1. Introduction

Welcome to the series of E-learning modules on fitting of linear curve. In this module, we are going to cover the meaning and the purpose of fitting curve, methods of fitting linear curve and estimating the future value using the fitted curve.

By the end of this session, you will be able to:

- Explain the meaning of fitting curve and their purpose
- Explain the method of fitting linear curve using the method of least squares
- Estimate the future value using the fitted curve

The objective of curve fitting is to find a mathematical equation that describes a set of data that is minimally influenced by randomness.

Curve fitting is the process of finding the curve that best approximates a set of points within a set of curves. The least squares method does this by minimizing the sum of the squares of the differences between the actual and predicted values.

The linear least squares method, which is used here, restricts the set of curves to linear combinations of a set of basic functions. Problems on Linear least squares can be solved using standard methods of linear algebra.

The equation to fit a linear curve or a straight line y is given by:
Y is equal to a plus b into x.

Principle of least squares consists of minimizing the sum of squared deviations between the given values of y and their estimates is given by above equation.

In other words, we have to find a and b such that for given values of y corresponding to n, we get different values of x.
E is equal to summation over y i minus a minus b into x i whole square is minimum.

For a maxima or minima of E and for variations in a and b, we should differentiate E with respect to a and b. Then, equate them to zero and simplify to get a and b.

That is,
D by d a of E is equal to zero is equal to minus 2 summation over y i minus a minus b into x i
Implies summation y i is equal to n into a plus b into summation x i

D by d b of E is equal to zero is equal to minus 2 into summation x i into summation over y i minus a minus b into x i
Implies, summation x i into y i is equal to a into summation x i plus b into summation x i square, which are the normal equation for estimation a and b.

The values of summation y i, summation x i and summation x i square are obtained from the given data and the above equations can be solved for a and b.

With these values of a and b, line y is equal to a plus b into x which gives the desired straight line.
The solution of normal equations provides minima of E.

We can prove that normal equations provide minima of E as follows:
The necessary and sufficient condition for a minima of E for variations in a and b are:
First derivatives of E, d E by D a is equal to zero
D E by d b is equal to zero and
Delta is equal to determinant of
D square E by d a square, d square E by d a d b, d square E by d b d a, d square E by d b square is greater than zero.

Now using the expressions of first derivatives, we can find the second derivatives as follows:
D square e by d a square is equal to 2 into n, which is greater than zero,
D square e by d b square is equal to 2 into summation x square which is also greater than zero,
D square e by d a d b is equal to d square e by d b d a is equal to 2 into summation x.

Therefore, delta is equal to determinant of 2 into n, 2 into summation x, 2 into summation x 2 into summation x square.
Is equal to 4 into n into summation x square minus, summation x the whole square
Is equal to 4 into n square into summation x square by n minus summation x by n whole square, which is equal to 4 into n square into variance of x.

Since variance cannot be negative, the term 4 into square variance of x is greater than zero.
Hence, the solution of the least square equations provides a minimum of E.

2. Illustration 1

Illustration 1
Fit a linear equation for the following data assuming that there is a linear relation between the variable temperature (in degree Celsius) and germination time (in days) at various places. Also estimate the germination time required when temperature is 50 degree Celsius.

Data is given as follows:

Figure 1

Temperature	Germination time	Temperature	Germination time
57	10	42	27
42	26	44	19
40	30	40	18
38	41	46	19
42	29	44	31
45	27	43	29

Let X denote temperature in degree Celsius and Y denote the germination time in days required. To fit a straight line, first we find the following table.

Figure 2

Temperature	Germination time	x^2	xy
57	10	3249	570
42	26	1764	1092
40	30	1600	1200
38	41	1444	1558
42	29	1764	1218
45	27	2025	1215
42	27	1764	1134
44	19	1936	836
40	18	1600	720
46	19	2116	874
44	31	1936	1364
43	29	1849	1247
523	**306**	**23047**	**13028**

First two columns are written as it is, from the given problem. 3rd column is found by squaring the numbers of the first column. That is 57 square is equal to 3 thousand 2 hundred and 49, 42 square is equal to 1 thousand 7 hundred and 64. Similarly, all the other numbers are squared.

The last column is found by multiplying 1st and 2nd column. That is 57 into 10 is equal to 570, 42 into 26 is equal to 1 thousand 92. Similar calculations are done to get the remaining numbers.

Once we find all the elements in different columns, we find the totals of all the columns, which are given in bold numbers.

The linear equation is given by, Y is equal to a plus b into x.
The normal equations are given by,
Summation y is equal to n into a plus b into summation x.
And summation x into y is equal to, a into summation x plus b into summation x square.

By substituting the different values from the above two equations we get,
306 is equal to 12 into a plus b into 523 and
13 thousand 28 is equal to a into 523 plus b into 23 thousand 47.

Now, we will solve the above two simultaneous equations using Cramer's rule to get the values of a and b.

First we consider the determinant where, delta is equal to determinant of 12, 523, 523, 23 thousand 47 is equal to 12 into 23 thousand 47 minus 523 into 523 is equal to 3 thousand 35.

Delta 1 is equal to determinant of 306, 523, 13 thousand 28, 23 thousand 47 is equal to 306 into 23 thousand 47 minus 13 thousand 28 into 523 is equal to 2 lakh 38 thousand 738.

Delta 2 is equal to determinant of 12, 306, 523, 13 thousand 28 is equal to 12 into 13 thousand 28 minus 523 into 306 is equal to minus 3 thousand 7hundred and 2.

Therefore, a is equal to delta one by delta is equal to 2 lakh 38 thousand 7hundred and 38 divided by 3 thousand 35 is equal to 78 point six six.

And B is equal to delta 2 by delta is equal to minus 3 thousand 7hundred and 2 divided by 3 thousand 35 is equal to minus 1 point two two.

Therefore, the equation for the straight line is given by,
Y is equal to 78 point six six minus 1 point two two into x

Hence, the germination time needed when the temperature is 50 degree Celsius is given by,
Y is equal to 78 point six six minus 1 point two two into 50 is equal to 17 point six six nearly equal to 18 days.

3. Illustrations 2 - 3

Illustration 2

Fit a linear curve by the method of least squares for the data regarding sales in lakhs for different years. Estimate the sales for the year 2013.

Figure 3

Year	Sales	Year	Sales
2005	65	2009	73
2006	68	2010	67
2007	70	2011	73
2008	72		

Since X values are very large, we cannot take x as it is. We take time deviation from 2008 as the x value, which is the middle most value. Hence, when we add the values of x, we get zero. These x values were written in the column number 3.

Figure 4

Year	Sales	X=2008-year	x^2	xy
2005	65	-3	9	-195
2006	68	-2	4	-136
2007	70	-1	1	-70
2008	72	0	0	0
2009	73	1	1	73
2010	67	2	4	134
2011	73	3	9	219
Total	**488**	**0**	**28**	**25**

Column 4 is obtained by squaring the values of the column 3, that is minus 3 square is equal to 9, minus 2 square is equal to 4. Similarly, all the other numbers are squared. The last column is obtained by multiplying the numbers in 2nd and 3rd column. That is, 65 into minus 3 is equal to minus 195, 68 into minus 2 is equal to minus 136. Similar calculations are done to get the remaining numbers. Finally, we find the totals of all the columns.

The linear equation is given by, Y is equal to a plus b into x

The normal equations are given by,
Summation y is equal to n into a plus b into summation x.
And summation x into y is equal to a into summation x plus b into summation x square.

By substituting the different values in the above two equations we get,
488 is equal to 8 into a or a is equal to 488 divided by 8 is equal to 61, and
25 is equal to a into zero plus b into 28. That is 25 is equal to b into 28, implies, b is equal to 25 divided by 28 is equal to zero point 8 nine.

Therefore, the linear equation is given by,
Y is equal to 61 plus zero point 8 nine into x.

To estimate the sales for the year 2013, we find the value of x corresponding to 2013. That is x is equal to 2013 minus 2008 is equal to 5.

Therefore, sales for the year 2013 is given by,
Y is equal to 61 plus zero point 8 nine into 5 is equal to 65 point 45 lakhs.

Illustration 3
In a certain industry, the production of certain commodity in thousand units during the years 2000 to 2010 is given in the following table.

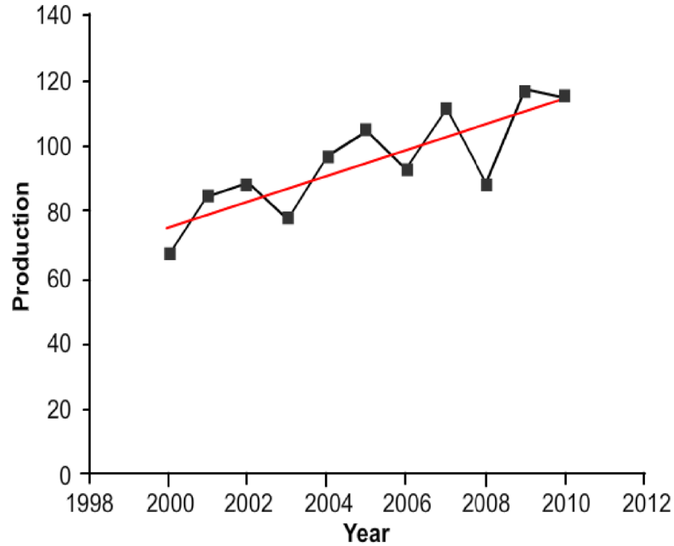
Figure 5

Year	Production	Year	Production
2000	66.6	2006	93.2
2001	84.9	2007	111.6
2002	88.6	2008	88.3
2003	78.0	2009	117.0
2004	96.8	2010	115.2
2005	105.2		

i. Graph the data
ii. Obtain the least square line fitting the data and obtain the estimated line
iii. Estimate the production of commodity during the year 2012 if the present trend continues

We can plot the given data as follows:

Figure 6



Observe that in the graph, the black line with dots shows the actual data and the red line shows the fitted straight line.

To fit a straight line for the given data, first we construct the following table.

Figure 7

Year	Production	x=year-2005	x^2	xy
2000	66.6	-5	25	-333.0
2001	84.9	-4	16	-339.6
2002	88.6	-3	9	-265.8
2003	78.0	-2	4	-156.0
2004	96.8	-1	1	-96.8
2005	105.2	0	0	0
2006	93.2	1	1	93.2
2007	111.6	2	4	223.2
2008	88.3	3	9	264.9
2009	117.0	4	16	468.0
2010	115.2	5	25	576.0
Total	**1050.4**	**0**	**110**	**434.1**

As the variable year (x) is very large, instead of using it directly, we take the deviation from the year 2005, which in turn makes summation x is equal to zero.

Hence, it is easy to solve the normal equation. Therefore, we take the values of x as mentioned in 3rd column. Fourth column is obtained by squaring the values of the 3rd column. That is minus 5 square is equal to 25, minus 4 square is equal to 16. Similarly, all the numbers are squared.

5th column is obtained by multiplying 2nd and 3rd column. That is 66 point 6 into minus 5 is equal to minus 333 point zero, 84 point 9 into minus four is equal to minus 339 point 6. Similar calculations are done to get the remaining numbers.
Once we complete multiplying different numbers, we find the total of all the columns.

The linear equation is given by, Y is equal to a plus b into x
The normal equations are given by,
Summation y is equal to n into a plus b into summation x.
And summation x into y is equal to a into summation x plus b into summation x square.

By substituting the different values in the above two equations we get,
1050.4 is equal to 11 into a implies a is equal to 1050.4 divided by 11 is equal to 95 point 4 nine, and
434 point 1 is equal to a into zero plus b into 110 that is, 434 point 1 is equal to b into 110.
Implies b is equal to 434 point 1 divided by 110 is equal to 3.95.

Therefore, the linear equation to be fitted is given by,
Y is equal to 95 point 4 nine plus 3 point 9 five into x.

To obtain the estimated values for different years, we substitute x is equal to minus 5, minus 4 etc., 4, 5 in the above equation.

That is y is equal to 95 point 4 nine plus 3 point 9 five into minus 5 is equal to 75 point 7 four.
y is equal to 95 point 4 nine plus 3 point 9 five into minus 4 is equal to 79 point 6 nine.
y is equal to 95 point 4 nine plus 3 point 9 five into minus 3 is equal to 83 point 6 four.
Proceeding like this, we get estimated values of the production for all the years.
These values can be tabulated as follows.

Figure 8

Year	Estimated Production	Year	Estimated Production
2000	75.74	2006	99.44
2001	79.69	2007	103.39
2002	83.64	2008	107.34
2003	87.59	2009	111.29
2004	91.54	2010	115.24
2005	95.49		

To estimate the production for the year 2012, the value of x is given by,
X is equal to 2012 minus 2005 is equal to 7
Therefore, y is equal to 95 point 4 nine plus 3 point 9 five into 7 is equal to 123 point 1 four thousand units.

4. Illustration 4

Illustration 4
Fit a straight line by the method of least squares to the following data regarding the production of a firm in tons and estimate the production for the year 2013.

Figure 9

Year	Production	Year	Production
1997	80	2005	94
1999	90	2007	99
2001	92	2009	92
2003	83	2011	110

As the variable year (x) is very large, we take time deviation 2004, which is chosen such that the sum of x becomes zero.

Figure 10

Year	Production (Y)	X= year - 2004	x²	xy
1997	80	-7	49	-560
1999	90	-5	25	-450
2001	92	-3	9	-276
2003	83	-1	1	-83
2005	94	1	1	94
2007	99	3	9	297
2009	92	5	25	460
2011	110	7	49	770
Total	**740**	**0**	**168**	**252**

Hence, the third column gives the deviations of the year from 2004.

Column 4 is found by squaring the values in the column 3. That is minus 7 square is equal to 49, minus five square is equal to 25. Similarly, all the numbers are squared.

The last column is found by multiplying the values in the 2nd and 3rd column.
That is 80 into minus 7 is equal to minus 560. Similar calculations are done to get the remaining numbers.

Once we find all the columns, we find the totals of 2nd, 3rd, 4th and 5th columns.

The linear equation is given by, Y is equal to a plus b into x
The normal equations are given by,
Summation y is equal to n into a plus b into summation x.
And summation x into y is equal to a into summation x plus b into summation x square.

By substituting the different values in the above two equations we get,
740 is equal to 8 into a or a is equal to 740 divided by 8 is equal to 92 point 5, and 252 is equal to a into zero plus b into summation x square
That is 252 is equal to b into 168, implies, b is equal to 1 point 5.

Therefore, the linear equation is given by,
Y is equal to 92 point 5 plus 1 point 5 into x.

To estimate the production for the year 2013, we find the value of x corresponding to 2013 that is x is equal to 2013 minus 2004 is equal to 9.
Therefore, production for the year 2013 is given by,
Y is equal to 92 point 5 plus 1.5 into 9 is equal to 106 tons.

5. Illustration 5

Illustration 5
Fit a straight line by the method of least squares to the following data regarding the sales in thousand rupees during different years. Also estimate the sales for the year 2015.

Figure 11

Year	Sales	Year	Sales
1970	107	1990	115
1975	110	1995	113
1980	114	2000	118
1985	112	2005	115

As the variable year (x) is very large, we take time deviation 1987 point 5 and divide by 2 point 5, which is chosen such that the sum of x becomes zero.

Figure 12

Year	Sales	X= (1987.5-year)/2.5	x^2	xy
1970	107	-7	49	-749
1975	110	-5	25	-550
1980	114	-3	9	-342
1985	112	-1	1	-112
1990	115	1	1	115
1995	113	3	9	339
2000	118	5	25	590
2005	115	7	49	805
Total	**904**	**0**	**168**	**96**

Hence, the third column gives the deviations from the year.

The fourth column is found by squaring the elements of the third column. That is minus 7 square is equal to 49, minus 5 square is equal to 25. Similarly, all the numbers are squared.

The last column is found by multiplying the second and third column. That is 107 into minus 7 is equal to minus 749, minus 5 into 110 is equal to minus 550. Finally, we find the sum of all the columns.

The linear equation for fitting the above data is given by, y is equal to a plus b into x

The normal equations are given by,
Summation y is equal to n into a plus b into summation x and summation x into y is equal to a into summation x plus b into summation x square.

By substituting the different values in the above two equations we get,
904 is equal to 8 into a implies a is equal to 904 divided by 8 is equal to 113
Similarly, for the second equation,
96 is equal to a into zero plus b into 168, that is 96 is equal to b into 168, implies b is equal to 96 by 168 is equal to zero point 5 seven.

Therefore, the linear equation is given by, y is equal to 113 plus zero point 5 seven into x.

To find the sales for the period, we need to find the value of x that is 2015 minus 1987 point five whole divided by 2.5 is equal to, eleven.

Therefore, the estimate of sales for the year 2015 is given by y is equal to 113 plus zero point 5 seven into 11 is equal to 119 point two seven thousand rupees.
Here’s a summary of our learning in this session:

- Meaning and purpose of curve fitting
- Method of fitting linear curve using the method of least squares
- Estimating the future value using the fitted curve