}$ and it is written against $y_3$ which is the center of these three values. Similarly, all the means are calculated till the last value from the given values of the variable is included. The averages thus calculated are called three yearly moving averages which indicate the trend.

It is not necessary that the unit of time in every time series is year and time interval for repetitions in the pattern of the values of the variable may not be necessarily three years. The moving averages will be denoted according to the unit of time. For example, 5 days moving averages, three monthly moving averages, 4 weekly moving averages, etc. The unit for time is taken as 'year' for the discussion here.

While calculating for the given data, we first find the total of values of the variable in accordance with the time interval for the average. After finding the first total $y_1 + y_2 + y_3$, the next total namely $y_2 + y_3 + y_4$ is found by subtracting $y_1$ from the above total and then adding $y_4$ to it. All the successive totals are found in this manner and each total is divided by 3 to obtain three yearly moving averages.

**Note :** The first three yearly moving average is written against $y_2$ and thus the moving average against $y_1$ i.e. the trend value at that time cannot be obtained. Similarly, the trend value corresponding to $y_n$ cannot be obtained.

**Illustration 8 : The number of accounts opened in different weeks in a branch of a certain bank are given below. Find the trend using three-weekly moving averages.**

| **Week** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **No. of accounts opened** | 26 | 27 | 26 | 25 | 22 | 24 | 25 | 23 | 22 | 21 |

Calculation for three weekly moving averages

| Week $t$ | No. of accounts opened $y$ | Three weekly moving total | Three weekly moving average |
|---|---|---|---|
| 1 | 26 | — | — |
| 2 | 27 | $26 + 27 + 26 = 79$ | $\frac{79}{3} = 26.33$ |
| 3 | 26 | $79 - 26 + 25 = 78$ | $\frac{78}{3} = 26$ |
| 4 | 25 | $78 - 27 + 22 = 73$ | $\frac{73}{3} = 24.33$ |
| 5 | 22 | $73 - 26 + 24 = 71$ | $\frac{71}{3} = 23.67$ |
| 6 | 24 | $71 - 25 + 25 = 71$ | $\frac{71}{3} = 23.67$ |
| 7 | 25 | $71 - 22 + 23 = 72$ | $\frac{72}{3} = 24$ |
| 8 | 23 | $72 - 24 + 22 = 70$ | $\frac{70}{3} = 23.33$ |
| 9 | 22 | $70 - 25 + 21 = 66$ | $\frac{66}{3} = 22$ |
| 10 | 21 | — | — |

**Illustration 9 : Find the trend using five yearly moving averages for the following data about yearly production (in tons) of a factory.**

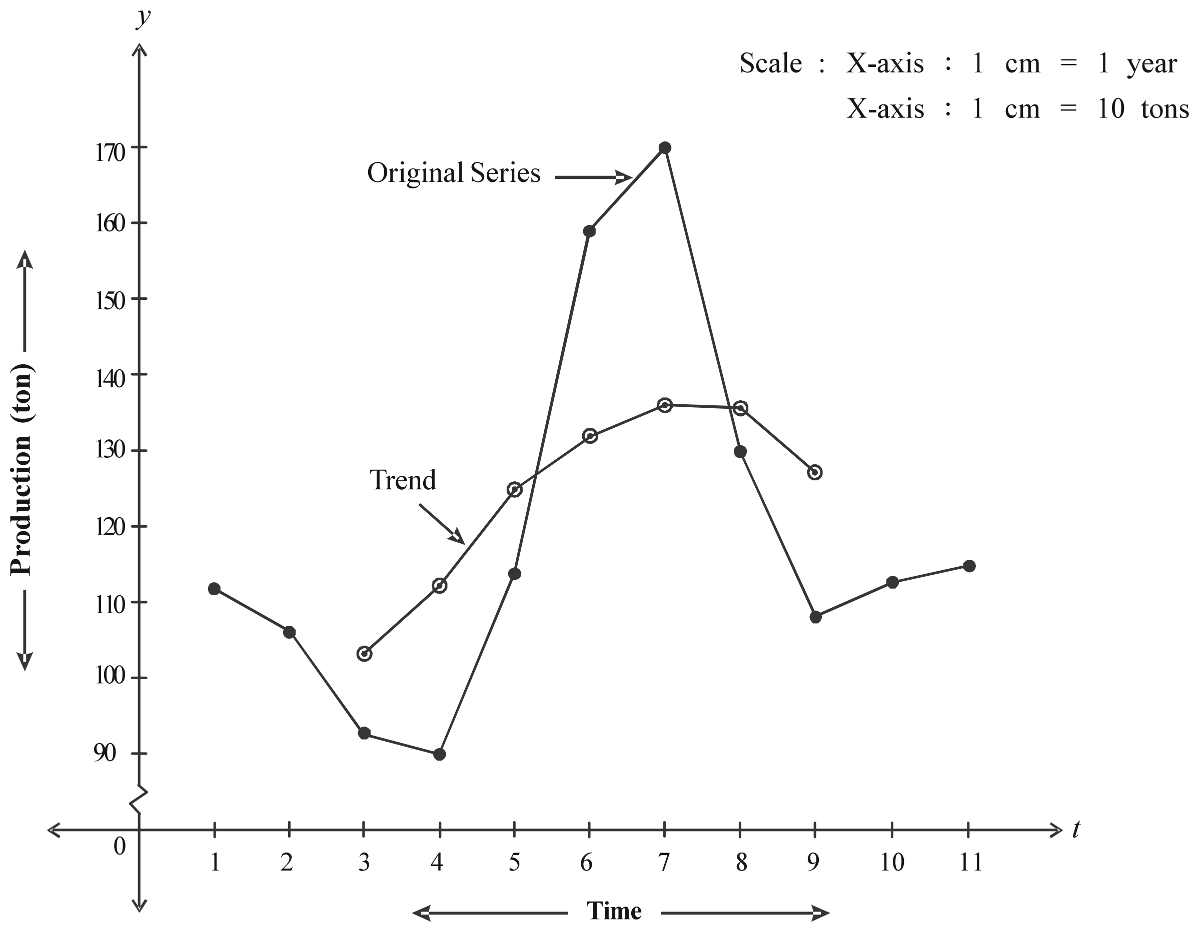
| Year | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Production (tons)** | 112 | 106 | 93 | 90 | 114 | 159 | 170 | 130 | 108 | 113 | 115 |

Calculation for five yearly moving averages

| Year | Production $y$ | $t$ | Five yearly moving total | Five yearly moving average (trend) |
|---|---|---|---|---|
| 2006 | 112 | 1 | — | — |
| 2007 | 106 | 2 | — | — |
| 2008 | 93 | 3 | $112+106+93+90+114=515$ | $\frac{515}{5}=103$ |
| 2009 | 90 | 4 | $515-112+159=562$ | $\frac{562}{5}=112.4$ |
| 2010 | 114 | 5 | $562-106+170=626$ | $\frac{626}{5}=125.2$ |
| 2011 | 159 | 6 | $626-93+130=663$ | $\frac{663}{5}=132.6$ |
| 2012 | 170 | 7 | $663-90+108=681$ | $\frac{681}{5}=136.2$ |
| 2013 | 130 | 8 | $681-114+113=680$ | $\frac{680}{5}=136$ |
| 2014 | 108 | 9 | $680-159+115=636$ | $\frac{636}{5}=127.2$ |
| 2015 | 113 | 10 | — | — |
| 2016 | 115 | 11 | — | — |

**Additional information for understanding**

We shall show the values of the variable and the trend obtained by five yearly moving averages to understand the trend found by this method.




The curve passing through all the averages shows the trend.

If the interval of time for the moving averages is odd number like 3, 5, 7, .... then the trend is found as shown earlier. But the calculation of moving averages becomes difficult if this interval is an even number.

Suppose four yearly moving averages are to be found. First four yearly average will be found as $\frac{y_1 + y_2 + y_3 + y_4}{4}$. As the center for these four values is between $y_2$ and $y_3$, this average will be written at that position. Similarly, the successive averages namely $\frac{y_2 + y_3 + y_4 + y_5}{4}$, $\frac{y_3 + y_4 + y_5 + y_6}{4}$, ..... will be found and written between $y_3$ and $y_4$, between $y_4$ and $y_5$, .... respectively. Since these averages are in between two years, the average of each pair of averages is found and it is written between two moving averages. Thus, the average value of the first two averages shown above will be written against $y_3$. The averages thus obtained are called as four yearly moving averages. The processes of finding an average is to be done twice here. To simplify these calculations, four yearly totals are obtained first and then totals of pairs of years are found. As these totals involve 8 values, each total is divided by 8 which gives the four yearly averages mentioned above.

Whenever the time of the cycles of short-term variations is an even number, moving averages are obtained by first finding the moving totals and then the pairwise totals as shown in the above method.

**Illustration 10 : Find the trend using four monthly moving averages for the following data showing monthly sales (in lakh ₹) of a shop.**

| **Month** | March | April | May | June | July | August | Sept. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sales (lakh ₹)** | 5 | 3 | 7 | 6 | 4 | 8 | 9 | 10 | 8 | 9 |

Calculation of four monthly moving averages

| **Month** | **Sales (lakh ₹)** $y$ | $t$ | **Four monthly moving total** | **Pairwise total** | **Four monthly moving average** |
|---|---|---|---|---|---|
| March | 5 | 1 | | — | — |
| | | | — | | |
| April | 3 | 2 | | — | — |
| | | | $5+3+7+6=21$ | | |
| May | 7 | 3 | | $21+20=41$ | $\frac{41}{8}=5.13$ |
| | | | $21-5+4=20$ | | |
| June | 6 | 4 | | $20+25=45$ | $\frac{45}{8}=5.63$ |
| | | | $20-3+8=25$ | | |
| July | 4 | 5 | | $25+27=52$ | $\frac{52}{8}=6.5$ |
| | | | $25-7+9=27$ | | |
| August | 8 | 6 | | $27+31=58$ | $\frac{58}{8}=7.25$ |
| | | | $27-6+10=31$ | | |
| September | 9 | 7 | | $31+35=66$ | $\frac{66}{8}=8.25$ |
| | | | $31-4+8=35$ | | |
| October | 10 | 8 | | $35+36=71$ | $\frac{71}{8}=8.88$ |
| | | | $35-8+9=36$ | | |
| November | 8 | 9 | | — | — |
| | | | — | | |
| December | 9 | 10 | | — | — |

The trend of the time series is shown by the four monthly moving averages.

**Merits and limitations of the method of moving averages are as follows :**

**Merits :**

(1) The effect of short-term component is eliminated to a large extent using the averages and the trend of the series is obtained.

(2) The calculation is easy to understand as it is comparatively less and simple.

**Limitations :**

(1) The trend obtained by this method is not accurate if the interval for the moving averages is not chosen correctly.

(2) The estimates of trend for some initial and last time periods cannot be obtained.

(3) A specific mathematical formula is not obtained for future estimates.

## EXERCISE 4.3

**1.** Find the trend by three yearly moving averages from the following data about the sales (in ten lakh ₹) of a company.

| **Year** | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sales (ten lakh) ₹** | 3 | 4 | 8 | 6 | 7 | 11 | 9 | 10 | 14 | 12 |

**2.** The average monthly closing prices of shares of a company in the year 2016 are given in the following table. Find the trend using four monthly moving averages.

| **Month** | January | February | March | April | May | June | July | August | Sept. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Share price (₹)** | 253 | 231 | 350 | 261 | 262 | 266 | 263 | 261 | 281 | 278 | 278 | 272 |

**3.** Find the trend using five yearly moving averages from the following data of profit (in lakh ₹) of a trader in different years.

| **Year** | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|
| **Profit (lakh ₹)** | 15 | 14 | 18 | 20 | 17 | 24 | 27 | 25 | 23 |

**4.** The wholesale price index numbers for different quarters ($Q$) of a year are obtained as follows. Find the trend by four quarterly moving averages.

| **Year** | 2013 | | | | 2014 | | | | 2015 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Quarter** | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ |
| **Index No.** | 110 | 110 | 125 | 135 | 145 | 152 | 155 | 168 | 131 | 124 | 132 | 153 |

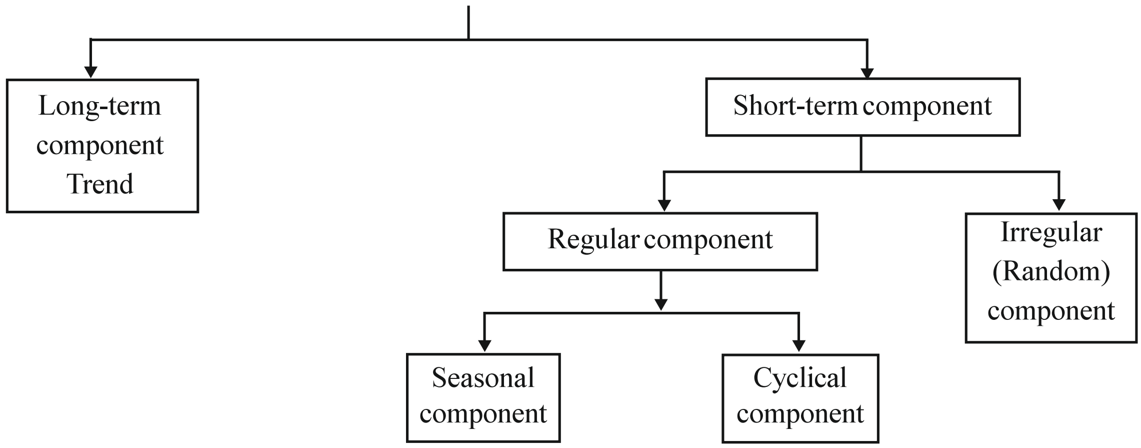
*

## Summary

- The data collected and arranged according to time is called Time Series.
- It is necessary to analyse the time series to find the future estimates of the given variable.
- There are four main components affecting the values of the variable in a time series :
  (1) Long-term Component (Trend) (2) Seasonal Component
  (3) Cyclical Component (4) Random (Irregular) Component
- Short-term fluctuations are found in the time series due to seasonal component, cyclical component and random component.
- Seasonal and cyclical fluctuations repeat almost regularly.
- Three methods of measuring trend :
  (1) Graphical Method (2) Method of least squares (3) Method of moving average

## Chapter at a glance

**Time Series**




**List of formulae :**

For fitting a linear equation $\hat{y} = a + bt$ to the given data

$$b = \frac{n\Sigma ty - (\Sigma t)(\Sigma y)}{n\Sigma t^2 - (\Sigma t)^2}, \qquad a = \bar{y} - b\bar{t}$$

## Exercise 4

### Section A

**Find the correct option for the following multiple choice questions :**

**1.** Which type of variations are produced in the time series variable due to seasonal component ?
(a) Long-term (b) Irregular (c) Regular (d) Zero

**2.** Which variation is shown in 'decrease in the production of a company' due to strike ?
(a) Random (b) Trend (c) Seasonal (d) Cyclical

**3.** Name the method for fitting the linear equation to find linear trend.
(a) Graphical Method (b) Method of least squares
(c) Method of moving average (d) Method of partial average

**4.** How do you show the additive model of the time series ?

(a) $y_t = T_t + S_t + C_t - R_t$ (b) $y_t = T_t + S_t + C_t + R_t$

(c) $y_t = T_t \times S_t + C_t \times R_t$ (d) $y_t = S_t + C_t + R_t$

**5.** State the independent variable of time series.

(a) $y_t$ (b) $S_t$ (c) $t$ (d) $x_t$

**6.** Which component of the time series is impossible to predict ?

(a) Random component (b) Trend (c) Seasonal component (d) Cyclical component

**7.** Which of the following variations are due to cyclical component ?

(a) Rise in demand during winter

(b) Decrease in the share prices due to recession in share market

(c) Decrease in the agricultural produce due to excessive rains

(d) Continuously decreasing death rate

**8.** The trend equation obtained from a time series from January 2016 to December 2016 is $\hat{y} = 30.1 + 1.5\,t$. Find the value of trend for April 2016.

(a) 30.1 (b) 34.6 (c) 36.1 (d) 33.1

**9.** Which of the following fluctuations is the effect of seasonal component ?

(a) Increase in the migration to cities from rural areas

(b) Increasing number of vehicles on roads in a city

(c) Increase in the number of tourists during school vacation

(d) Increased death rate during a certain epidemic

**10.** Which method of finding trend is best to eliminate the effect of repetitive short-term variations ?

(a) Graphical Method (b) Method of least squares

(c) Karl Pearson's method (d) Method of moving average

## Section B

**Answer the following questions in one sentence :**

**1.** Give an example of time series having decreasing trend.

**2.** What is a time series ?

**3.** Which of the components of time series produce short-term variations ?

**4.** What is meant by analysis of time series ?

**5.** What is the notation to show the cyclical component of the time series ?

**6.** State the names of methods of measuring trend.

**7.** The effect of which component indicates fluctuations repeating within one year ?

**8.** State the components of time series.

**9.** When is the method of moving average more useful to find trend ?

**10.** The linear equation fitted using the data of 7 weeks for a variable $y$ is $\hat{y} = 25.1 - 1.3\,t$. Estimate the value of $y$ for the eighth week.

## Section C

**Answer the following questions :**

**1.** Describe the additive model of time series.

**2.** What is meant by cyclical component ?

**3.** How does seasonal component differ from the cyclical component ?

**4.** Explain the irregular component.

**5.** State the limitations of graphical method.

**6.** Explain the meaning of moving average.

**7.** Define time series.

**8.** State the merits of the method of moving average to measure trend.

**9.** Describe the graphical method to measure trend.

## Section D

**Answer the following questions :**

**1.** Explain the importance of time series.

**2.** State the uses of analysis of time series.

**3.** What is meant by trend of a time series ? Explain with an illustration.

**4.** Write a short note on seasonal component.

**5.** Explain the method of fitting a linear equation to the given data using the method of least squares.

**6.** State the merits and limitations of the method of least squares.

**7.** Describe the method of moving average to find trend.

**8.** Discuss the limitations of the method of moving average.

**9.** The following time series shows the daily production of a factory. Find the trend using graphical method.

| **Day** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Production (units)** | 21 | 22 | 23 | 25 | 24 | 22 | 25 | 26 | 27 | 26 |

**10.** Fit a linear equation from the following data for variable ($y$) of a time series.

$n = 4, \quad \Sigma y = 270, \quad \Sigma ty = 734$

**11.** The data collected about the demand of a commodity from a store are as follows. Find the trend using three monthly moving averages.

| **Month** | January | February | March | April | May | June | July |
|---|---|---|---|---|---|---|---|
| **Demand (units)** | 15 | 16 | 18 | 18 | 23 | 23 | 20 |

## Section E

**Solve the following :**

**1.** The data about exports (in crore ₹) of ready-made garments of a textile manufacturer are shown below :

| Year | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|
| **Export (crore ₹)** | 22 | 25 | 23 | 26 | 20 | 25 |

Fit a linear trend to these data and estimate the trend for the export in the year 2017.

**2.** The following data are available for the number of passengers who travelled in the last 5 years by the aircrafts of an airline company. Estimate the trend for the year 2016 by fitting linear trend.

| Year | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|
| **No. of passengers (thousands)** | 45 | 47 | 44 | 40 | 38 |

**3.** The data about closing prices of shares of a company registered in a stock exchange for different months is given in the following table. Find the trend using three monthly moving averages.

| Month | 2015 April | May | June | July | August | Sept. | Oct. | Nov. | Dec. | 2016 January |
|---|---|---|---|---|---|---|---|---|---|---|
| **Share price (₹)** | 76 | 73 | 65 | 68 | 67 | 60 | 63 | 67 | 65 | 66 |

**4.** The following data show the sales (in thousand ₹) of a commodity. Find the trend by graphical method.

| Year | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sales (thousand ₹)** | 200 | 216 | 228 | 235 | 230 | 232 | 236 | 235 | 230 | 233 |

**5.** The quantity index numbers of consumption of edible oil in a state are given in the following table. Find the trend using five yearly moving averages.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Index No.** | 115 | 121 | 119 | 120 | 117 | 119 | 120 | 118 | 116 | 124 | 125 |

## Section F

**Solve the following:**

**1.** Find a linear equation using the method of least squares for the trend of production from the following data about sugar production of a country recorded for the last 6 years. Find the trend estimates for the production of the year 2016-17 and 2017-18.

| Year | 2009 – 10 | 2010 – 11 | 2011 – 12 | 2012 – 13 | 2013 – 14 | 2014 – 15 |
|---|---|---|---|---|---|---|
| **Sugar production (crore tons)** | 29.2 | 34.2 | 35.4 | 36.4 | 33.6 | 37.7 |

**2.** The number of students studying in a college are shown in the following table. Find the trend by four yearly moving averages.

| Year | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|
| **No. of students** | 332 | 317 | 357 | 392 | 402 | 405 | 410 | 427 | 405 | 438 |

**3.** The birth rates of a state in different years are given in the following table. Fit a linear trend for these data. Also find the estimates for birth rates in the year 2016 and 2017.

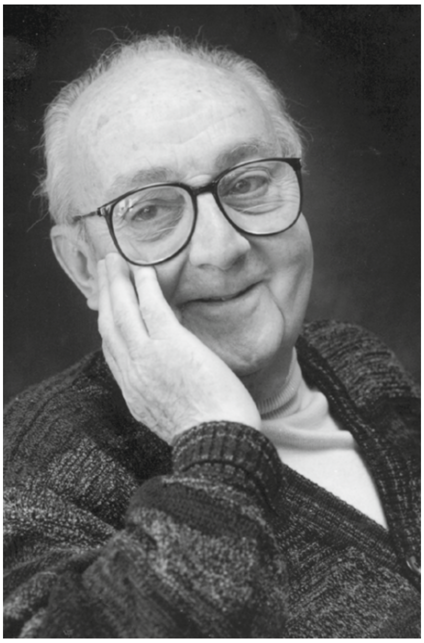
| **Year** | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|
| **Birth rate** | 22.2 | 21.8 | 21.3 | 20.9 | 20.6 | 20.2 | 19.9 |

**4.** The data about goods transported in different years by a division of railways are given below. Find the estimates for each year by fitting a linear equation and represent it by a graph. Also find the estimate for the year 2016.

| **Year** | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|
| **Goods transported (tons)** | 180 | 192 | 195 | 204 | 202 |

**5.** The data of weekly prices (in USD per barrel) of crude oil are given in the following table. Find the trend using four weekly moving averages.

| **Month** | March 2016 | | | | April 2016 | | | | May 2016 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Week** | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| **Price of Crude oil** | 35.92 | 38.50 | 39.44 | 39.46 | 36.79 | 39.72 | 40.36 | 43.73 | 45.92 | 44.66 | 46.21 | 48.45 |



**George Edward Pelham Box**
**(1919 -2013)**

George E. P. Box worked in the areas of quality control, time series analysis, design of experiments and Bayesian inference. He has been called "one of the greatest statistical minds of the 20th century." He has been associated with University at Raleigh (now North Carolina State University), Princeton University, University of Wisconsin–Mandison. Box has published numerous articles and papers and he is an author of many books. He is a recipient of prestigious honours, medals and was the president of American Statistical Association in 1978 and of the Institute of Mathematical Statistics in 1979. His name is associated with results in statistics such as Box–Jenkins models, Box–Cox transformations, Box–Behnken designs, and others. Box was elected a member of the American Academy of Arts and Sciences in 1974 and a Fellow of the Royal Society (FRS) in 1985.

●

# Answers

## Exercise 1.1

**1.** (1) Fixed base index numbers : 100, 103.27, 105.09, 106.55, 108, 113.82, 119.27, 125.45

(2) Chain base index numbers : 100, 103.27, 101.76, 101.38, 101.37, 105.39, 104.79, 105.18

(3) Index numbers using average wage : 91.36, 94.35, 96.01, 97.34, 98.67, 103.99, 108.97, 114.62

**2.** (1) Fixed base index numbers : 100, 101.79, 105.36, 107.14, 110.71, 114.29,121.43, 128.57

(2) Chain base index numbers : 100, 101.79, 103.51, 101.69, 103.33, 103.23, 106.25, 105.88

(3) Index numbers using average price : 96.55, 98.28, 101.72, 103.45, 106.90, 110.34, 117.24, 124.14

**3.** (1) Fixed base index numbers : 100, 108.70, 112.78, 115.19, 119.44

(2) Chain base general price index numbers : 100, 108.70, 103.65,102.26, 103.71

**4.** General index number of $n$ items : 126.45; Overall increase in the price of fuel items is 26.45 %

### Exercise 1.2

1. Fixed base index numbers : 100, 110, 104.5, 112.86, 135.43, 143.56, 157.92
2. Chain base index numbers : 117.4, 100.51, 102.80, 103.13, 102.64, 102.49, 102.28
3. Chain base index numbers : 100, 99.63, 99.26, 100, 103.73, 101.80, 100, 103.53, 100, 102.05
4. Fixed base index numbers : 110, 123.2, 134.29, 145.03, 152.28, 169.03

### Exercise 1.3

1. $I = 307$, prices have increased by 207 %.
2. $I = 123.80$, prices have increased by 23.80 %.
3. $I_L = 126.72$, $I_P = 126.85$, $I_F = 126.78$
4. $I_L = 141.13$, $I_P = 140.15$, $I_F = 140.64$ **5.** $I_F = 142.57$ **6.** $I_P = 115.2$, $I_F = 115.14$

### Exercise 1.4

1. Index number by family budget method = 135.64 and total expenditure has increased by 35.64 %. Average monthly disposable income = ₹ 20,346.
2. Index number $I = 128.53$ and rise in total expenditure is 28.53 %.
3. Index number $I = 132.51$ and rise in total expenditure is 32.51 %.
4. Index number $I = 213.20$ and rise in total expenditure is 113.20 %.
5. Index number by family budget method = 129.64 and by total expenditure method $I = 129.64$ Thus, both index numbers are same.

### Exercise 1

#### Section A

**1.** (c) **2.** (a) **3.** (d) **4.** (c) **5.** (d)

**6.** (d) **7.** (c) **8.** (c) **9.** (c) **10.** (c)

**11.** (a) **12.** (c)

#### Section B

**12.** The statement is false. Price index number of oil is 500.

#### Section C

**7.** Real wage ₹ 16,392.85 and loss to worker ₹ 1642.85 (Decrease in purchasing power)

**8.** Real wages ₹ 29166.67, 26666.67, 32307.69, 31250

**9.** Rate of inflation for year 2015 : 2.03 %

**10.** 449.55

**11.** Average monthly disposable income = ₹ 30,000

**12.** Index number of income = 125 **13.** Index number of production = 280 **14.** $I_p = 222.5$

## Section D

**7.** 161.87

**8.** Fixed base index numbers = 100, 111.11, 133.33, 144.44, 166.67, 222.22, 263.89

**9.** Chain base index numbers = 100, 104, 100.96, 102.86, 100.93, 116.51

**10.** Fixed base index numbers = 120, 108, 151.20, 189

**11.** Chain base index numbers = 100, 112.5, 106.67, 114.58, 109.09, 116.67

**12.** Index number = 226.6 **13.** $I_L = 166.67$, $I_p = 150$, $I_F = 158.12$ **14.** $I_p = 167.71$

## Section E

**1.** General index number = 122.32 **2.** Index number by total expenditure method = 149.41

**3.** Index number by total expenditure method = 115.69

**4.** Fixed base index numbers = 100, 118.75, 125, 131.25, 140.63, 187.5, 203.13;

Index numbers using average price = 91.43, 108.57, 114.29, 120, 128.57, 171.43, 185.71

**5.** Index number of industrial production I = 379.19

**6.** Index number I = 126.79 and rise in price is 26.79 %.

**7.** Real wages = 12,500, 10,000, 9268.29, 9090.91, 9361.7, 9615.38 Purchasing power of money = ₹ 0.38

## Section F

**1.** $I_L = 113.65$, $I_P = 113.94$, $I_F = 113.79$ and rise in price is 13.79 %.

**2.** $I_P = 191.53$, $I_F = 211.52$

**3.** $I_F = 84.84$

**4.** $I_L = 109.52$, $I_P = 110.29$, $I_F = 109.90$

**5.** Index number by family budget method = 118.58 and index number by total expenditure method = 118.58. Thus, both index numbers are same.

**6.** Index number for year 2014 $I_1 = 239.41$ and index number for year 2015 $I_2 = 253.44$. The rise in cost of living in the current year is 14.03 %. The percentage rise in the price index number is 5.86 % and rise in wage is 5 %. Hence, wage rise is 0.86 % less.

**7.** Index number $I = 231.44$ Income should be ₹ 13,886.40 to maintain earlier standard of living.

**8.** Index number of industrial production $= 100.10$, which indicates a rise of 0.10 % with respect to the base year.

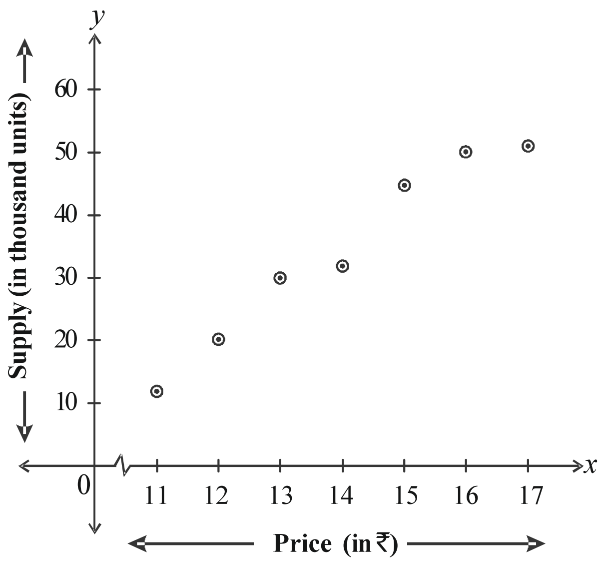
**9.** Index number $I = 128.75$.

**10.** Cost of living index number $= 196.35$ and the rise is $(196.35 - 100) = 96.35$ % as compared to the base year.

●

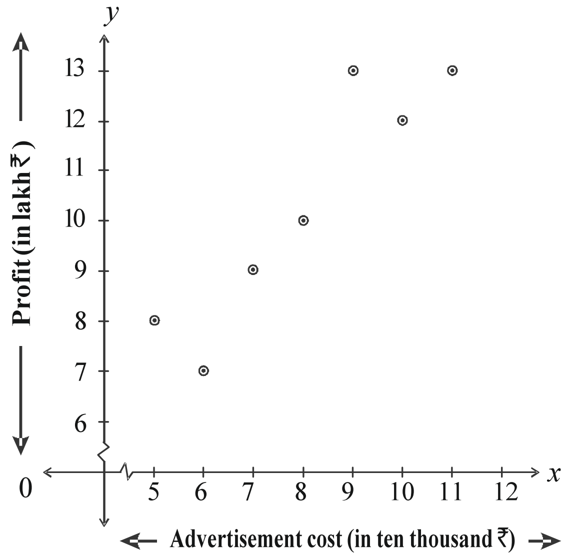
### Exercise 2.1

1.



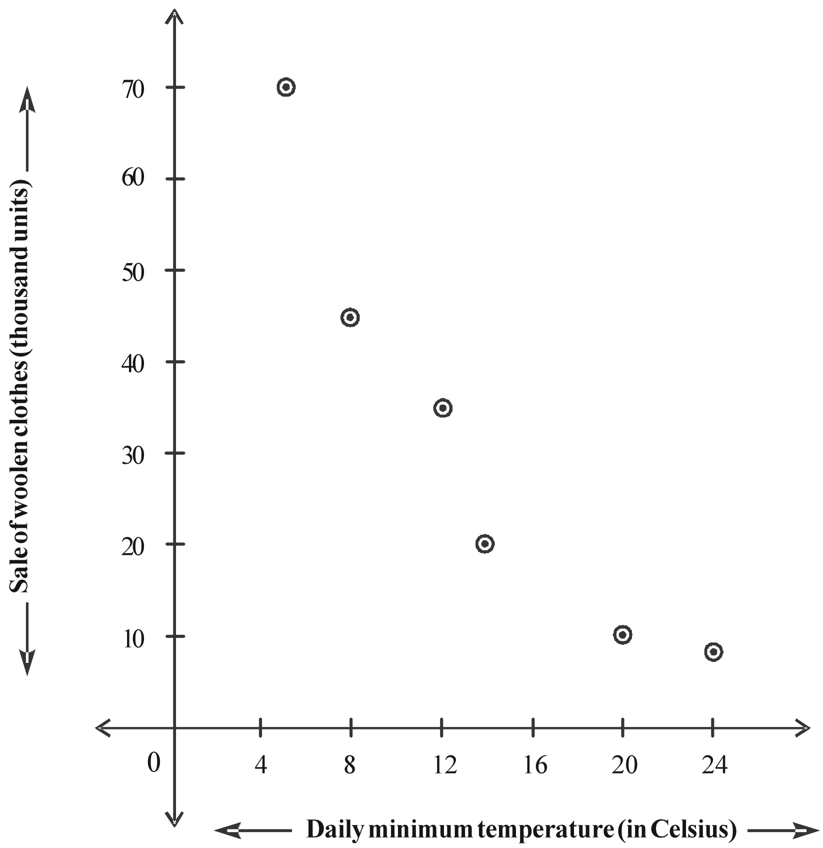

There is partial positive corelation between price and supply

2.




There is partial positive correlation between advertisement cost and profit

3.




There is partial negative correlation between daily minimum temperature and sale of woolen clothes

## Exercise 2.2

| | | | | |
|---|---|---|---|---|
| **1.** $r = 0.81$ | **2.** $r = -0.90$ | **3.** $r = 0.90$ | **4.** $r = 0.24$ | **5.** $r = 0.82$ |
| **6.** $r = -0.96$ | **7.** $r = 0.67$ | **8.** $r = -0.92$ | **9.** $r = 0.99$ | **10.** $r = 0.80$ |
| **11.** $r = 0.84$ | **12.** $r = 0.5$ | **13.** $r = 0.8$ | **14.** (1) $r = 0.94$ | (2) $r = 0.96$ |
| **15.** $r = -0.55$ | | | | |

## Exercise 2.3

| | | | | |
|---|---|---|---|---|
| **1.** $r = 0.49$ | **2.** $r = 0.78$ | **3.** $r = 0.7$ | **4.** $r = 0.82$ | **5.** $r = 0.91$ |
| **6.** $r = 0.90$ | **7.** $r = -0.30$ | **8.** Corrected $\Sigma d^2 = 82.5$, $r = 0.26$ | | |

## Exercise 2

### Section A

| | | | | |
|---|---|---|---|---|
| **1.** (c) | **2.** (a) | **3.** (d) | **4.** (b) | **5.** (c) |
| **6.** (d) | **7.** (b) | **8.** (b) | **9.** (b) | **10.** (c) |
| **11.** (a) | **12.** (b) | **13.** (c) | **14.** (c) | **15.** (a) |
| **16.** (b) | **17.** (b) | **18.** (a) | | |

### Section B

**3.** Positive **4.** Positive **5.** Negative **6.** Negative **7.** Nonsense correlation

**8.** $r$ remain unchanged due to change of origion, so $r = 0.4$ **10.** $r = 0$ **11.** Negative

### Section C

**11.** $r = 0.67$ **12.** $r = -0.54$ **13.** $r = 0.27$

### Section D

**10.** $r = 0.75$ **11.** $r = 0$ **12.** $r = -0.5$ **13.** $r = 0.2$

### Section E

| | | | | |
|---|---|---|---|---|
| **1.** $r = -0.81$ | **2.** $r = 0.43$ | **3.** $r = 0.79$ | **4.** $r = 0.77$ | **5.** $r = 0.54$ |
| **6.** $r = 0.13$ | | | | |

**Section F**

1. $r = 0.99$ 2. $r = -0.96$ 3. $r = 0.88$ 4. $r = 0.81$ 5. $r = 0.38$

6. $r = 0.79$ 7. $r = 0$ 8. $r = 0.6$ 9. $r = 0.3$ 10. $r = 0.79$

11. Corrected $\Sigma d^2 = 78$ ; $r = 0.53$ 12. $r = 0.73$

●

## Exercise 3.1

1. $\hat{y} = 31.44 - 1.34\,x$ and for price $x$ = 20 ₹, estimate of demand $\hat{y} = 4.64$ (hundred units)

2. $\hat{y} = 3.35 + 1.93\,x$ and for usage time of car $x$ = 5 year, Estimate of annual maintenance cost $\hat{y} = 13$ (thousand ₹)

   $\therefore$ Error $e = y - \hat{y} = 13 - 13 = 0$ (Here for $x$ = 5, the observed value of $y$ given in the table is 13)

3. $\hat{y} = 64.27 + 0.83\,x$ and for average rain $x$ = 35 cm, estimate of yield of crop $\hat{y} = 93.32$ (ton)

4. $\hat{y} = 69.7 + 1.13\,x$ and for experience of worker, $x$ = 7 year, estimate of performance index $\hat{y} = 77.61$

## Exercise 3.2

1. $\hat{y} = 54.84 + 2.52\,x$ and for 300 kg usage of fertilizer [$\therefore$ $x$ = 30 (ten kg.)], estimate of crop of cotton $\hat{y} = 130.44$ (Quintal per Hectare)

2. $\hat{y} = 52.84 + 0.68\,x$ and for a father's height $x$ = 170 cm, estimate of height of the son $\hat{y} = 168.44$ cm

3. $\hat{y} = 20.72 - 0.71\,x$ and for altitude $x$ = 7 thousand feat, estimate of effective Oxygen $\hat{y} = 15.75$ %

4. $\hat{y} = -3495.7 + 327.73\,x$ and for carpet area $x$ = 110 sq. meter estimated monthly rent $\hat{y} = 32554.6$ ₹

5. $\hat{y} = 0.53 + 0.02\,x$ and for $x$ = 80 customers, estimated sales $\hat{y} = 2.13$ (thousand ₹)

6. $\hat{y} = 7.6 + 0.29\,x$ ; $x$ = Profit (lakh ₹) and $y$ = Administrative cost (lakh ₹)

7. $\hat{y} = 53.72 + 1.54\,x$ and for rainfall $x$ = 60 cm, estimate of yield of corn $\hat{y} = 146.12$ Quintal

8. $\hat{y} = 8.74 + 1.02\,x$ and for price $x$ = 16 ₹, estimated supply $\hat{y} = 25.06$ (hundred units)

9. $\hat{y} = -4.8 + 0.15\,x$ and for maximum daily temperature $x$ = 42 celcius, estimate of sale of icecream $\hat{y} = 1.5$ (lakh ₹)

## Exercise 3

### Section A

| | | | | |
|---|---|---|---|---|
| **1.** (b) | **2.** (a) | **3.** (c) | **4.** (d) | **5.** (a) |
| **6.** (a) | **7.** (c) | **8.** (c) | **9.** (d) | **10.** (b) |
| **11.** (c) | **12.** (c) | **13.** (c) | **14.** (c) | **15.** (b) |

### Section B

**8.** Error = 0

**9.** Both variables are multiplied by 2 there fore $c_x = \frac{1}{2}$ and $c_y = \frac{1}{2}$. $\therefore$ Regression coefficient will not change

**10.** $b_{yx} = 0.5 \times \frac{4}{2} = 1$ **11.** $\hat{y} = 50$ **12.** $r = 1$ **13.** $r = -1$

### Section C

**2.** Error $e = 1$ **3.** $a = 2$ and $\hat{y} = 2 + 0.6\,x$

**4.** $b_{yx} = 5$ So, it can be said that because of increase of 1 unit in *x*, there is appropriate 5 units of increase in *y*.

**5.** $s_y = 3$ **6.** $R^2 = 1$ **7.** $s_x = 5$ **8.** 5 Units

**9.** $b_{yx} = 1.2$ and $a = 13$ **10.** $b_{vu} = b_{yx} \times \frac{c_x}{c_y} = 0.75 \times \frac{\frac{1}{6}}{\frac{1}{2}} = 0.25$

### Section D

**8.** $\hat{y} = 4 + 0.75\,x$ **9.** $\hat{y} = -10 + 2\,x$

**10.** $R^2 = 0.81$; 81 % variation of the total variation in *y*, can be explained by the regression model.

**11.** $b_{yx} = 2.52$ so it can be said that because of increase of 1 unit in *x*, there is approximate 2.52 units of increase in y.

**12.** (i) $b_{vu} = 0.8$ (ii) $b_{vu} = 1.6$ (iii) $b_{vu} = 0.08$ **13.** $\hat{y} = 12 + 0.88\,x$

### Section E

**1.** $\hat{y} = 2 + 0.75\,x$ **2.** $\hat{y} = 38.8 + 0.67\,x$ **3.** $\hat{y} = 58 + 3.2\,x$

**4.** $\hat{y} = 764.8 + 11.4\,x$ and for $x = 20$ cm, estimate of yield of crop is $\hat{y} = 992.8$ kg.

**5.** $\hat{y} = 18 + 0.8\,x$ and for $x =$ ₹ 45 lakh, estimate of market price is $\hat{y} = 54$ (thousand ₹)

## Section F

1. $\hat{y} = 73.29 - 1.59\,x$ and for price $x$ = 40 ₹, estimate of demand $\hat{y} = 9.69$ (hundred units)

2. $\hat{y} = 73.43 + 0.9\,x$ and for price $x$ = 17 years, the estimate of performance rating $\hat{y} = 88.73$

3. $\hat{y} = 34.8 + 0.74\,x$ and for daily income $x$ = 500 ₹, estimate of expenditure $\hat{y} = 404.8$ ₹

4. $\hat{y} = 3.73 + 0.13\,x$ and for advertisement cost $x = 50$ (ten thousand ₹), estimate of sales $\hat{y} = 10.23$ crore ₹

5. $\hat{y} = -122.94 + 91.67\,x$ and $R^2 = 0.97$ $\therefore$ Regression model is reliable.

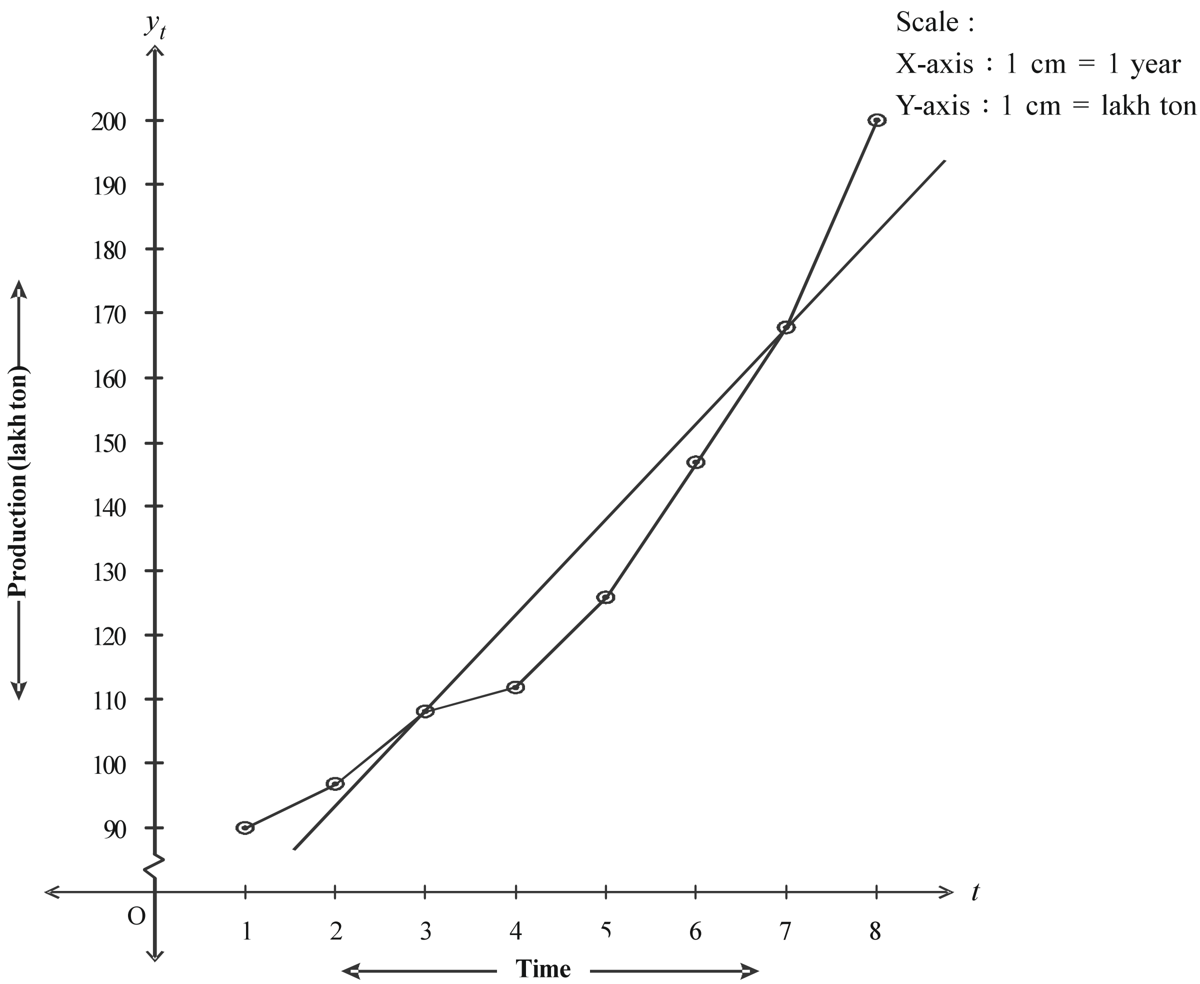
6. $\hat{y} = -10 + 1.6\,x$ and for $x$ = 30 $\hat{y} = 38$

7. $\hat{y} = -0.44 + 0.7\,x$ and for $x$ = 5, $\hat{y} = 3.06$

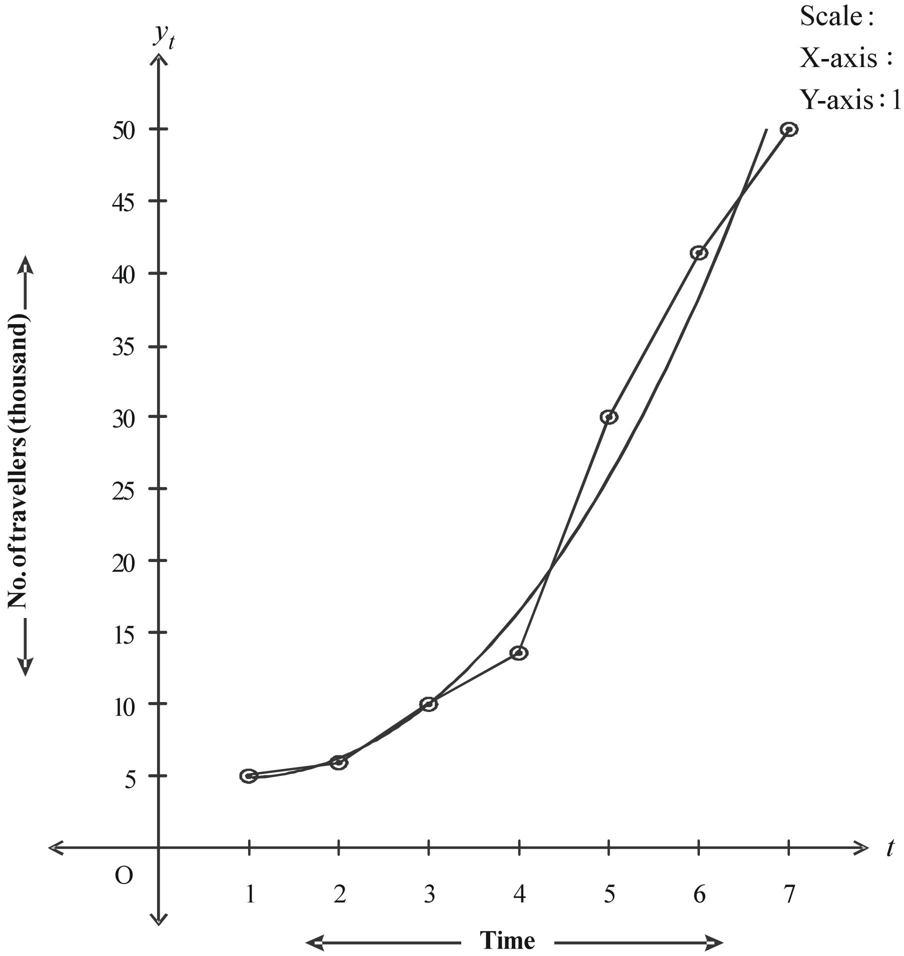
●

### Exercise 4.1

1.

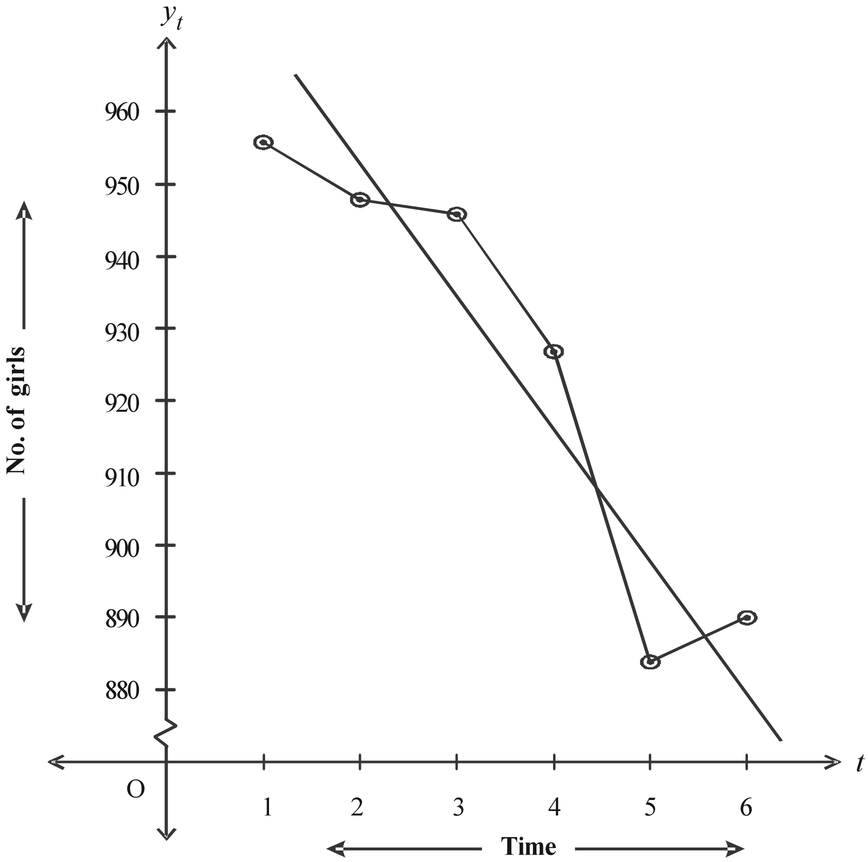


**2.**



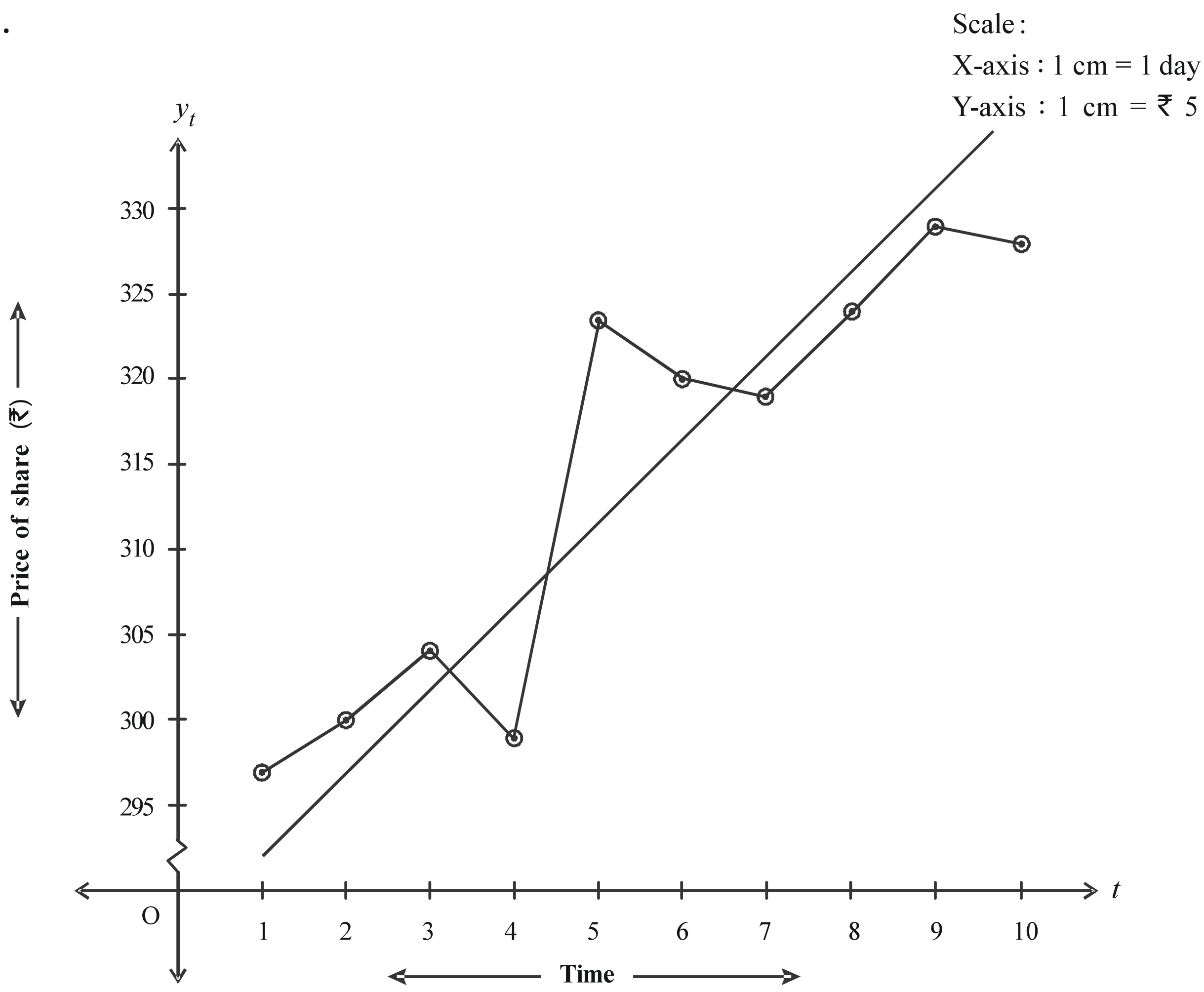

Scale :
X-axis : 1 cm = 1 year
Y-axis : 1 cm = 5 thousand travellers

**3.**




Scale :
X-axis : 1 cm = 10 year
Y-axis : 1 cm = 10 units

**4.**




**Exercise 4.2**

**1.** $\hat{y} = 7.41 - 0.07\,t$, for year 2017 $\hat{y} = 6.78$

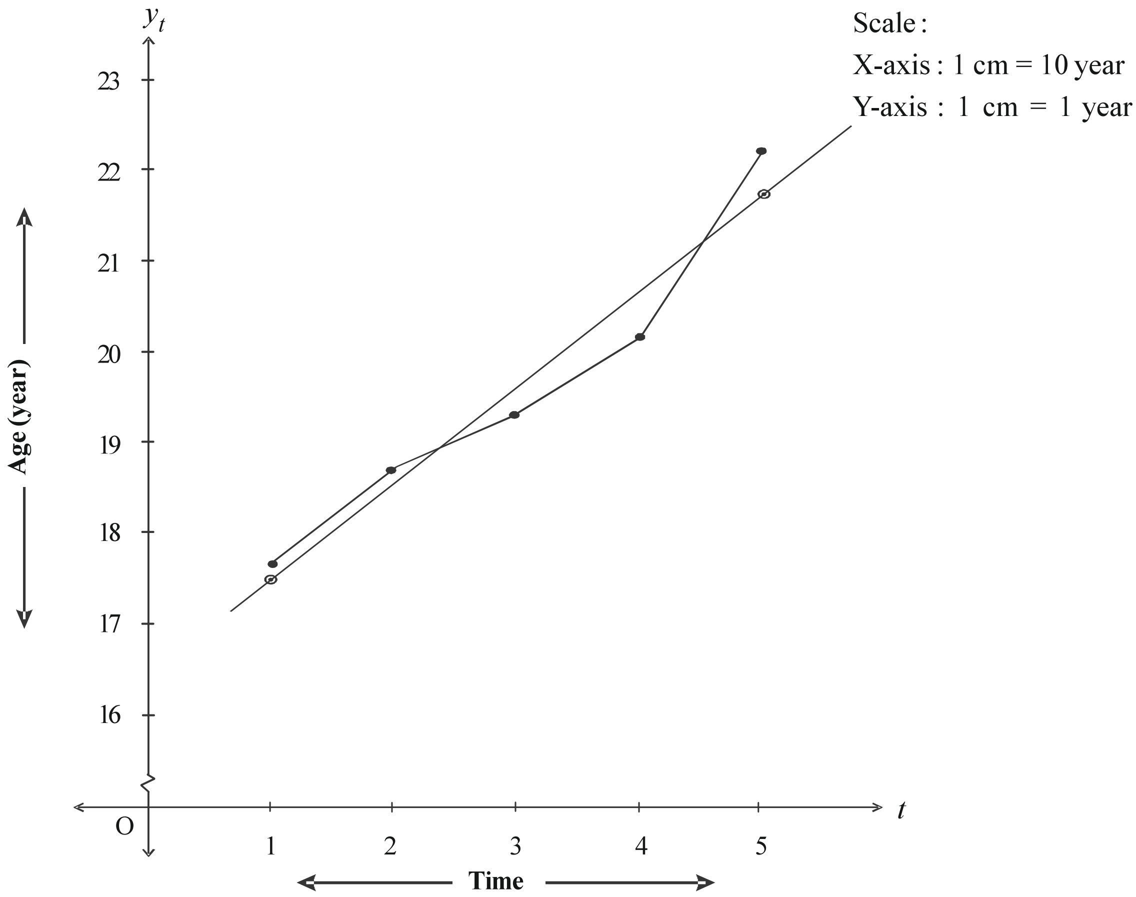
**2.** $\hat{y} = 447.2 + 69.4\,t$, for year 2015-16 $\hat{y} = 1071.8$

**3.** $\hat{y} = 57.12 + 9.06\,t$, $\hat{y} = 120.54$ thousand for year 2016

$\hat{y} = 129.6$ thousand for year 2017

| **Year** | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|
| **Estimated values of trend (thousand vehicles)** | 66.18 | 75.24 | 84.3 | 93.36 | 102.42 | 111.48 |

**4.** $\hat{y} = 16.47 + 1.05\,t$, $\hat{y} = 22.77$ years for year 2021




**Exercise 4.3**

**1.**

| **Year** | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Three yearly moving average** | – | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | – |

**2.**

| **Month** | Jan. | Feb. | March | April | May | June | July | August | Sept. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Four yearly moving average** | – | – | 274.88 | 280.38 | 273.88 | 263 | 265.38 | 269.25 | 272.63 | 275.88 | – | – |

**3.**

| **Year** | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|
| **Four yearly moving average** | – | – | 16.8 | 18.6 | 21.2 | 22.6 | 23.2 | – | – |

4.

| Year | 2013 | | | | 2014 | | | | 2015 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Quarter | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ |
| Four Quarterly moving average | – | – | 124.38 | 134 | 143 | 150.88 | 153.25 | 148 | 141.63 | 136.88 | – | – |

## Exercise 4

### Section A

**1.** (c) **2.** (a) **3.** (b) **4.** (b) **5.** (c)

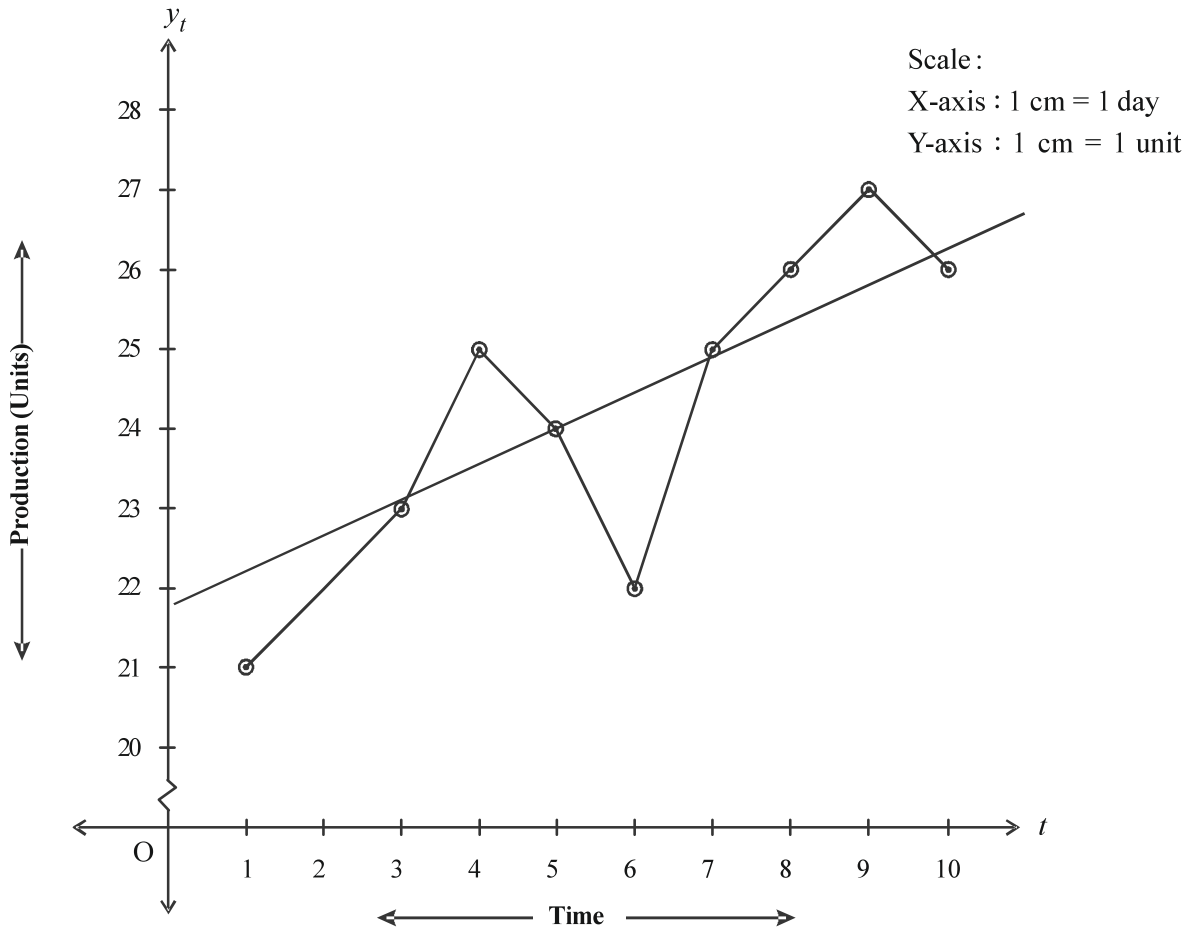
**6.** (a) **7.** (b) **8.** (c) **9.** (c) **10.** (d)

### Section B

**9.** $\hat{y} = 14.7$ for eighth week

### Section D

**9.**




**10.** $\hat{y} = 38 + 11.8\,t$

**11.**

| Month | January | February | March | April | May | June | July |
|---|---|---|---|---|---|---|---|
| Three monthly moving average | – | 16.33 | 17.33 | 19.67 | 21.33 | 22 | – |

## Section E

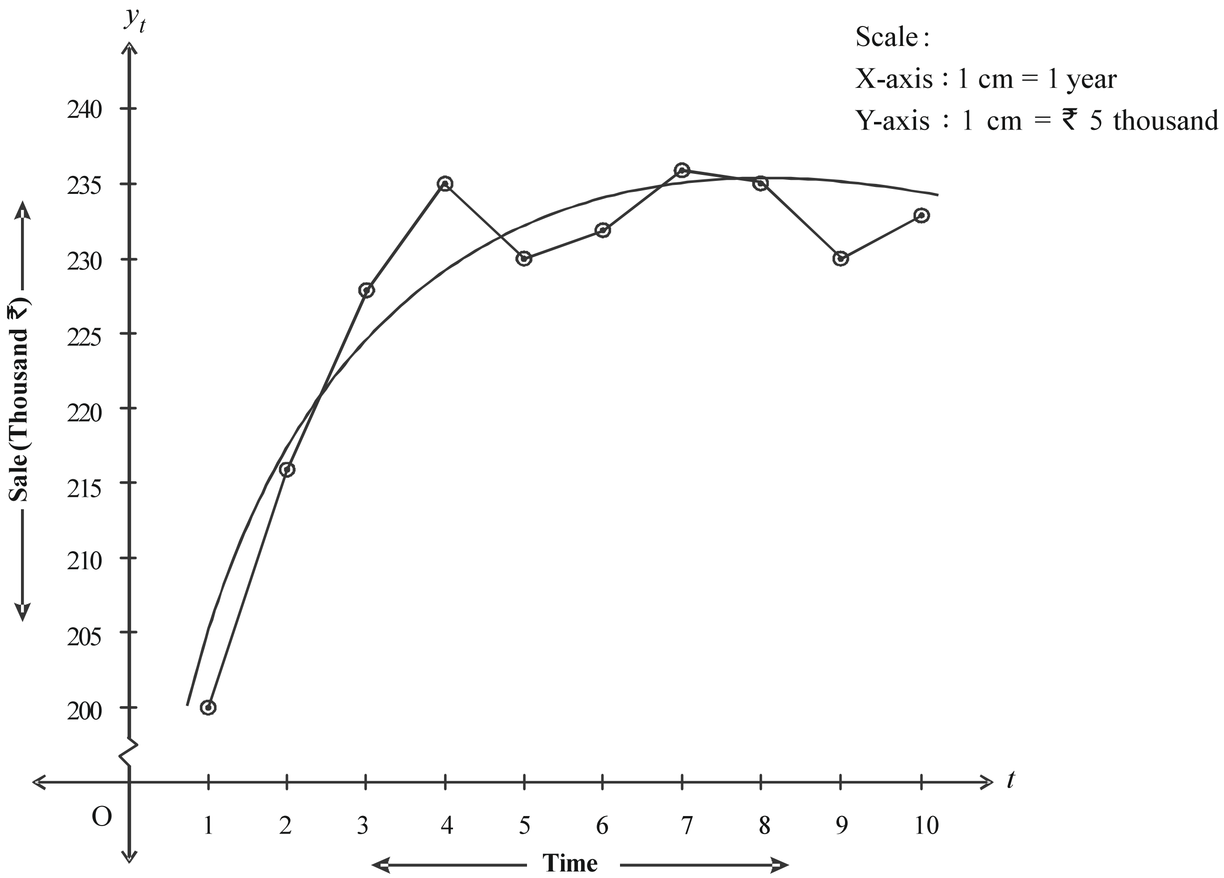
**1.** $\hat{y} = 23.18 + 0.09\,t$, $\hat{y} =$ 23.91 crore for year 2017

**2.** $\hat{y} = 49.1 - 2.1\,t$, $\hat{y} =$ 36.5 thousand for year 2016

**3.**

| Month | April 2015 | May | June | July | August | Sept. | Oct. | Nov. | Dec. | Jan. 2016 |
|---|---|---|---|---|---|---|---|---|---|---|
| Three monthly moving average | – | 71.33 | 68.67 | 66.67 | 65 | 63.33 | 63.33 | 65 | 66 | – |

**4.**




**5.**

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Five yearly moving average | – | – | 118.4 | 119.2 | 119 | 118.8 | 118 | 119.4 | 120.6 | – | – |

**Section F**

**1.** $\hat{y} = 30.26 + 1.19\,t$, $\hat{y} =$ 39.78 crore tons for year 2016-17

$\hat{y} =$ 40.97 crore tons for year 2017-18

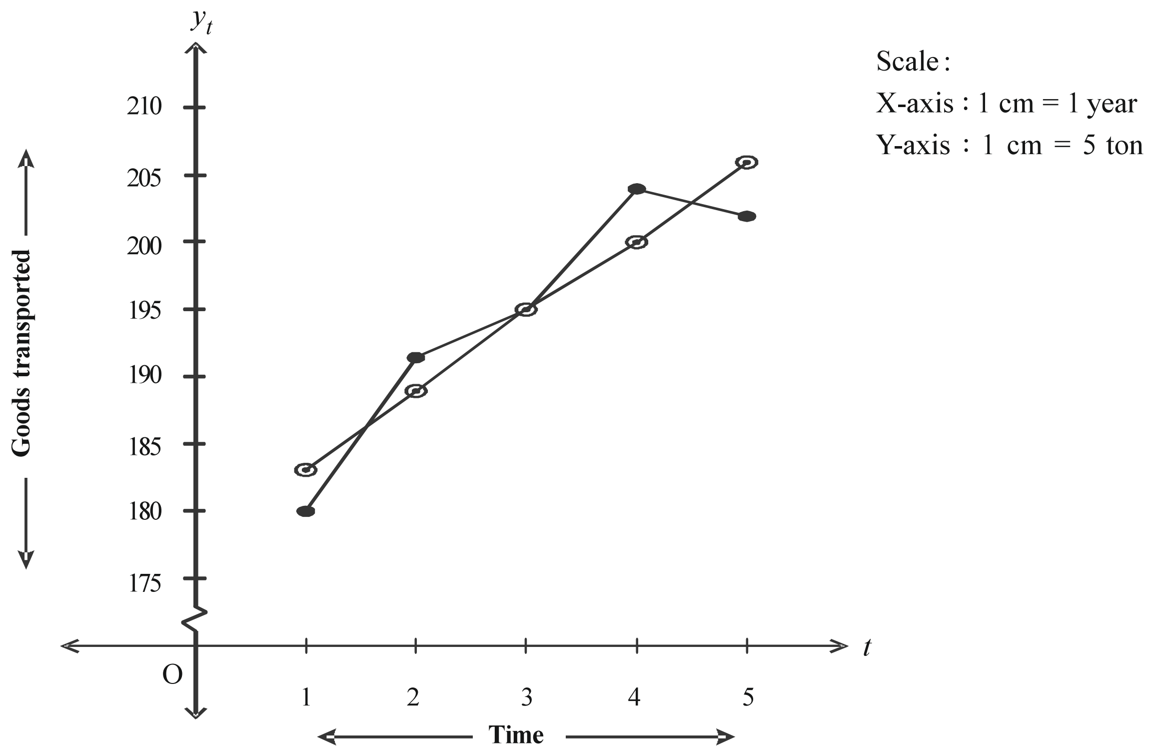
**2.**

| Year | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|---|---|---|---|---|
| Four yearly moving average | – | – | 358.25 | 378 | 395.63 | 406.63 | 411.38 | 415.88 | – | – |

**3.** $\hat{y} = 22.55 - 0.39\,t$, $\hat{y} =$ 19.43 for year 2016 $\hat{y} =$ 19.04 for year 2017

**4.** $\hat{y} = 177.8 + 5.6\,t$, $\hat{y} =$ 211.4 ton for year 2016

| Year | 2011 | 2012 | 2013 | 2014 | 2015 |
|---|---|---|---|---|---|
| Estimated value of trend (tons) | 183.4 | 189 | 194.6 | 200.2 | 205.8 |




**5.**

| Month | March | | | | April | | | | May | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Week | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Four weekly moving average | – | – | 38.44 | 38.7 | 38.97 | 39.62 | 41.29 | 43.05 | 44.4 | 45.72 | – | – |

●●●