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

$$y_t = T_t + S_t + C_t + R_t \qquad \ldots (2·1)$$

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,

$$y_t = T_t \times S_t \times C_t \times R_t \qquad \ldots (2·2)$$

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

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

2·4·3. 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 :* $y_t = a + bt$

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

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

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

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

(*b*) $y_t = ab^{c^t}$ (Gompertz curve)

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

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

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

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 :

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

From (2·4), we get

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

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

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

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 :

$$y_t = a + bt + ct^2 \qquad \text{...(2·5)}$$

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

$$\left.\begin{aligned} \sum y_t &= na + b\sum t + c\sum t^2 \\ \sum t\, y_t &= a\sum t + b\sum t^2 + c\sum t^3 \\ \sum t^2 y_t &= a\sum t^2 + b\sum t^2 + c\sum t^4 \end{aligned}\right\}, \qquad \text{...(2·6)}$$

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

Fitting of Exponential Curve :

$$y_t = a\, b^t \qquad \text{...(2·7)}$$

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

$$\Rightarrow \quad Y = A + Bt \text{ (say)}, \qquad \text{...(2·7}a\text{)}$$

where $Y = \log y_t$, $A = \log a$, $B = \log b$. ...(2·7b)

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

$$\left.\begin{aligned} \sum Y &= nA + B\sum t, \\ \sum tY &= A\sum t + B\sum t^2 \end{aligned}\right\} \qquad \text{...(2·7}c\text{)}$$

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

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

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

$$Y_t = a\, b^t c^{t^2} \qquad \text{... (2·8)}$$

Taking logarithms of both sides, we get

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

$$\Rightarrow \quad Y_t = A + B\,t + C\,t^2 \qquad \text{...(2·8}a\text{)}$$

where $Y_t = \log y_t$; $A = \log a$; $B = \log b$ and $C = \log c$ ...(2·8b)

Now, (2·8a) 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

$$x = t - 1999 \qquad \dots (1)$$

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

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

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

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

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)

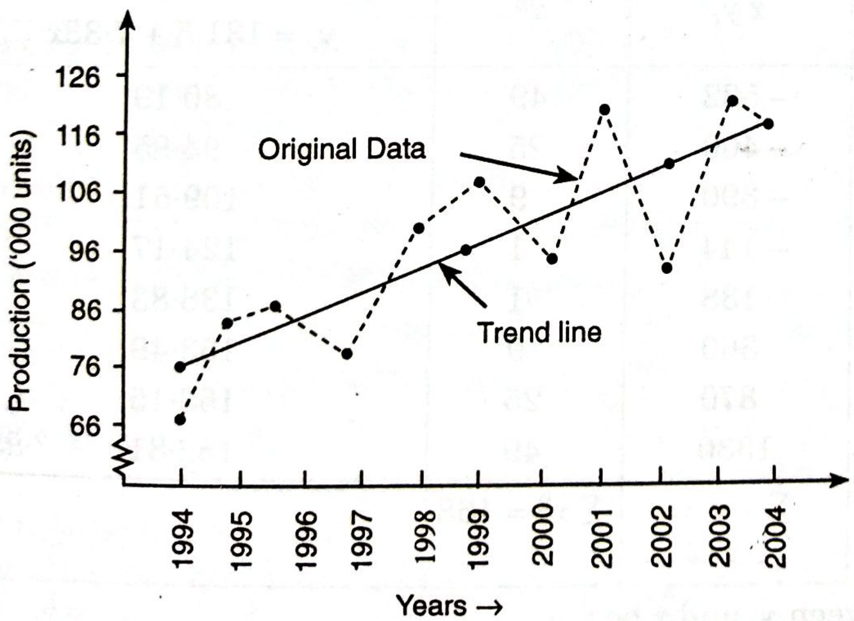

Fig. 2·2.

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	$\Sigma y_t = 1052$	$\Sigma x = 0$	$\Sigma x\,y_t =$ 1,232	$\Sigma x^2 = 168$	

Let the linear trend equation between y_t and x be :

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

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

$$a = \frac{\Sigma y_t}{n} = \frac{1052}{8} = 131\cdot5 \ , \quad b = \frac{\Sigma x y_t}{\Sigma x^2} = \frac{1232}{168} = 7\cdot33$$

Hence the least square trend line becomes : $y_t = 131\cdot5 + 7\cdot33x$ ... (3)

where $b = 7\cdot33$ 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 $x = -7, -5, ..., 5, 7$ respectively, as shown in the last column of the above *Table 2·3*.

Estimate for 2006 : When $t = 2006$, we get from (1), $x = 2(2006 - 2000\cdot5) = 11$

Hence the predicted sales for 2006 are : $y_e = 131\cdot5 + 7\cdot33 \times 11 = 212\cdot13$ (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*) $y_e = 88\cdot8 + 1\cdot37x$	*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

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

Solving for a and b, we get : $a = 88\cdot80$ and $b = 1.37$

$\therefore$ Trend equation is : $y_t = 88\cdot8 + 1\cdot37x \ ; x = t - 1999$ ... (*)

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.

$$\therefore \quad \text{Monthly increase in production} = \frac{1{\cdot}37}{12} = 0{\cdot}114 \text{ thousand tonnes.}$$

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	xy	x^2y	*Trend Values* $y_e = 34 + 6.28x + 0{\cdot}6x^2$
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	$\Sigma x = 0$	$\Sigma x^2 = 110$	$\Sigma x^3 = 0$	$\Sigma x^4 = 1{,}958$	$\Sigma xy = 691$	$\Sigma x^2y = 4{,}917$	

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

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

From (2), we get $b = (691/110) = 6{\cdot}28.$

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 $\quad a = \dfrac{440 - 110c}{11} = \dfrac{440 - 110 \times 0{\cdot}60}{11} = \dfrac{374}{11} = 34$

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 :

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

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

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

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

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

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

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