2·1. INTRODUCTION

A set of ordered observations of a quantitative variable taken at successive points in time is known as '*Time Series*'. In other words, arrangement of statistical data in chronological order, *i.e.*, in accordance with occurrence of time, is known as '*Time Series*'. Time, in terms of years, months, days, or hours, is simply a device that enables one to relate all phenomenon to a set of common, stable reference points.

Such series have a unique important place in the field of Economic and Business Statistics since the data relating to prices, consumption and production of various commodities; money in circulation; bank deposits and bank clearings; sales and profits in a departmental store, agricultural and industrial production; national income and foreign exchange reserves; prices and dividends of shares in a stock exchange market, etc., are all time series data spread over a long period of time. A time series depicts the relationship between two variables, one of them being time, *e.g.*, the population (y_t) of a country in different years (t); temperature (y_t) of a place on different days (t), etc.

According to ***Ya-lun Chou,***

> "*A time series may be defined as a collection of readings belonging to different time periods, of some economic variable or composite of variables.*"

Mathematically, a time series is defined by the functional relationship

$$y_t = f(t)$$

where y_t is the value of the phenomenon (or variable) under consideration at time t. For example, (*i*) the population (y_t) of a country or a place in different years (t), (*ii*) the number of births and deaths (y_t) in different months (t) of the year, (*iii*) the sale (y_t) of a departmental store in different months (t) of the year, (*iv*) the temperature (y_t) of a place on different days (t) of the week, and so on, constitute time series. Thus , if the values of a phenomenon or variable at times $t_1, t_2, ..., t_n$ are $y_1, y_2, ..., y_n$ respectively, then the series

$$\begin{array}{ccccccc} t & : & t_1, & t_2, & t_3, & ..., & t_n \\ y_t & : & y_1, & y_2, & y_3, & ..., & y_n \end{array}$$

constitute a time series. Thus, a time series invariably gives a bivariate distribution, one of the two variables being time (t) and the other being the value (y_t) of the phenomenon at different points of time. The values of t may be given yearly, monthly, weekly, daily or even hourly, usually but not always at equal intervals of time.

If the data are segregated by time (days, months, years, etc.) the value of the variable under consideration changes from time to time. These fluctuations are affected not by a single force but are due to the net effect of multiplicity of forces pulling it up and down and if these forces were in a state of equilibrium, the series would remain constant. For example, the retail prices of a particular commodity are influenced by a number of factors, *viz.*, the crop yield which further depends on weather conditions, irrigation facilities, fertilizers used, transportation facilities, consumer demand, etc.

2·2. COMPONENTS OF TIME SERIES

The various forces at work, affecting the values of a phenomenon in a time series, can be broadly classified into the following four categories, commonly known as the *components of a time series,* some or all of which are present (in a given time series) in varying degrees.

(*a*) *Secular Trend or Long-term Movement.*

(*b*) *Periodic Changes or Short-term Fluctuations.*

(*i*) *Seasonal variations, and* (*ii*) *Cyclic variations.*

(*c*) *Random or Irregular Movements.*

The value of a time series y_t at any time t is regarded as the resultant of the combined impact of above components. In the following section we shall briefly explain them one by one.

Remarks **1.** Not all economists agree as to the classification of the components used here or as to the manner in which they are related. Some argue that there are more than four components, and some think that trend and cyclical movements are produced by the same set of forces.

2. During the past two or three decades, the attempt to make forecasting more scientific has led to the development of alternate approaches such as econometrics. *Econometrics* attempts to express economic theories in mathematical models that can be tested and verified by statistical methods. It tries to measure the impact of one economic variable upon another in the hope of being able to predict future events.

3. The approach in this chapter is to present the classical statistical approach to time-series analysis, but at the same time to point out that many other possible models exist which are based on different assumptions and which may lead to different results.

2·2·1. Trend. By secular trend or simply trend we mean the general tendency of the data to increase or decrease during a *long period* of time. This is true of most of series of Business and Economic Statistics. For example, an upward tendency would be seen in data pertaining to population, agricultural production, currency in circulation etc., while a downward tendency will be noticed in data of births and deaths, epidemics etc., as a result of advancement in medical sciences, better medical facilities, literacy and higher standard of living.

1. It may be clearly noted that trend is the *general, smooth, long-term, average tendency*. It is not necessary that the increase or decline should be in the same direction throughout the given period. It may be possible that different tendencies of increase, decrease or stability are observed in different sections of time. However, the overall tendency may be upward, downward or stable. Such tendencies are the result of the forces which are, more or less, constant for a long time or which change very gradually and continuously over a long period of time such as the change in the population, tastes, habits and customs of the people in a society and so on. They operate in an evolutionary manner and do not reflect sudden changes. For example, the effect of population increase over a long period of time on the expansion of various sectors like agriculture, industry, education, textiles, etc., is a continuous but a gradual process. Similarly, the growth or decline in a number of economic time series is the interaction of forces like advances in production technology, large-scale production, improved marketing management and business organization, the invention and discovery of new natural resources and the exhaustion of the existing resources and so on—all of which are gradual processes.

2. It should not be inferred that all the series must show an upward or downward trend. We might come across certain series whose values fluctuate round a constant reading which does not change with time, *e.g.*, the series of barometric readings or the temperature of a particular place.

3. *Linear and Non-linear (Curvi-linear) Trend.* If the time series values plotted on graph cluster more, or less, round a straight line, then the trend exhibited by the time series is termed as *Linear* otherwise *Non-linear* (curvi-linear). In a straight line trend, the time-series values increase or decrease more or less by a constant absolute amount, *i.e.*, the rate of growth (or decline) is constant. Although, in practice, linear trend is commonly used, it is rarely obtained in economic and business data. In an economic and business phenomenon, the rate of growth or decline is not of constant nature throughout but varies considerably in different sectors of time. Usually, in the beginning the growth is slow, then rapid which is further accelerated for quite sometime, after which it becomes stationary or stable for some period and finally retards slowly.

4. The term *'long period of time'* is a relative term and cannot be defined exactly. In some cases a period as small as a week may be fairly long while in some cases, a period as long as 2 years may not be enough. For example, if the data of agricultural production for 24 months shows an increase it won't be termed as secular change over a period of 2 years whereas if the count of bacterial population of a culture every five minutes, for a week shows an increase, then we would regard it as a secular change.

2·2·2. Periodic Changes. It would be observed that in many social and economic phenomena, apart from the growth factor in a time series there are forces at work which prevent the smooth flow of the series in a particular direction and tend to repeat themselves over a period of time. These forces do not act continuously but operate in a regular spasmodic manner. The resultant effect of such forces may be classified as :

(1) *Seasonal variations,* *and* (2) *Cyclic variations.*

(1) *Seasonal Variations.* These variations in a time series are due to the rhythmic forces which operate in a regular and periodic manner over a span of less than a year, *i.e.*, during a period of 12 months and have the same or almost same pattern year after year. Thus seasonal variations in a time series will be there if the data are recorded quarterly (every three months), monthly, weekly, daily, hourly, and so on. Although in each of the above cases, the amplitudes of the seasonal variations are different, all of them have the same period *viz.*, 1 year. Thus, in a time series data where only annual figures are given, there are no seasonal variations. Most of economic time series are influenced by seasonal swings, *e.g.*, prices, production and consumption of commodities; sales and profits in a departmental store; bank clearings and bank deposits, etc., are all affected by seasonal variations. The seasonal variations may be attributed to the following two causes :

(*i*) *Those resulting from natural forces.* As the name suggests, the various seasons or weather conditions and climatic changes play an important role in seasonal movements. For instance, the sale of umbrellas pick up very fast in rainy season; the demand for electric fans goes up in summer season; the sale of ice and ice-cream increases very much in summer; the sales of woollens go up in winter—all being affected by natural forces, *viz.*, weather or seasons. Likewise, the production of certain commodities such as sugar, rice, pulses, eggs, etc., depends on seasons. Similarly, the prices of agricultural commodities always go down at the time of harvest and then pick up gradually.

(*ii*) *Those resulting from man-made conventions.* These variations in a time series within a period of 12 months are due to habits, fashions, customs and conventions of the people in the society. For instance, the sale of jewellery and ornaments goes up in marriages; the sales and profits in departmental stores go up considerably during marriages, and festivals like Diwali, Dussehra (Durga Pooja), Christmas, etc. Such variations operate in a regular spasmodic manner and recur year after year.

The main objective of the measurement of seasonal variations is to isolate them from the trend and study their effects. A study of the seasonal patterns is extremely useful to businessmen, producers, sales managers, etc., in planning future operations and in formulation of policy decisions regarding purchase, production, inventory control, personnel requirements, selling and advertising programmes. In the absence of any knowledge of seasonal variations, a seasonal upswing may be mistaken as indicator of better business conditions while a seasonal slump may be mis-interpreted as deteriorating business conditions. Thus, to understand the behaviour of the phenomenon in a time series properly, the time series data must be adjusted for seasonal variations. [This is done by isolating them from trend and other components by dividing the given time series values (y_t) by the seasonal variations (S_t). This technique is called *de-seasonalisation* of data and is discussed in detail later [See § 2·5·4].

(2) *Cyclic Variations.* The oscillatory movements in a time series with period of oscillation more than one year are termed as cyclic fluctuations. One complete period is called a *'cycle'*. The cyclic movements in a time series are generally attributed to the so-called *'Business Cycle'*, which may also be referred to as the 'four-phase cycle' composed of prosperity (period of boom), recession, depression and recovery, and normally lasts from seven to eleven years. The upswings and downswings in business depend upon the cumulative nature of the economic forces (affecting the equilibrium of demand and supply) and the interaction between them. Most of the economic and commercial series, *e.g.*, series relating to prices, production and wages, etc., are affected by business cycles. Cyclic fluctuations, though more or less regular, are not periodic.

2·2·3. Irregular (or Random) Component. Apart from the regular variations, almost all the series contain another factor called the *random or irregular or residual fluctuations*, which are not accounted for by secular trend and seasonal and cyclic variations. These fluctuations are purely random, erratic, unforeseen, unpredictable and are due to numerous non-recurring and irregular circumstances which are beyond the control of human hand but at the same time are a part of our system such as earthquakes, wars, floods, famines, revolutions, epidemics, etc. These isolated or irregular but powerful fluctuations due to floods, revolution, political upheavals, famines, etc., are also called *episodic fluctuations*. In some cases the importance of irregular fluctuations may not be significant while in others these may be very effective and might give rise to cyclic movements.

Remark. It may be noted that because of their absolutely random character, it is not possible to isolate irregular variations and study them exclusively, nor one can forecast or estimate them precisely. Only rough estimates from past experience can be obtained and accordingly one may make some provisions for such abnormalities.

2·3. ANALYSIS OF TIME SERIES

The main problems in the analysis of time series are :

(*i*) To identify the forces or components at work, the net effect of whose interaction is exhibited by the movement of a time series, and

(*ii*) To isolate, study, analyse and measure them independently, *i.e.*, by holding other things constant.

2·3·1. Mathematical Models for Time Series. The following are the two models commonly used for the decomposition of a time series into its components.

***(i) Decomposition by Additive Hypothesis* (or Additive Model).** According to the *additive model*, a time series can be expressed as

$$y_t = T_t + S_t + C_t + R_t \qquad \dots (2\cdot1)$$

where y_t is the time-series value at time t, T_t represents the trend value, S_t, C_t and R_t represent the seasonal, cyclic and random fluctuations at time t. Obviously, the term S_t will not appear in a series of annual data. The additive model implicitly implies that seasonal forces (in different years), cyclical forces (in different cycles) and irregular forces (in different long term period) operate with equal absolute effect irrespective of the trend value. As such C_t (and S_t) will have positive or negative values, according as whether we are in an above normal or below normal phase of the cycle (and year) and the total of positive and negative values for any cycle (and any year) will be zero. R_t will also have positive or negative value and in the long-term (ΣR_t) will be zero. Occasionally, there may be a few isolated occurrences of extreme R_t of episodic nature.

The additive model assumes that all the four components of the time series operate independently of each other so that none of these components has any effect on the remaining three.

***(ii) Decomposition by Multiplicative Hypothesis* (*or Multiplicative Model*).** On the other hand, if we have reasons to assume that the various components in a time series *operate proportionately* to the general level of the series, the traditional or classical multiplicative model is appropriate. According to the multiplicative model,

$$y_t = T_t \times S_t \times C_t \times R_t \qquad \dots (2\cdot2)$$

where S_t, C_t and R_t , instead of assuming positive and negative value, are indices fluctuating above or below unity and the geometric means of S_t in a year, C_t in a cycle and R_t in a long-term period are unity. In a time series with both positive end negative values, the multiplicative model can not be applied unless the time series is translated by adding a suitable positive value. It may be pointed out that the multiplicative decomposition of a time series is same as the additive decomposition of logarithmic values of the original time series, *i.e.*,

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

In practice, most of the series relating to economic data conform to multiplicative model

***Remarks* 1.** *Limitations of the Hypothesis of Decomposition of a Time Series.* Hypothesis of decomposition presupposes that the trend and periodic components are determined by separate forces acting independently so that simple aggregation of the components could constitute the series. But in reality, it is possible that this year's value of the series will depend to some extent on last year's value so that trend and periodic movement will get inextricably mixed up and no meaningful separation of them will be possible. In such a case any variations of this year may affect the whole future course of the series and no meaningful separation of trend and periodic components will be possible.

2. Mixed Models. In addition to the additive and multiplicative models discussed above, the components in a time series may be combined in a large number of other ways. The different models, defined under different assumptions will yield different results. Some of the mixed models resulting from different combinations of additive and multiplicative models are given below :

$$\left.\begin{aligned} y_t &= T_t C_t + S_t R_t \\ y_t &= T_t + S_t C_t R_t \\ y_t &= T_t + S_t + C_t R_t \end{aligned}\right\} \qquad \dots (2\cdot2a)$$

3. The model (2·1) or (2·2) can be used to obtain a measure of one or more of the components by elimination, *viz.*, subtraction or division. For example, if trend component (T_t) is known, then using multiplication model, it can be isolated from the given time series to give :

$$S_t \times C_t \times R_t = \frac{y_t}{T_t} = \frac{\text{Original values}}{\text{Trend values}} \qquad \dots (2\cdot2b)$$

Thus, for the *annual data*, for which the seasonal component S_t is not there, we have

$$y_t = T_t \times C_t \times R_t \quad \Rightarrow \quad C_t \times R_t = \frac{y_t}{T_t} \qquad \dots (2\cdot2c)$$

2·3·2. Uses of Time Series. The time series analysis is of greater importance not only to businessman or an economist but also to people working in various disciplines in natural, social and physical sciences. Some of its uses are enumerated below :

1. It enables us to *study the past behaviour of the phenomenon* under consideration, *i.e.*, to determine the type and nature of the variations in the data.

2. The segregation and study of the various components is of paramount importance to a businessman in the planning of future operations and in the formulation of executive and policy decisions.

3. It helps to compare the actual current performance of accomplishments with the expected ones (on the basis of the past performances) and analyse the causes of such variations, if any.

4. It enables us to predict or estimate or forcast the behaviour of the phenomenon in future which is very essential for business planning.

5. It helps us to compare the changes in the values of different phenomenon at different times or places, etc.

In the following sections we shall discuss various techniques for the measurement of different components

2·4. MEASUREMENT OF TREND

Trend can be studied and/or measured by the following methods :

(*i*) Graphic (or *Free-hand* Curve Fitting) Method,
(*ii*) Method of *Semi-Averages*,
(*iii*) Method of Curve Fitting by *Principle of Least Squares*, and
(*iv*) Method of *Moving Averages*.

We shall now discuss each of these methods in detail.

2·4·1. Graphic Method. A free-hand smooth curve obtained on plotting the values y_t against 't' enables us to form an idea about the general 'trend' of the series. Smoothing of the curve eliminates other components, *viz*. regular and irregular fluctuations.

This method does not involve any complex mathematical techniques and can be used to describe all types of trend, linear and non-linear. Thus, simplicity and flexibility are strong points of this method. Its main drawbacks are :

(*i*) The method is very subjective, *i.e.*, the bias of the person handling the data plays a very important role and as such different trend curves will be obtained by different persons for the same set of data. As such 'trend by inspection' should be attempted only by skilled and experienced statisticians and this limits the utility and popularity of the method.

(*ii*) It does not enable us to measure trend.

2·4·2. Method of Semi-averages. In this method, the whole data is divided into two parts with respect to time, *e.g.*, if we are given y_t for t from 1991-2002, *i.e.*, over a period of 12 years, the two equal parts will be the data from 1991 to 1996 and 1997 to 2002. In case of odd number of years the two parts are obtained by omitting the value corresponding to the middle year, *e.g.*, for the data from 1991-2001, the value corresponding to middle year, *viz.* 1996 being omitted. Next we compute the arithmetic mean for each part and plot these two averages (means) against the mid-values of the respective time-periods covered by each part. The line obtained on joining these two points is the required trend line and may be extended both ways to estimate intermediate or future values.

Remark. For even number of years like 8, 12, 16, etc. the centering of average of each part would create problems, *e.g.*, from the data 1997-2002 (n = 12), let the two averages be $\overline{X}_1$, (say) for period 1991-1996 and $\overline{X}_2$ (say), for the period 1991-2002. Here $\overline{X}_1$, will be plotted against the mean of two mid-values, *viz.* 1993 and 1994 for the period 1991-1996, *i.e.*, against 1st July 1993. Similarly, for the period 1997-2002.

***Merits* 1.** As compared with graphic method, the obvious advantage of this method is its objectivity in the sense that everyone who applies it would get the same results. Moreover, we can also estimate the trend values.

2. It is readily comprehensible as compared to the 'method of least squares' or the 'moving average method'.

Limitations. This method assumes linear relationship between the plotted points — which may not exist. Moreover, the limitations of arithmetic mean as an average also stand in its way.

Example 2·1. *Fit a trend line to the following data by the method of semi-averages :*

Year	*Bank Clearances (Rs. Crores)*	*Year*	*Bank Clearances (Rs. Crores)*
1992	*53*	*1999*	*87*
1993	*79*	*2000*	*79*
1994	*76*	*2001*	*104*
1995	*66*	*2002*	*97*
1996	*69*	*2003*	*92*
1997	*94*	*2004*	*101*
1998	*105*		

Solution. Here since n = 13 (odd), the two parts would consist of 1992 to 1997 and 1999 to 2004, the year 1998 being omitted.

$\overline{X}_1$ = Average sales for first part

$= \dfrac{437}{6} = 72{\cdot}83$ (Rs. crores)

$\overline{X}_2$ = Average sales for second part

$= \dfrac{560}{6} = 93{\cdot}33$ (Rs. crores)

As explained in Remark to § 2·4·2, $\overline{X}_1$ and $\overline{X}_2$ will be plotted against 1st July 1994 and 1st July 2001 respectively, as given in *Fig. 2·1*

Joining the points A [1994, $\overline{X}_1$] and B [2001, $\overline{X}_2$], we get the trend line [*Fig. 2·1*].

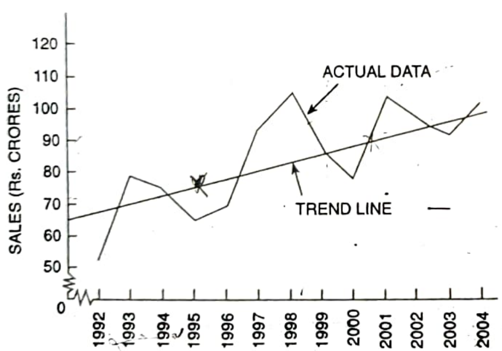

Fig. 2.1 : Trend by the Method of Semi-averages

2·4·3. Method of Curve Fitting by Principle of Least Squares. The principle of least squares is the most popular and widely used method of fitting mathematical functions to a given set of data. The method yields very correct results if sufficiently good appraisal of the form of the function to the fitted is obtained either by a scrutiny of the graphical plot of the values over time or by a theoretical understanding of the mechanism of the variable change. An examination of the plotted data often provides an adequate basis for deciding upon the type of trend to use. Apart from the usual arithmetic scales, semi-logarithmic or doubly-logarithmic scales may be used for the graphical representation of the data. The various types of curves that may be used to describe the given data in practice are :

(If y_t is the value of the variable corresponding to time t)

(*i*) *A straight line :* $\quad y_t = a + bt$

(*ii*) *Second degree parabola :* $\quad y_t = a + bt + ct^2$

(*iii*) *kth-degree polynomial :* $\quad y_t = a_0 + a_1 t + a_2 t^2 + \ldots + a_k t^k$

(*iv*) *Exponential curves :* $\quad y_t = a\, b^t$

$$\Rightarrow \quad \log y_t = \log a + t \log b = A + Bt, \text{ (say)}.$$

(*v*) *Second degree curve fitted to logarithms :*

$$y_t = a\, b^t c^{t^2}$$

$$\Rightarrow \quad \log y_t = \log a + t \log b + t^2 \log c = A + B\,t + C\,t^2, \text{ (say)}.$$

(*vi*) *Growth curves :*

(*a*) $\quad y_t = a + b\,c^t \quad$ (Modified Exponential Curve)

~~(*b*)~~ $\quad \cancel{y_t = ab^{c^t}} \quad$ ~~(Gompertz curve)~~

$$\Rightarrow \quad \log y_t = \log a + c^t \log b = A + Bc^t \text{ (say)}$$

(*c*) $\quad y_t = \dfrac{k}{1 + \exp\,(a + bt)} \quad$ (*Logistic curve*)

Remark. For deciding about the type of trend to be fitted to a given set of data, the following points may be helpful :

(*i*) When the time series is found to be increasing or decreasing by equal absolute amounts, the straight line trend is used. In this case, the plotting of the data will give a straight line graph.

(*ii*) The logarithmic straight line (exponential curve $y_t = ab^t$) is used as an expression of the secular movement, when the series is increasing or decreasing by a constant percentage rather than a constant absolute amount. In this case, the data plotted on a semi-logarithmeic scale will give a straight line graph.

(*iii*) Second degree curve fitted to logarithms may be tried for trend fitting if the data plotted on a semi-logarithmic scale is not a straight lone graph but shows curvature, being concave either upward or downward.

Alternatively, approximations about the type of the curve to be fitted can be made by use of the following theorem based on finite differences :

"The nth differences $\Delta^n y_t$, $\Delta^n (\log y_t)$, $\Delta^n (1/y_t)$ of any general polynomial y_t of nth degree in t is constant and (n + 1)th differences are equal to zero."

For further guidelines, the following statistical tests based on the calculus of finite differences may be applied.

We know that for a polynomial y_t of nth degree in t,

$$\left.\begin{aligned} \Delta^r y_t &= \text{constant}, & r &= n \\ &= 0, & r &> n \end{aligned}\right\}$$

where Δ is the difference operator given by $\Delta y_t = y_{t+h} - y_t$, h being the interval of differencing and $\Delta^r y_t$ is the rth differences of y_t.

1. If Δy_t = constant, use straight line trend.
2. If $\Delta^2 y_t$ = constant, use a second degree (parabolic) trend.
3. If $\Delta(\log y_t)$ = constant, use exponential trend curve.
4. If $\Delta^2(\log y_t)$ = constant, use second degree curve fitted to logarithms.
5. The growth curves, *viz.*, modified exponential, Gompertz and Logistic curves can be approximated by the constancy of the ratios :

$$\frac{\Delta y_t}{\Delta y_{t-1}}, \quad \left\{\frac{\Delta(\log y_t)}{\Delta(\log y_{t-1})}\right\}, \quad \left\{\frac{\Delta(1/y_t)}{\Delta(1/y_{t-1})}\right\}$$

respectively, for all possible values of t.

The following tests may also be used :

6. If Δy_t tends to decrease by a constant percentage, use modified exponential curve.
7. if Δy_t resembles a skewed frequency curve, use a Gompertz curve or Logistic curve.

Fitting of Straight Line by Least Squares Method. Let the straight line trend between the given time-series values (y_t) and time t be given by the equation :

$$y_t = a + bt \qquad \ldots (2\cdot3)$$

Principle of least squares consists in minimizing the sum of squares of the deviations between the given values of y_t and their estimates given by (2·3). In other words, we have to find a and b such that for given values of y_t corresponding to n different values of t,

$$E = \sum_t (y_t - a - bt)^2$$

is minimum. For a maxima or minima of E, for variations in a and b, we should have

$$\left.\begin{aligned}\frac{\partial E}{\partial a} &= 0 = -2\sum(y_t - a - bt)\\ \frac{\partial E}{\partial b} &= 0 = -2\sum t\,(y_t - a - bt)\end{aligned}\right\} \Rightarrow \left.\begin{aligned}\sum y_t &= na + b\sum t\\ \sum t_t &= a\sum t + b\sum t^2\end{aligned}\right\}, \qquad \ldots (2\cdot4)$$

which are the normal equations for estimating a and b.

The values of $\sum y_t$, $\sum t$, $\sum t^2$ are obtained from the given data and the equations (2·4) can now be solved for a and b. With these values of a and b, the line (2·3) gives the desired trend line.

Remark. The solution of normal equations (2·4) provides a minima of E. The proof is given below : The necessary and sufficient condition fo a minima of E for variations in a and b are :

$$(i)\ \frac{\partial E}{\partial a} = 0,\ \frac{\partial E}{\partial b} = 0 \ \ldots (*) \quad \text{and} \quad (ii)\ \Delta = \begin{vmatrix}\dfrac{\partial^2 E}{\partial a^2} & \dfrac{\partial^2 E}{\partial a\,\partial b}\\ \dfrac{\partial^2 E}{\partial b\,\partial a} & \dfrac{\partial^2 E}{\partial b^2}\end{vmatrix} > 0 \ \text{and}\ \frac{\partial^2 S}{\partial a^2} > 0 \qquad \ldots (**)$$

From (2·4), we get

$$\frac{\partial^2 S}{\partial a^2} = 2n > 0 \ ; \ \frac{\partial^2 S}{\partial b^2} = 2\sum t^2 > 0 \ ; \ \frac{\partial^2 S}{\partial a\,\partial b} = \frac{\partial^2 E}{\partial b\,\partial a} = 2\sum t$$

$$\therefore \quad \Delta = \begin{vmatrix} 2n & 2\sum t\\ 2\sum t & 2\sum t^2\end{vmatrix} = 4\left[n\sum t^2 - (\sum t)^2\right]$$

$$= 4n^2\left[\frac{\sum t^2}{n} - \left(\frac{\sum t}{n}\right)^2\right] = 4n^2\,\mathrm{Var}(t) > 0$$

Hence, the solution of the least square equations (2·4), satisfies (*) and (**) and, therefore, provides a minima of E.

Fitting of Second Degree (Parabolic) Trend. Let the second degree parabolic trend curve be :

$$y_t = a + bt + ct^2 \qquad \ldots (2\cdot5)$$

Proceeding similarly as in the case of a straight line, the normal equations for estimating a, b and c are given by :

$$\left.\begin{aligned}\sum y_t &= na + b\sum t + c\sum t^2\\ \sum t\,y_t &= a\sum t + b\sum t^2 + c\sum t^3\\ \sum t^2 y_t &= a\sum t^2 + b\sum t^2 + c\sum t^4\end{aligned}\right\}, \qquad \ldots (2\cdot6)$$

the summation being taken over the values of the time series.

Fitting of Exponential Curve :

$$y_t = a\,b^t \qquad \ldots (2\cdot7)$$

$$\Rightarrow \quad \log y_t = \log a + t\log b$$

$$\Rightarrow \quad Y = A + Bt \text{ (say)}, \qquad \ldots (2\cdot7a)$$

where $\quad Y = \log y_t \ , \ A = \log a \ , \ B = \log b. \qquad \ldots (2\cdot7b)$

(2·7*a*) is a straight line in t and Y and thus the normal equations for estimating A and B are

$$\left.\begin{aligned}\sum Y &= nA + B\sum t,\\ \sum tY &= A\sum t + B\sum t^2\end{aligned}\right\} \qquad \ldots (2\cdot7c)$$

These equations can be solved for A and B and finally on using (2·7*b*), we get

$$a = \text{antilog}(A)\ ;\ b = \text{antilog}(B).$$

Second Degree Curve Fitted to Logarithms. Suppose the trend curve is :

$$Y_t = a\,b^t c^{t^2} \qquad \ldots (2\cdot8)$$

Taking logarithms of both sides, we get

$$\log y_t = \log a + t\log b + t^2 \log c$$

$$\Rightarrow \quad Y_t = A + Bt + Ct^2 \qquad \ldots (2\cdot8a)$$

where $\quad Y_t = \log y_t \ ; \ A = \log a \ ; \ B = \log b$ and $C = \log c \qquad \ldots (2\cdot8b)$

Now, (2·8*a*) is a second degree parabolic curve in Y_t and t and can be fitted by the technique already explained. We can finally obtain

$$a = \text{Antilog}(A) \ ; \ b = \text{Antilog}(B) \ \text{and} \ c = \text{Antilog}(C).$$

With these values of a, b and c, the curve (2·8) becomes the best second degree curve fitted to logarithms.

Remark. The method of curve fitting by the principle of least squares is used quite often in trend analysis particularly when one is interested in making projections for future times. Obviously, the reliability of the estimated (projected) values primarily depends upon the appropriatness of the form of the mathematical function fitted to the given data. If the function is determined on the ad-hoc basis by the scrutiny of the plotted values, the projections based on it may be valid for the near future while, if the study of physical mechanism of the variable change forms the basis of the selection of function, then there is very little likelihood that the function will change for sufficiently long period and hence in this case reliable long term projections can be made.

Merits and Drawbacks of Trend Fitting by the Principle of Least Squares.

Merits. The method of least squares is the most popular and widely used method of fitting mathematical functions to a given set of observations. It has the following advantages :

1. Because of its mathematical or analytical character, this method completely eliminates the element of subjective judgement or personal bias on the part of the investigator.

2. Unlike the method of moving averages [discussed in § 2·4·5], this method enables us to compute the trend values for all the given time periods in the series.

3. The trend equation can be used to estimate or predict the values of the variable for any period t in future or even in the intermediate periods of the given series and the forecast values are also quite reliable.

4. The curve fitting by the principle of least squares is the only technique which enables us to obtain the rate of growth per annum, for yearly data, if linear trend is fitted.

Drawbacks **1. The method is quite tedious** and time-consuming as compared with other methods. It is rather difficult for a non-mathematical person (layman) to understand and use.

2. The addition of even a single new observation necessitates all calculations to be done afresh.

3. Future predictions or forecasts based on this method are based only on the long term variation, *i.e.*, trend and completely ignore the cyclical, seasonal and irregular fluctuations.

4. The most serious limitation of this method is the determination of the type of the trend curve to be fitted, *viz.*, whether we should fit a linear or a parabolic trend or some other more complicated trend curve.

5. It cannot be used to fit growth curves like Modified Exponential curve, Gompertz curve and Logistic curve, to which most of the economic and business time series data conform.

Example 2·2. *In a certain industry, the production of a certain commodity (in '000 units) during the years 1994—2004 is given in the adjoining table :*

Year	Production ('000 units)	Year	Production ('000 units)
1994	66·6	2000	93·2
1995	84·9	2001	111·6
1996	88·6	2002	88·3
1997	78·0	2003	117·0
1998	96·8	2004	115·2
1999	105·2		

(i) *Graph the data.*

(ii) *Obtain the least square line fitting the data and construct the graph of the trend line.*

(iii) *Compute the trend values for the year 1994-2004 and estimate the production of commodity during the years 2005 and 2006, if the present trend continues.*

(iv) *Eliminate the trend.*

Solution. Here $n = 11$, *i.e.*, odd and, therefore, we shift the origin to the middle time period, *viz.*, the year 1999. Let $x = t - 1999$ ... (1)

TABLE 2·1 : COMPUTATION OF TREND LINE

Year (t)	Production ('000 units) (y_t)	x	$x\,y_t$	x^2	Trend values ('000 units) (t_e) $y_e = 95{\cdot}49 + 3{\cdot}95x$
1994	66·6	– 5	– 333·0	25	75·74
1995	84·9	– 4	– 339·6	16	79·69
1996	88·6	– 3	– 265·8	9	83·64
1997	78·0	– 2	– 156·0	4	87·59
1998	96·8	– 1	– 96·8	1	91·54
1999	110·2	0	0	0	95·49
2000	93·2	1	93·2	1	99·44
2001	111·6	2	223·2	4	103·39
2002	88·3	3	264·9	9	107·34
2003	117·0	4	468·0	16	111·29
2004	115·2	5	576·0	25	115·24
Total	1,050·4	0	434·1	110	

Let the least square line of y_t on x be : $y_t = a + bx$ (origin : July 1999) ... (2)

The normal equations for estimating a and b are

$$\sum y_t = na + b\sum x \quad \text{and} \quad \sum x.y_t = a\sum x + b\sum x^2$$

$$\Rightarrow \quad 1050 = 11a \qquad \Big| \quad \Rightarrow \quad 434{\cdot}1 = 110b$$

$$\Rightarrow \quad a = \frac{1050{\cdot}4}{11} = 95{\cdot}49 \qquad \Big| \quad \Rightarrow \quad b = \frac{434{\cdot}1}{110} = 3{\cdot}95$$

Hence, the least square line fitting the data is : $y_t = 95{\cdot}49 + 3{\cdot}95x$, ... (3)

where origin is July 1999 and x unit = 1 year.

Trend values for the years 1994 to 2004 are obtained on putting $x = -5, -4, -3, \ldots, 4, 5$ respectively in (3) and have been tabulated in the last column of the *Table 2·1*.

Estimate for 2005. Taking $x = 2005$ in (1), we get $x = 2005 - 1999 = 6$

Hence the estimate production of the commodity for 2005 is obtained on putting $x = 6$ in (***) and is given by :

$$(\hat{y}_e)_{2005} = 95.49 + 3.95 \times 6 = 119.19 \text{ ('000 units)}.$$

Similarly, $(\hat{y}_e)_{2006} = 95{\cdot}49 + 3{\cdot}95 \times 7 = 123{\cdot}14$ ('000 units)

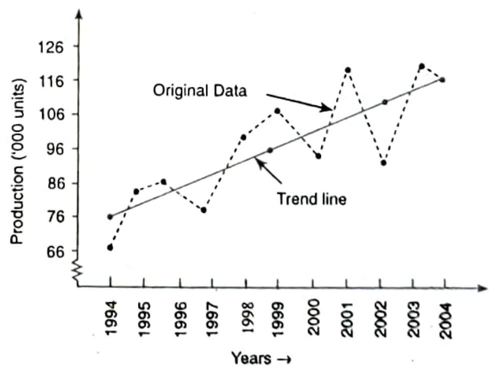

Fig. 2·2.

The graph of the original data and the trend line is given in *Fig. 2·2* :

Assuming multiplicative model, the trend values are eliminated on dividing the given values (y_t) by the corresponding trend values (y_e). However, if we assume the additive model, the trend eliminated values are given by $(y_t - y_e)$. The resulting values contain short-term (seasonal and cyclic) variations and irregular variations. Trend eliminated values are given in *Table 2·2*.

TABLE 2·2 : ELIMINATION OF TREND

Year	Trend Eliminated Values Based on	
	Additive Model $(y_t - y_e)$	*Multiplicative Model* $(y_t \div y_e)$
1994	66·6 – 75·74 = – 9·14	66·6/75·74 = 0·880
1995	84·9 – 79·69 = 5·21	84·9/79·69 = 1·065
1996	88·6 – 83·64 = 4·96	1·059
1997	78·0 – 87·59 = – 9·59	0·891
1998	96·8 – 91·54 = 5·26	1·057
1999	110·2 – 95·49 = 14·71	1·154
2000	93·2 – 99·44 = – 6·24	0·937
2001	111·6 – 103·39 = 8·21	1·079
2002	88·3 – 107·34 = – 19·04	0·823
2003	117·0 – 111·29 = 5·71	1·051
2004	115·2 – 115·24 = – 0·04	0·999

***Example 2·3.** Fit a straight line trend by the method of least squares to the following data relating to the sales of a leading departmental store. Assuming that the same rate of change continues, what would be predicted earnings for the year 2006 ?*

Year :	*1997*	*1998*	*1999*	*2000*	*2001*	*2002*	*2003*	*2004*
Sales (Crores Rs.) :	*76*	*80*	*130*	*144*	*138*	*120*	*174*	*190*

Solution. Here $n = 8$, *i.e.*, even. Hence we shift the origin to the arithmetic mean of the two middle years, *viz.*, 2000 and 2001. We define

$$x = \frac{t - \frac{1}{2}(2000 + 2001)}{\frac{1}{2}\text{(Interval)}} = \frac{t - 2000{\cdot}5}{\frac{1}{2} \times 1} = 2t - 4001 \qquad \ldots (1)$$

where x values are in units of six months (half year).

TABLE 2·3 : COMPUTATION OF LINEAR TREND

Year (t)	Sales (Crores Rs.) y_t	x	$x\,y_t$	x^2	Trend values (Crores Rs.) $y_e = 131{\cdot}5 + 7{\cdot}33x$
1997	76	– 7	– 532	49	80·19
1998	80	– 5	– 400	25	94·85
1998	130	– 3	– 390	9	109·51
2000	144	– 1	– 144	1	124·17
2001	138	1	138	1	138·83
2002	120	3	360	9	153·49
2003	174	5	870	25	168·15
2004	190	7	1330	49	182·81
Total	$\sum y_t = 1052$	$\sum x = 0$	$\sum x\,y_t =$ 1,232	$\sum x^2 = 168$	

Let the linear trend equation between y_t and x be :

$$y_t = a + bx, \quad x = 2(t - 2000{\cdot}5) \qquad \ldots (2)$$

Since $\sum x = 0$, the normal equations for estimating a and b are :

$$a = \frac{\sum y_t}{n} = \frac{1052}{8} = 131{\cdot}5\,, \quad b = \frac{\sum x\,y_t}{\sum x^2} = \frac{1232}{168} = 7{\cdot}33$$

Hence the least square trend line becomes : $y_t = 131{\cdot}5 + 7{\cdot}33x$ ... (3)

where $b = 7{\cdot}33$ units represent half yearly increase in the earnings.

The trend values for the year 1997 to 2004 can now be obtained from (3) on putting it $x = -7, -5, \ldots, 5, 7$ respectively, as shown in the last column of the above *Table 2·3*.

Estimate for 2006 : When $t = 2006$, we get from (1), $x = 2(2006 - 2000{\cdot}5) = 11$

Hence the predicted sales for 2006 are : $y_e = 131{\cdot}5 + 7{\cdot}33 \times 11 = 212{\cdot}13$ (Crores Rs.)

***Example 2·4.** Below are given the figures of production (in thousand tonnes) of a fertiliser factory :*

Year :	*1995*	*1997*	*1998*	*1999*	*2000*	*2001*	*2004*
Production ('000 tonnes) :	*77*	*88*	*94*	*85*	*91*	*98*	*90*

(i) Fit a straight line by the 'Least Squares Method' and tabulate the trend values.

(ii) Eliminate the trend, assuming additive model. What components of the time series are thus left over ?

(iii) What is the monthly increase in the production ?

Solution. (i)

TABLE 2·4 : COMPUTATION OF TREND VALUES

Year (t)	Production (y_t)	$x = t - 1999$	xy_t	x^2	Trend values ('000 tonnes) $y_e = 88{\cdot}8 + 1{\cdot}37x$	Elimination of Trend
1995	77	– 4	– 308	16	83·32	– 6·32
1997	88	– 2	– 176	4	86·06	+ 1·94
1998	94	– 1	– 94	1	87·43	+ 6·57
1999	85	0	0	0	88·80	– 3·80
2000	91	1	91	1	90·17	+ 0·83
2001	98	2	196	4	91·54	+ 6·46
2004	90	5	450	25	96·65	– 5·65
Total	623	1	159	51	622·97	

Let the trend equation be $y_t = a + bx$, [origin : July 1999]

Normal equations for estimating a and b are

$$\left.\begin{aligned} \sum y_t &= na + b\sum x \\ \sum x\,y_t &= a\sum x + b\sum x^2 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} 623 &= 7a + b \\ 159 &= a + 51b \end{aligned}\right.$$

Solving for a and b, we get : $a = 88{\cdot}80$ and $b = 1{\cdot}37$

$\therefore$ **Trend equation is :** $y_t = 88{\cdot}8 + 1{\cdot}37x\,;\ x = t - 1999$ ... (*)

Substituting the values of x, *viz.*, – 4, – 2, etc. successively, we get the required trend values as shown in the last but one column of *Table 2·4*.

(ii) Assuming additive model for the time series, the trend values are eliminated by subtracting them from the given values, as shown in the last column of *Table 2·4*. The resulting values give the short-term fluctuations which change with a period of more than one year.

(*iii*) Yearly increase in the production of fertiliser, as provided by linear trend $y_t = a + bx$ is 'b' = 1·37 thousand tonnes.

$\therefore$ Monthly increase in production $= \frac{1.37}{12} = 0.114$ thousand tonnes.

Example 2·5. Fit a straight line trend to the following data by the method of least squares and obtain two monthly trend values for Nov. 2000 and Sept. 2001.

Year	1996	1997	1998	1999	2000	2001	2002	2003	2004
Average Monthly Profit (crores Rs.)	12·6	14·8	18·6	14·8	16·6	21·2	18·0	17·4	15·8

Solution. Let the straight line trend of y_t on x be given by :

$y_t = a + bx$, where the origin is July 2000 and x unit = 1 year. ... (1)

The normal equations for estimating a and b in (1) are :

$$\sum y_t = na + b\sum x \quad \text{and} \quad \sum x y_t = a\sum x + b\sum x^2 \qquad \dots (2)$$

TABLE 2·5 : FITTING STRAIGHT LINE TREND

Year (t)	$x = t - 2000$	Average monthly profit (in crores Rs.) (y_t)	x^2	$x\,y_t$	Trend values (crores Rs.) $y_e = 16.64 + 0.43x$
1996	– 4	12·6	16	– 50·4	14·92
1997	– 3	14·8	9	– 44·4	15·35
1998	–2	18·6	4	– 37·2	15·78
1999	– 1	14·8	1	– 14·8	16·21
2000	0	16·6	0	0	16·64
2001	1	21·2	1	21·2	17·07
2002	2	18·0	4	36·0	17·50
2003	3	17·4	9	52·2	17·93
2004	4	15·8	16	63·2	18·36
Total	0	149·8	60	25·8	

Substituting the values in (2), we get

$$149.8 = a(9) + b(0) \quad \Rightarrow \quad a = (149.8/9) = 16.64$$

and

$$25.8 = a(0) + b(60) \quad \Rightarrow \quad b = (25.8/60) = 0.43$$

$\therefore$ The trend equation is : [From (1)]

$y_t = 16.64 + 0.43x$; (Origin : July 2000, x unit = 1 year)

Since y represent the *monthly average* for each year and the unit of x is 12 months, the trend of monthly average increases by 0·43 in 12 months, *i.e.*, (0·43/12) per month. So the trend equation for *monthly values* is :

$$y_t = 16.64 + \frac{0.43}{12}x \quad \Rightarrow \quad y_t = 16.64 + 0.036x \qquad \dots (3)$$

(origin : 1st July 2000, x unit = 1 month).

In order to make this equation useful for estimating monthly trend values, the origin is to be shifted to the middle of a month. Since July 2000 is selected as origin, and we have to shift the origin half a month later, x should be replaced by x + (1/2). The transformed monthly trend equation is :

$$y_t = 16.64 + 0.036\left(x + \tfrac{1}{2}\right) \quad \Rightarrow \quad y_t = 16.658 + 0.036\,x$$

[(origin : 15 July 2000 ; unit of x = 1 month ; unit of y_t = Monthly (crores Rs.)]

Now we find the trend values, for November 2000, and September 1999. Since, Nov. 2000 is 4 months, (*i.e.*, 4 units) ahead origin, putting $x = 4$ in the trend equation (3) :

$$(y_e)_{Nov.\ 2000} = 16.658 + 0.036 \times 4 = 16.802) \text{ (crores Rs.)}$$

Similarly, Sept. 1999 is 10 months behind the origin, putting $x = -10$, we have (in 3) :

$$(y_e)_{Sept.\ 1999} = 16.658 + 0.036\,(-10) = 16.298 \text{ (crores Rs.)}$$

Example 2·6. *The following figures are the production data of a certain factory manufacturing air-conditioners :*

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Production ('000 units)	17	20	19	26	24	40	35	55	51	74	79

Fit the second degree parabolic trend curve to the above data and obtain the trend values.

Solution. Let the second degree parabolic tend curve be :

$y_t = a + bx + x^2$, where $x = t - 1995$ ... (*)

TABLE 2·6 : COMPUTATION OF PARABOLIC TREND VALUES

Year (t)	Production ('000 units) (y_t)	$x = t - 1995$	x^2	x^3	x^4	$x\,y$	$x^2 y$	Trend Values $y_e = 34 + 6.28x + 0.6x^2$
1990	17	– 5	25	– 125	625	– 85	425	17·60
1991	20	– 4	16	– 64	256	– 80	320	18·48
1992	19	– 3	9	– 27	81	– 57	171	20·56
1993	26	– 2	4	– 8	16	– 52	104	23·90
1994	24	– 1	1	– 1	1	– 24	24	28·32
1995	40	0	0	0	0	0	0	34·00
1996	35	1	1	1	1	35	35	40·88
1997	55	2	4	8	16	110	220	48·96
1998	51	3	9	27	81	153	459	58·24
1999	74	4	16	64	256	296	1184	68·72
2000	79	5	25	125	625	395	1975	80·40
Total	440	$\Sigma x = 0$	$\Sigma x^2 = 110$	$\Sigma x^3 = 0$	$\Sigma x^4 = 1{,}958$	$\Sigma xy = 691$	$\Sigma x^2 y = 4{,}917$	

The normal equations for estimating a, b and c in (*) are :

$$\left.\begin{aligned} \sum y_t &= na + b\sum x + c\sum x^2 \\ \sum xy_t &= a\sum x + b\sum x^2 + c\sum x^3 \\ \sum x^2 y_t &= a\sum x^2 + b\sum x^3 + c\sum x^4 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} 440 &= 11a + 110\,c & \dots (1) \\ 691 &= 110b & \dots (2) \\ 4{,}917 &= 110\,a + 1{,}958\,c & \dots (3) \end{aligned}\right.$$

From (2), we get $\quad b = (691/110) = 6.28$.

Multiplying (1) by 10 and then subtracting from (3), we get

$$4{,}917 - 440 \times 10 = (110a + 1{,}958c) - (110a + 1{,}100c)$$

$$\Rightarrow \quad 517 = 858c \quad \Rightarrow \quad c = 0{\cdot}60.$$

Substituting in (1), we get $a = \dfrac{440 - 110c}{11} = \dfrac{440 - 110 \times 0{\cdot}60}{11} = \dfrac{374}{11} = 34$

Substituting the values of a, b and c in (**), we get the required trend equation as :

$$y_t = 34 + 6{\cdot}28\,x + 0{\cdot}06\,x^2 \;;\quad x = t - 1995 \qquad \ldots(**)$$

The trend values y_e can be computed on putting $x = -5, -4, -1, 0, 1, \ldots, 4, 5$ in (**) and are given in the last column of the *Table. 2·6*

Example 2·7. *You are given the population figures of India as follows :*

Census year (x) :	1911	1921	1931	1941	1951	1961	1971
Population (in crores) :	25·0	25·1	27·9	31·9	36·1	43·9	54·7

Fit an exponential trend $y = ab^x$ to the above data by the method of least squares and find the trend values. Estimate the population in 1981, 2001 and 2011

Solution. Taking logarithm of both sides of the equation $y = a\,b^x$, we get

$$\log y = \log a + x \log b \quad \Rightarrow \quad v = A + Bx \qquad \ldots(1)$$

where $v = \log y$, $A = \log a$ and $B = \log b$. Now (1) represents a linear trend between v and x.

The arithmetic for fitting the linear trend (1) to the given data can be reduced to a great extent if we shift the origin in x to 1941 and change the scale by defining a new variable u as follows :

$$u = [(x - 1941)/10], \quad \text{so that } \textstyle\sum u = 0$$

Thus the linear trend $v = A + Bu$ between v and u is equivalent to the exponential trend

$$y = ab^u, \quad [(u = (x - 1941)/10] \qquad \ldots[(2)]$$

where $A = \log a$ and $B = \log b$.

By the principle of least squares, the normal equations for estimating A and B in (2) are given by :

$$\textstyle\sum v = nA + B\sum u \quad \text{and} \quad \sum uv = A\sum u + B\sum u^2$$

Since $\sum u = 0$, these equations give

$$A = \frac{\sum v}{n} = \frac{\sum v}{7}, \quad B = \frac{\sum uv}{\sum u^2} \qquad (3)$$

TABLE 2·7 : FITTING OF EXPONENTIAL TREND

Year (x)	Population (in crores) (y_t)	$u = \dfrac{x - 1941}{10}$	$v = \log y$	u^2	uv
1911	25·0	−3	1·3979	9	−4·1937
1921	25·1	−2	1·3997	4	−2·7994
1931	27·9	−1	1·4456	1	−1·4456
1941	31·9	0	1·5038	0	0
1951	36·1	1	1·5575	1	1·5575
1961	43·9	2	1·6425	4	3·2850
1971	54·7	3	1·7380	9	5·2140
Total		0	10·6850	28	1·6178

On using (3), we get

$$A = \frac{10{\cdot}6850}{7} = 1{\cdot}5264$$

$$\Rightarrow \quad a = \text{Antilog } A = 33{\cdot}60$$

$$B = \frac{1{\cdot}6178}{18} = 0{\cdot}0577$$

$$\Rightarrow \quad b = \text{Antilog } B = 1{\cdot}142$$

Substituting the values of a and b in (2), the exponential trend fitted to the given data is :

$$y = 33{\cdot}60\,(1{\cdot}142)^{(x - 1941/10)}$$

To obtain the trend values y for different x, we use the linear trend (**),

$$v = A + Bu \Rightarrow v = 1{\cdot}5264 + 0{\cdot}0577u$$

Substituting the appropriate values of u from −3 to 3 in the above equations, we get the corresponding values of v *and* finally the trend values y are obtained from the fact that

$$v = \log y \quad \Rightarrow \quad y = \text{Antilog}\,(v),$$

as shown in the *Table 2·8* :

Hence, on assuming the exponential trend $y = ab^x$, the estimated population for 1981, 2001 and 2011 is 57·18 crores, 74·57 crores and 85·17 crores respectively.

TABLE 2·8 : COMPUTATION OF EXPONENTIAL TREND

Year	u	$0{\cdot}0577\,u$	$u = 1{\cdot}5264 + 0{\cdot}0577u$	Trend Values y_e = Antilog (v)
1911	−3	−0·1731	1·3533	22·56
1921	−2	−0·1154	1·4160	25·76
1931	−1	−0·0577	1·4687	29·43
1941	0	0	1·5264	33·50
1951	1	0·0577	1·5841	38·38
1961	2	0·1154	1·6418	43·83
1971	3	0·1731	1·6995	50·06
1981	4	0·2308	1·7572	57·18
2001	6	0·3462	1·8726	74·57
2011	7	0·4039	1·9303	85·17

2·4·4. Growth Curves and Their Fitting. The various growth curves, *viz.* the modified exponential, Gompertz and Logistic curves as given in (*vi*) § 2·4·3 cannot be determined by the principle of least squares. Special techniques have been devised for fitting these curves to the given set of data. In the following sections we shall discuss these curves and their fitting in detail.

Modified Exponential Curve and its Fitting. As already pointed out modified exponential curve is given by

$$y_t = a + bc^t, \; a > 0 \; ; \qquad \ldots(2{\cdot}9)$$

where y_t represents the time series value at the time t and a, b, c are constants, called its parameters.

Taking first difference of (2·9), we get

$$\Delta y_t = y_{t+h} - y_t = bc^t\,(c^h - 1)$$

where 'h' is the interval of differencing.

Similarly

$$\Delta y_{t-h} = y_t - y_{t-h} = bc^{t-h}\,(c^h - 1)$$

$$\therefore \quad \frac{\Delta y_t}{\Delta y_{t-h}} = c^h, \text{ a constant.}$$

Thus, the most striking feature of the modified exponential curve is that the first differences of the consecutive value of y_t corresponding to equivalent values of t change by a constant ratio. This implies that the first differences of y_t when plotted on a semi-logarthmic graph paper, lie on a straight line. It may be pointed out that in (2·9), the constant 'a' is always positive and $y_t = a$ is the only asymptote of the curve.

We discuss below two methods of fitting modified exponential curve.

1. *Method of Three Selected Points.* We take three ordinates y_1, y_2, y_3, (say), corresponding to three equidistant values of t, (say) t_1, t_2 and t_3 respectively such that

$$t_2 - t_1 = t_3 - t_2$$

Substituting the values of $t = t_1$, t_2 and t_3 in (2·9), we get respectively

$$y_1 = a + bc^{t_1}, \quad y_2 = a + bc^{t_2}, \quad y_3 = a + bc^{t_3} \qquad \ldots(2{\cdot}10)$$

$$\Rightarrow \quad y_2 - y_1 = b(c^{t_2} - c^{t_1}) = bc^{t_1}(c^{t_2 - t_1} - 1) \qquad \ldots(2{\cdot}10a)$$

Now

$$(-3\delta^4 - \tfrac{1}{2}\delta^6)\,y_0 = \left[-3(E^{1/2} - E^{-1/2})^4 - \tfrac{1}{2}(E^{1/2} - E^{-1/2})^6\right] y_0$$

$$= \left[-\frac{1}{2}E^{-3} + 0.E^{-2} + \frac{9}{2}E^{-1} - 8 + \frac{9}{2}E + 0.E^2 - \frac{1}{2}E^3\right] y_0$$

$$= -\frac{1}{2}y_{-3} + 0y_{-2} + \frac{9}{2}y_{-1} - 8y_0 + \frac{9}{2}y_1 + 0.y_2 - \frac{1}{2}y_3$$

$$= \left[-\frac{1}{2},\ 0,\ \frac{9}{2},\ -8,\ \frac{9}{2},\ 0,\ -\frac{1}{2}\right] y_0$$

$$\therefore\quad [-4, 9, -4]\,y_0 + [-3\delta^4 - \tfrac{1}{2}\delta^6]\,y_0 = \left[-\frac{1}{2}, 0, \frac{9}{2} - 4, -8 + 9, \frac{9}{2} - 4, 0, -\frac{1}{2}\right] y_0$$

$$= \tfrac{1}{2}[-1, 0, 1, 2, 1, 0, -1]\,y_0$$

Therefore, the trend value y_0, correct upto 3rd difference becomes

$$y_0 = \frac{1}{350}[5]^2\,[7]\ \ [-1, 0, 1, 2, 1, 0, -1]\,y_0 \qquad \ldots (2\cdot 38a)$$

Proceeding exactly as in the case of Spencer's 15-point formula, the full weights of Spencer's 21-point formula will be found to be :

$$\frac{1}{350}[-1, -3, -5, -5, -2, 6, 18, 33, 47, 57, 60, 57, \ldots]\,y_0 \qquad \ldots (2\cdot 39)$$

It may be noted that weights are symmetric about the mid-value and the sum of weights is unity. Since the formula requires 21 points to find a single trend value, it is named as Spencer's 21-point formula.

***Example* 2·14.** *Calculate the trend of the following time series by applying Spencer's 15-point formula :*

DEATHS FROM CANCER AND OTHER TUMOURS (*In Million*)

Year	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Deaths for persons	*1401*	*1437*	*1460*	*1493*	*1516*	*1586*	*1615*	*1634*	*1654*	*1678*
Year	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
Deaths for persons	*1867*	*1998*	*2078*	*2197*	*2208*	*2263*	*2343*	*2395*	*2920*	*2481*

Solution. In order to obtain the trend values by Spencer's 15-point formula (2·36a), we prepare table whose different columns are explained below :

(1) The first column gives the t, *i.e.*, the year.

(2) The second column gives y_t, *i.e.*, the deaths per year in million.

(3) Third column gives $[4]\,y_t$, *i.e.*, the moving totals of column (2) of extent 4, writing the first moving total between 2nd and 3rd value of t.

(4) Column four gives $[4]^2\,y_t$, *i.e.*, the moving totals of column (3) of extent four, entering the first total against $t = 4$.

(5) Column five gives $[4]^2\,[5]\,y_t$, *i.e.*, the moving totals of column (4) of extent 5, entering the first total against $t = 6$.

(6) Column six gives $[4]^2\,[5]\,[-3, 3, 4, 3, -3]\,y_t$.

(7) Column seven obtained on dividing the values in column six by 320 gives the trend values.

TABLE 2·16 : TREND VALUES BY SENCER'S 15—POINT FORMULA

t (1)	y_t (2)	$[4]\,y_t$ (3)	$[4]^2\,y_t$ (4)	$[4]^2[5]\,y_t$ (5)	$[4]^2[5][-3, 3, 4, 3, -3]\,y_t$ (6)	*Trend Values* (7) = (6) ÷ 320
1	1,401					
2	1,437					
3	1,460	5,791				
4	1,493	5,906	23,962			
5	1,516	6,055	24,522			
6	1,586	6,210	25,105	1,25,474		
7	1,615	6,351	25,631	1,28,612		
8	1,634	6,489	26,254	1,32,322	5,21,278	1,628·99
9	1,654	6,581	27,100	1,37,008	5,38,960	1,684·25
10	1,678	6,833	28,232	1,42,816	5,63,488	1,760·90
11	1,867	7,197	29,791	1,49,550	5,93,889	1,855·90
12	1,998	7,621	31,439	1,56,828	6,26,226	1,956·96
13	2,078	8,140	32,988	1,64,073	6,57,603	2055·01
14	2,197	8,481	34,378	1,71,169		
15	2,208	8,746	35,477	1,78,010		
16	2,263	9,011	36,887			
17	2,343	9,209	38,280			
18	2,395	9,921				
19	2,920	10,139				
20	2,481					

2·5. MEASUREMENT OF SEASONAL VARIATIONS

It has already been pointed out that one of the types of fluctuations found in time series data is the seasonal component. Many economic and business series have distinct seasonal patterns that are pronounced enough to predict future behaviour of the series. The objectives for studying seasonal patterns in a time series are necessitated by the following reasons :

(*i*) To isolate the seasonal variations, *i.e.*, to determine the effect of seasonal swings on the value of the given phenomenon, and

(*ii*) To eliminate them, *i.e.*, to determine the value of the phenomenon if there were no seasonal ups and downs in the series. This is known as *de-seasonalising the given data* and is necessary for the study of cyclic variations.

The determination of seasonal effects is of paramount importance in planning (*i*) business efficiency, or (*ii*) a production programme. For example, the head of a departmental store would be interested to study the variations in the demands of different articles for different months in order to plan his future stocks to cater to the public demands due to seasonal swings. Moreover, the isolation and elimination of seasonal factor from the data is necessary to study the effect of cycles. Obviously, for the study of seasonal variations, the data must be given for 'parts' of year, *viz.*, monthly or quarterly, weekly, daily or hourly. Different methods for measuring seasonal variations are discussed below.

2·5·1. Method of Simple Averages. This is the simplest method of measuring seasonal variations in a time series and involves the following steps :

(*i*) Arrange the data by years and months (or quarters if quarterly data are given).

(*ii*) Compute the average $\bar{x}_i$, (i = 1, 2, ..., 12) for the ith month for all the years. [ith month, i = 1, 2, ..., 12 represents January, February, ..., December respectively.]

(*iii*) Compute the average $\bar{x}$ of the monthly averages, *i.e.*, $\bar{x} = \frac{1}{12} \sum_{i=1}^{12} \bar{x}_i$.

(*iv*) Seasonal indices for different months are obtained by expressing monthly averages as percentage of $\bar{x}$. Thus.

$$\text{Seasonal Index for } i\text{th month} = (\bar{x}_i / \bar{x}) \times 100; \; i = 1, 2, ..., 12.$$

***Remarks* 1.** If instead of monthly averages, we use monthly totals for all the years, the result remains the same.

2. Total of seasonal indices is 12 × 100 = 1,200 for monthly data and 4 × 100 = 400 for quarterly data.

Merits and Demerits. This method is based on the basic assumption that the data do not contain any trend and cyclic components and consists in eliminating irregular components by averaging the monthly (or quarterly) values over different years. Since most of the economic time series have trends, these assumptions are not in general true and as such this method, though simple, is not of much practical utility.

Example 2·13. *Use the method of monthly averages to determine the monthly indices for the following data of production of a commodity for the years 2002, 2003, 2004 :*

Month	(Production in lakhs of tonnes) 2002	2003	2004	Month	(Production in lakhs of tonnes) 2002	2003	2004
January	12	15	16	July	16	17	16
February	11	14	15	August	13	12	13
March	10	13	14	September	11	13	10
April	14	16	16	October	10	12	10
May	15	16	15	November	12	13	11
June	15	15	17	December	15	14	15

TABLE 2·17 : COMPUTATION OF SEASONAL INDICES

Month	(Production in lakhs of tonnes) 2002	2003	2004	Total	Monthly Average	Seasonal Index
(1)	(2)	(3)	(4)	(5)	(6) = (5) ÷ 3	(7) = (6) ÷ $\bar{x}$
Jan.	12	15	16	43	14·33	104·886
Feb.	11	14	15	40	13·33	97·566
March	10	13	14	37	12·33	90·247
April	14	16	16	46	15·33	112·205
May	15	16	15	46	15·33	112·205
June	15	15	17	47	15·66	114·620
July	16	17	16	49	16·33	119·524
Aug.	13	12	13	38	12·66	92·662
Sept.	11	13	10	34	11·33	82·928
Oct.	10	12	10	32	10·66	78·024
Nov.	12	13	11	36	12·00	87·832
Dec.	15	14	15	44	14·66	107·301
Total				492	163·95	1,200
Averages				41	13·6625	100

Average of averages,

$$\bar{\bar{x}} = \frac{163·95}{12} = 13·6625$$

Seasonal Index for Jan.

$$= \frac{14·33}{13·6625} \times 100 = 104·886$$

Seasonal Index for Feb.

$$= \frac{13·33}{13·6625} \times 100 = 97·566$$

and so on.

2·5·2. Ratio to Trend Method. This method is an improvement over the simple averages method and is based on the assumption that seasonal variation for any given month is constant factor of the trend. The measurement of seasonal variation by this method consists of the following steps :

(*i*) Compute the trend values by the principle of least squares by fitting an appropriate mathematical curve (straight line, 2nd degree parabolic curve or exponential curve, etc.).

(*ii*) Express the original data as the percentage of the trend values. Assuming the multiplicative model, these percentages will, therefore contain the seasonal, cyclic and irregular components.

(*iii*) The cyclic and irregular components are then wiped out by averaging the percentages for different months (quarters) if the data are monthly (quarterly), thus leaving us with indices of seasonal variations. Either arithmetic mean or median can be used for averaging, but median is preferred to arithmetic mean since the latter gives undue weightage to extreme values which are not primarily due to seasonal swings. If there are few abnormal values, modified mean (which consists of calculating arithmetic mean after dropping out the extreme or abnormal values) may be used with advantage.

(*iv*) Finally, these indices, obtained in step (*iii*), are adjusted to a total of 1200 for monthly data or 400 for quarterly data by multiplying them throughout by a constant k given by

$$k = \frac{1,200}{\text{Total of the indices}} \quad \text{and} \quad k = \frac{4,00}{\text{Total of the indices}}$$

for monthly and quarterly data respectively.

Merits and Demerits. Since this method attempts at ironing out the cyclical or irregular components by the process of averaging, the purpose will be accomplished only if the cyclical variations are known to be absent or they are not so pronounced even if present. On the other hand, if the series exhibits pronounced cyclical swings, the trend values obtained by the least square method can never follow the actual data as closely as 12-month moving average and as such the seasonal indices obtained by 'ratio to trend' method are liable to be more biased than those obtained by 'ratio to moving average' method discussed in § 2·5·3.

The obvious advantage of this method over the moving average method lies in the fact that 'ratio to trend' can be obtained for each month for which the data are available and as such, unlike the 'ratio to moving average' method, there is no loss of data.

Remark. The calculations are simplified to a great extent by first fitting a trend equation to the yearly totals (or averages) and then obtaining the monthly (or quarterly) trend values by a suitable modification of the trend equation, as illustrated in the following example :

Example 2·16. *Using Ratio to Trend method, determine the quarterly seasonal indices for the adjoining data :*

Year	I Qrt.	II Qrt.	III Qrt.	IV Qrt.
1995	30	40	36	34
1996	34	52	50	44
1997	40	58	54	48
1998	54	76	68	62
1999	80	92	86	82

TABLE 2·18 : COMPUTATION OF LINEAR TREND

Year (t)	Total of quarterly values (Million Rs.)	Average of quarterly values (y)	$x = t - 1997$	x^2	xy	Trend values (Million Rs.)
1995	140	35	−2	4	− 70	32
1996	180	45	−1	1	− 45	44
1997	200	50	0	0	0	56
1998	260	65	1	1	65	68
1999	340	85	2	4	170	80
Total		$\Sigma y = 280$	$\Sigma x = 0$	$\Sigma x^2 = 10$	$\Sigma xy = 120$	

First of all, we will determine the trend values for the quarterly averages by fitting a linear trend by the method of least squares (*c.f.* Remark above, § 2·5·2).

Let the straight line trend equation be :

$$y = a + bx \qquad \ldots(*)$$

(Origin : 1997, x unit : 1 year)

y : Quarterly averages (Million Rs.)]

The normal equations for estimating a and b in (*) are :

$$\left.\begin{aligned}\Sigma y &= na + b\Sigma x\\ \Sigma xy &= a\Sigma x + b\Sigma x^2\end{aligned}\right\} \Rightarrow \left\{\begin{aligned} a &= \frac{\Sigma y}{n} = \frac{280}{5} = 56\\ b &= \frac{\Sigma xy}{\Sigma x^2} = \frac{120}{12} = 10\end{aligned}\right.$$

Hence, the straight line trend is given by the equation :

$$y_e = 56 + 12x \qquad [\text{origin : 1997, } x \text{ unit} = 1 \text{ year}] \qquad \ldots(**)$$

Putting $x = -2, -1, 0, 1, 2$, we obtain the average quarterly trend values for the years 1995 to 1999 respectively, which are given in the last column of the *Table 2·18*.

From the trend equation (*), we observe that :

Yearly increment in trend values $= b = 12 \quad \Rightarrow \quad$ Quarterly increment $= \frac{12}{4} = 3$

The positive value of b implies that we have an increasing trend. Next, we determine the *quarterly trend values* as follows :

For the year 1995, the average quarterly trend value is 32 which is, in fact, the trend value for the middle quarter, *i.e.,* half of the 2nd quarter and half of 3rd quarter, of 1995. Since the quarterly increment is 3, we obtain the trend values for the 2nd and 3rd quarters of year 1995 as 32 − 1·5 and 32 + 1·5, *i.e.,* 30·5 and 33·5 respectively and consequently the trend value for first quarter is 30·5 − 3 = 27·5 and 4th quarter is 33·5 + 3 = 36·5. Similarly we can get the trend values for other years as given in the following table.

TABLE 2·19 : COMPUTATION OF SEASONAL INDICES

Year	Trend Values				Trend Eliminated Values (Given values as% of trend values)			
	I Qrt	II Qrt.	III Qrt.	IV Qrt.	I Qrt	II Qrt.	III Qrt.	IV Qrt.
1995	27·5	30·5	33·5	36·5	109·1	131·1	107·5	93·1
1996	39·5	42·5	45·5	48·5	86·1	122·4	109·9	90·7
1997	51·5	54·5	57·5	60·5	77·7	106·4	93·9	70·3
1998	63·5	66·5	69·5	72·5	85·0	114·3	97·8	85·5
1999	75·5	78·5	81·5	84·5	106·0	117·1	105·5	97·0
				Total	436·9	591·3	514·6	445·6
	Average (A.M.) (Seasonal Indices)				92·78	118·26	102·92	89·12
	Adjusted Seasonal Indices				92·07	117·36	102·14	88·44

The indices obtained above are adjusted to a total of 400 (since the sum of indices = 92·78 + 118·26 + 102·92 + 89·12 = 403·08, which is greater than 400) by multiplying each of them by a constant factor k, called *correction factor* given by : $k = \frac{400}{\text{Sum of indices}} = \frac{400}{403\cdot08} = 0\cdot9924$.

2·5·3. Ratio to Moving Average Method. As pointed out earlier, moving average eliminates periodic movements if the extent (period of moving average) is equal to the period of the oscillatory movements sought to be eliminated. Thus for a monthly data, a 12-month moving average should completely eliminate the seasonal movements if they are of constant pattern and intensity. The method of getting seasonal indices by moving average involves the following steps :

(*i*) Calculate the centred 12-month moving average of the data. These moving average values will give estimates of the combined effects of trend and cyclic variations.

(*ii*) Express the original data (except for 6 months in the beginning and 6 months at the end) as percentages of the centred moving average values. Using multiplicative model, these percentages would then represent the seasonal and irregular components.

(*iii*) The preliminary seasonal indices are now obtained by eliminating the irregular or random component by averaging these percentages. As discussed in § 2·5·2, Step (*iii*), either arithmatic mean or median (preferably median) can be used for averaging.

(*iv*) The sum of these indices = S (say) will not, in general, be 1,200 (or 400) for monthly (or quarterly) data. Finally, an adjustment is done to make the sum of the indices 1,200 (or 400) by multiplying throughout by a constant factor = 1,200/S (or 400/S), *i.e.,* by expressing the preliminary seasonal indices as the percentage of their arithmetic mean. The resultant gives the desired indices of seasonal variations.

Merits and Demerits. Of all the methods of measuring seasonal variations, the 'Ratio to the moving average 'method' is the most satisfactory, flexible and widely used method, for estimating the seasonal fluctuations in a time series because it irons out both trend and cyclical components from the indices of seasonal variations.

However, an obvious drawback of this method is that there is loss of some trend values in the beginning and at the end and accordingly seasonal indices for first six months (or 2 quarters) of the first year and last six months (or 2 quarters) of the last year cannot be determined.