Business Mathematics and Statistics

NCWEB Hansraj Centre

E-Class on 11/04/2020

By Ms.Komal Chhikara

Bcom Sem II, Section A

Learning Goals:

In this class we will discuss:

- Fitting of linear trend line using principle of least squares.

Least Square Method

This is also known as straight line method. This method is most commonly used in research to estimate the trend of time series data, as it is mathematically designed to satisfy two conditions. They are:

1) Sum of (Y + Yc) = 0 and

2) Sum of $(Y + Yc)^2$ = Least

The straight line method gives a line of best fit on the given data. The straight line which can satisfy the above conditions and make use of regression equation is given by:

Yc = a+ bX

where X represents time variable, Y. is the dependent variable for which trend values are to be calculated, a and b are the constants of the straight line to be found by the method of least squares. Constant **a** is the Y-intercept.

The values of constant, 'a' and 'b' are determined by the following two normal equations.

$$\sum Y = na + b\sum Y - - - - - (1)$$

$$\sum xy = \ a\sum x + b\sum x^2 - - - - (2)$$

The process of finding values of constants a and b can be made simple by using a shortcut method, that is, by taking the origin year in such a way that it gives the toal of 'x' (∑x) equal to 'zero'. This becomes possible if we take the median year as origin period. Thus, the negative values in the first half of the series balance out the positive values in the second half. Thus, the earlier normal equation shall be changed as follows, with reference to $\sum x = 0$.

$$\sum y = a \text{ (as } \sum bx \text{ becomes zero)}$$

$$\sum xy = b\sum x^2 \text{ (as } a\sum x \text{ becomes zero)}$$

Therefore, the values of two constants are obtained by the following formulae:

$$a = \frac{\sum y}{N} \text{ and } b = \frac{\sum xy}{x^2}$$

It is to be noted that when the number of time units involved is even, the point of origin will have to be chosen between the two middle time units.

Merits and demerits of method of Least square

Merits:

(i) This is a mathematical method of measuring trend and is free from subjectivity.

(ii) It provides the line of 'best fit' because it is this line from where the sum of positive and negative deviation is Zero and sum of the square of deviation is least,

i.e., $\sum(Y - Yc) = 0$ and $\sum(Y - Yc)^2$ is least.

(iii) It enables to compute the trend values for the whole time period.

(iv) The trend equation obtained by using least square method can be used to estimate or predict the values of the variable for any given time period't' in future and the predicted values are reliable.

Demerits:

(i) This method is time consuming as the tedious calculations are involved under this method.

(ii) All the calculation needs to be re-done in case of adding/changing of even a single observation.

(iii) Future prediction made by this method is based on long term variations and completely ignore the cyclical, seasonal and irregular fluctuations.

Illustration : Fit a straight line trend by the method of least square from the following data and find the trend values.

Year	1958	1959	1960	1961	1962
Sales (in lakhs of units)	65	95	80	115	105

Solution

We have n = 5, n is odd

Taking middle year i.e. 1960 as the origin, we gets

Year	Sales	X	X^2	XY
1958	65	–2	4	–130
1959	95	–1	1	–95
1960	80	0	0	0
1961	115	1	1	115
1962	105	2	4	210
Total	ΣY=460	ΣX=0	ΣX^2=10	ΣXY=100

N= 5 , $\sum x = 0$, $\sum x^2 = 10$, $\sum y = 460$ and $\sum xy = 100$

a= Σy/n = 460/5 = 92

b= Σxy/ Σx^2 = 100/10 = 10

The equation of the straight line trend is Yc= a+ bx i.e Yc = 92+ 10 x

For the year 1958, x= -2

⇒ Yc (1958) = 92 + 10(−2) = 92 − 20 = 72

For the year 1959, x= -1

⇒ Yc (1959) = 92 + 10(−1) = 92 − 10 = 82

For the year 1960, x= 0

⇒ Yc (1960) = 92 + 10(0) = 92 − 0 = 92

For the year 1961, x= 1

⇒ Yc (1961) = 92 + 10(1) = 92 + 10 = 102

For the year 1962, x= 2

⇒ Yc (1962) = 92 + 10(2) = 92 + 20

Thus we have,

Year	Trend Value	And the straight line trend is Yc = 92 + 10X
1958	72	
1959	82	
1960	92	
1961	102	
1962	112	

Illustration. The sales of a commodity (in ‘000 of Rs.) are given below:

Year	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Sales	82	86	81	86	92	90	99

(i) Using the method of least squares fit a straight line trend equation to the data.

(ii) What is the average annual change in the sales?

(iii) Obtain the trend value for the year 2014-2020 and show that the sum of difference between actual and trend value is equal to zero.

(iv) What are the expected sales for the year 2025.

Solution:

In the given problem n= No. of year covered = 7 (odd). By taking 2017 (i.e. the middle year) as origin and X unit as one year, we get the following:

FITTING STRAIGHT LINE TREND

Year	Sales (in '000 of Rs.)	X= t-2017)	XY	X^2	Trend Value (in '000 of Rs.)	$Y-Y_c$
2014	82	-3	-246	9	80.5	1.5
2015	86	-2	-172	4	83	3
2016	81	-1	-81	1	85.5	-4.5
2017	86	0	0	0	88	-2
2018	92	1	92	1	90.5	1.5
2019	90	2	180	4	93	-3
2020	99	3	297	9	95.5	3.5
	$\Sigma Y = 616$	$\Sigma X = 0$	$\Sigma XY = 70$	$\Sigma X^2 = 28$		$\Sigma(Y-Y_c) = 0$

The equation of the straight line trend is $Yc= a + bX$

Since $\sum X = 0$, the value of a and b can be obtained as follows:

$a= \sum Y/n = 616/7=88$

$b = \sum XY/\sum x^2 =70/28 = 2.5$

(i) Hence the equation of the straight line trend is: $Yc= 88 + 2.5X$ with origin :

2017; X units: one year and Y units : Annual sales (in '000 of Rs.).

ii) The average annual change in the sales = b = 2.5 X 1000 = Rs. 2,500

(iii) The trend value for the year 2014-2020 are obtained by substituting X - -3, -2, -1, 0, 1, 2, 3 in straight line trend equation. These have been shown in the second last column of the above table. The last column shows that the sum of difference between actual and trend value is equal to zero.

(iv) For 2025, X will be 8

$$Y_{2025} = 88 + 2.5\ (8) = 108$$

Thus the expected sales for the year 2025 are Rs. 1, 08,000.

Illustration: The decision making body of a fertilizer firm producing fertilizer wants to predict future sales trend for the year 2006 and 2008 based on the analyses of its past sales pattern. The sales of the firm for the last 7 years, for this purpose are given below:

Year	1998	1999	2000	2001	2002	2003	2004
Sales (in '000 tonnes)	70	75	90	98	85	91	100

Solution: To find the straight line equation (Yc= a + bx) for the given time series data, we have to substitute the values of already arrived expression, that is:

$$\boldsymbol{a} = \frac{\sum \boldsymbol{y}}{\boldsymbol{N}} \textbf{ and } \boldsymbol{b} = \frac{\sum \boldsymbol{xy}}{\boldsymbol{x^2}}$$

In order to make the total of x = 'zero', we must take median year (i.e. 2001) as origin. Study the following table carefully to understand the procedure for fitting the straight line.

Year	Sales (in '000 tonnes)	x	x^2	Xy	Trend (Y_c = a + bx)
1998	70	–3	9	–210	74.5
1999	75	–2	4	–150	78.6
2000	90	–1	1	–90	82.8
2001	98	0	0	0	87.2
2002	85	1	1	85	91.2
2003	91	2	4	182	95.4
2004	100	3	9	300	99.5
N = 7	$\sum y = 609$	$\sum x = 0$	$\sum x^2 = 28$	$\sum xy$ - 117	609.0

$\boldsymbol{a} = \frac{\sum \boldsymbol{y}}{\boldsymbol{N}}$ = 609/7= 87 and $\boldsymbol{b} = \frac{\sum \boldsymbol{xy}}{\boldsymbol{x^2}}$ = 117/28= 4.18

Thus, the straight line trend equation is: Yc = 87 + 4.18x

From the above equation, we can also find the monthly increase in sales as follows:

4.180/12= 348.33 tons

The reason for this is that the trend values increased by a constant amount 'b' every year. Hence the annual increase in sales in 4:18 thousand tons.

Trend values are to be obtained as follow:

Y_{1998} = 87+4.18 (–3) = 74.5

Y_{1999} = 87+4.18 (–2) = 78.6 and so on…

Predicting with decomposed components of the time series: The management wants to estimate fertilizer sales for the years 2006 and 2008.

Estimation of sales for 2006, ‘x’ would be 5 (because for 2004 ‘x’ was 3)

Y_{2006} = 87 + 4.18 (5) = 1.7.9 thousand tonnes

Estimation of sales for 2008, ‘x’ would be 7.

Y_{2008} = 87 + 4.18 (7) = 116.3 thousand tonnes