# LAB 1: Linear Curve Fitting

**THEORY:**  Given some data, it is very difficult to infer some logical and meaningful result by looking into it as it is. So after data collection, the next step is to express it graphically. Plot of X vs Y is called as a scatter plot.

If the points lie almost in a straight line then we say that X is related to Y linearly. Then we would like to find the equation of the line y = ax + b that describes the data perfectly. The coefficients of the line are computed using the principle of least squares. So, first form normal equations for each constant and then solve the equations to find the coefficients.

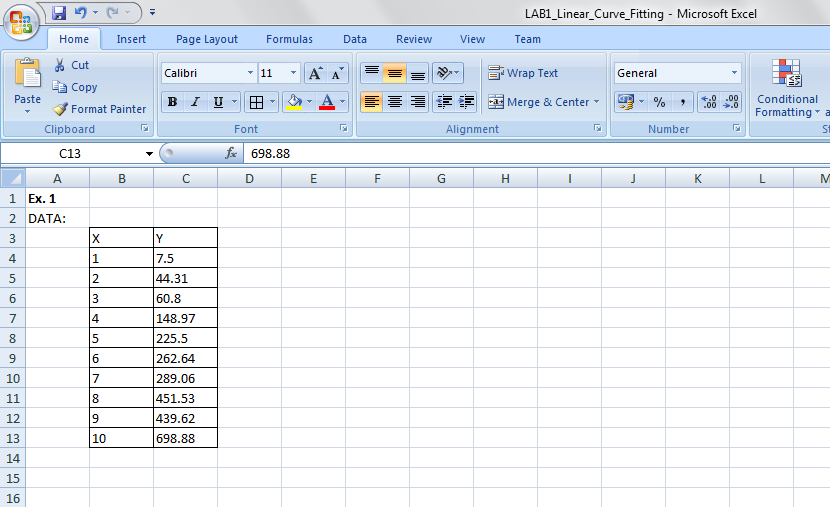
**Ex. 1** Consider the following data

|  |  |
| --- | --- |
| X | Y |
| 1 | 7.5 |
| 2 | 44.31 |
| 3 | 60.8 |
| 4 | 148.97 |
| 5 | 225.5 |
| 6 | 262.64 |
| 7 | 289.06 |
| 8 | 451.53 |
| 9 | 439.62 |
| 10 | 698.88 |

Draw scatter plot and observe whether the X and Y are linearly dependent or not. Find the equation of the straight line that fits best to the given data. Also compute the mean square error. Plot the best fit line.

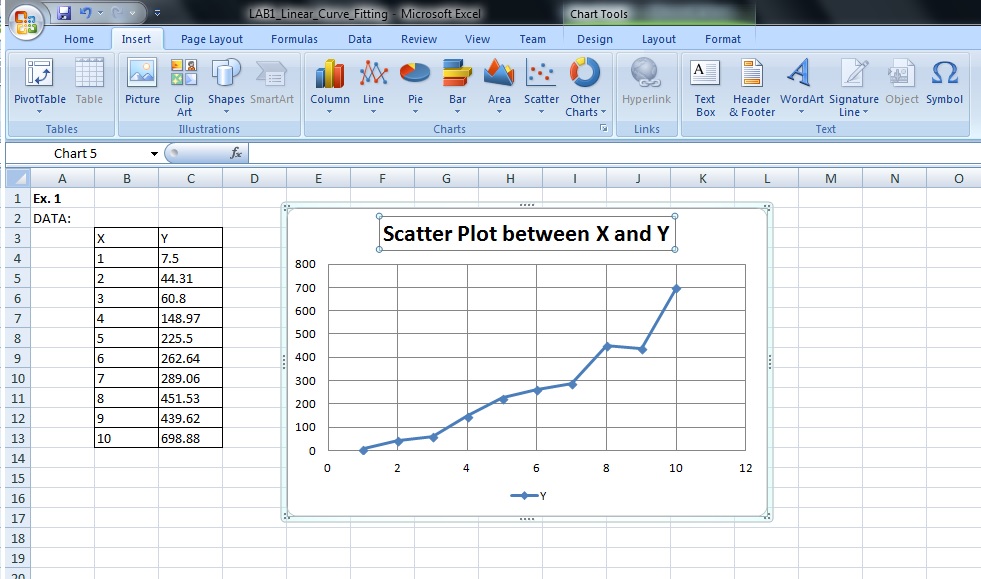
**PROCEDURE 1:**

* Go to files, open a new EXCEL workbook
* Enter the data in the workbook as shown in the 1.1.1

****

**Fig 1.1.1**

* Go to insert tab. Then select the data and choose scatter plot which is shown in Fig.1.1.2. One can edit the title



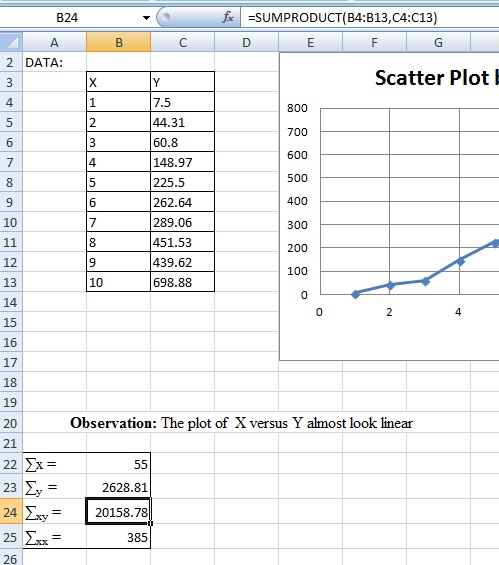
**Fig 1.1.2**

**Observation:** The plot of X versus Y almost looks linear

* Compute ∑x = sum of Xis by typing =SUM(B4:B13) in the cell or by typing =SUM(selecting the cells containing data points of X), press ENTER.

Note that SUM() is the function used to evaluate the sum of data points

* Compute ∑y = sum of Yis by typing =SUM(C4:C13) in the cell, press ENTER.
* Compute ∑xy = sum of product of YiXi s by typing =SUMPRODUCT(B4:B13,C4:C13), press ENTER. Here the SUMPRODUCT() is function used to compute element wise product of arrays or vectors of same dimension and then calculates the sum.
* Compute ∑xx = sum of Xi2s by typing =SUMPRODUCT(B4:B13,B4:B13), press ENTER



**Fig 1.1.3**

* The two normal equations are as follows:

. ∑y = a ∑x + n b

∑xy = a ∑xx + b ∑x

This can be written in the matrix form as

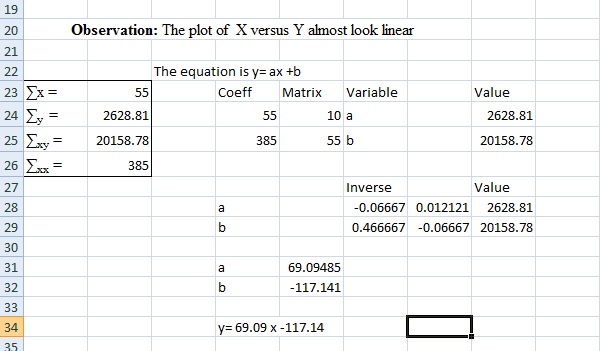
* Type normal equation in matrix form in the excel

|  |  |  |  |
| --- | --- | --- | --- |
| **Coeff** | **Matrix** | **Variable** | **Value** |
| **55** | **10** | **a** | **2628.81** |
| **385** | **55** | **b** | **20158.78** |

* Value of a and b are obtained by multiplying inv(coeff) to value matrix.

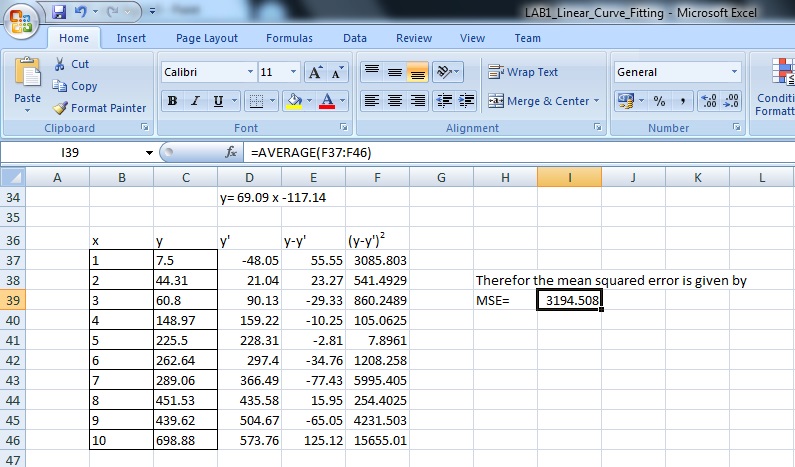
So, the inverse of the matrix is calculated by the given formula

* First select the cells where the inverse has to be entered.
* Then type in formula bar =MINVERSE(select the matrix whose inverse has to be computed) Then press CTRL+ SHIFT+ENTER. Clearly, MINVERSE() is the function used to compute inverse of a matrix.
* Therefore by multiplying inverse matrix by value vector, we get a, b
* First select the cells where the value of a and b has to be evaluated
* Then type in formula bar =MMULT(select the inverse, select the value matrix) Then press CTRL+SHIFT+ENTER. For matrix multiplication MMULT() is used.
* Thus we get ***y=69.09x – 117.14***



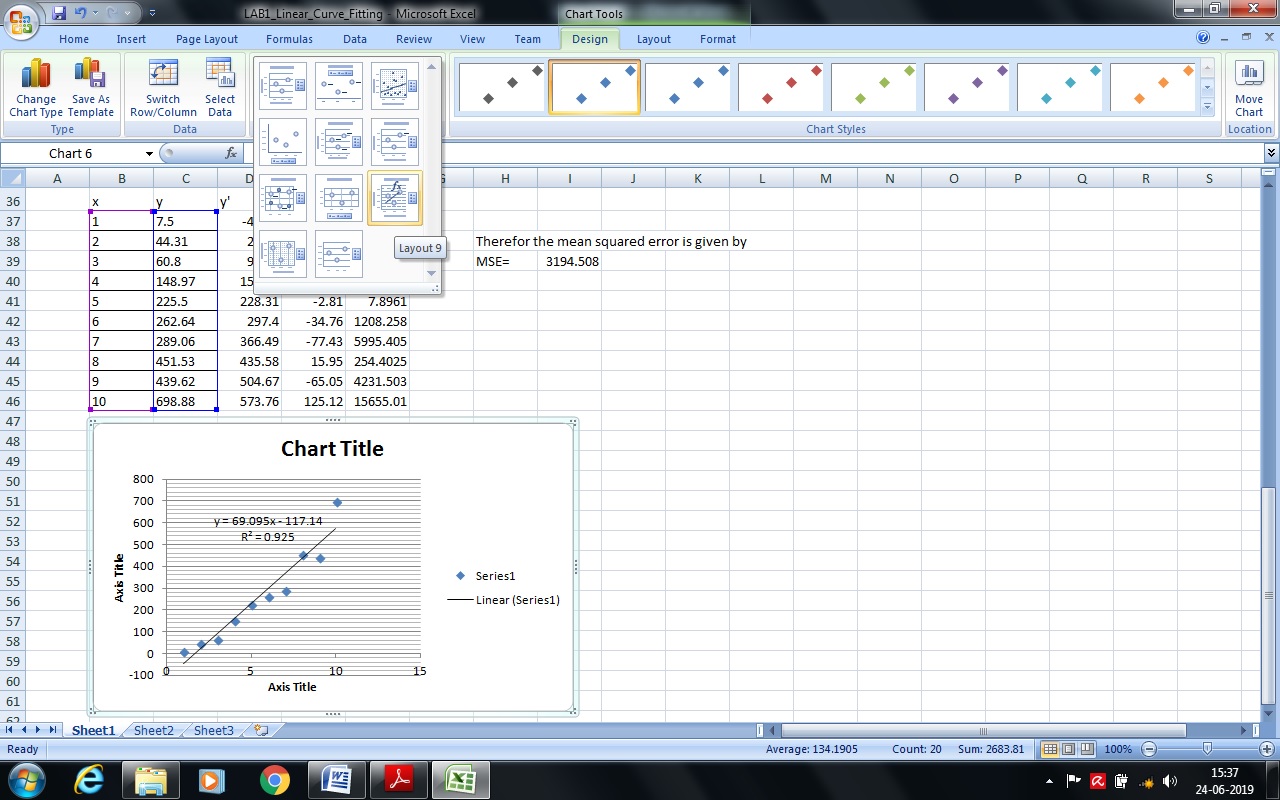
**Fig1.1.4**

* To evaluate the error first we will copy the data as it is once again. Then in next column we will compute y’ using the formula y= 69.09x -117.14. Then in next column we can find y-y’ and then (y-y’)2.
* In the column of y’, select the cells and in formula bar type =69.09\*(select the cells under x )-117.14, press CTRL+SHIFT+ENTER.
* In the column of y-y’, select the destination cells and in formula bar type =(select the cells under y)-(select the cells under y’), press CTRL+SHIFT+ENTER.
* In the column of (y-y’)2, select the destination cells and in formula bar type =(select the cells under (y-y’)2)^2, press CTRL+SHIFT+ENTER.
* Then the mean squared error is given finding the average of the last column.
* MSE = type in the formula bar =AVERAGE(select the cells under (y-y’)2), press ENTER. As it can be observed that function AVERAGE() is used to find average of data points.

****

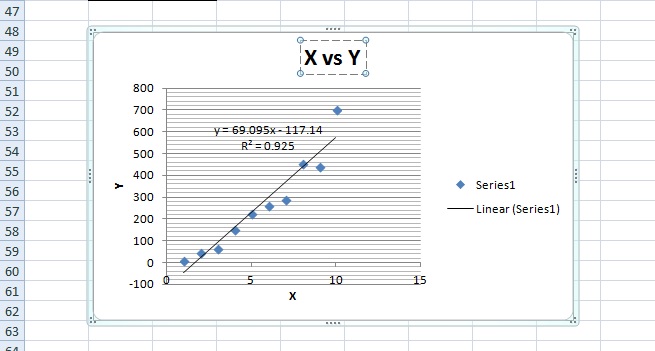
**Fig 1.1.5**

* Select cells under x and y. Then in insert tab choose scatter plot and in scatter plot option, choose layout with best fit line as shown in the Fig 1.1.6. Later we can edit the X axis and Y axis.

****

**Fig 1.1.6**

The final chart will look like as in Fig 1.1.7

****

**Fig 1.1.7**

This completes the problem.

**Ex. 2:** The weights of a calf taken at weekly intervals are given below

|  |  |
| --- | --- |
| Age (in weeks) | Weight (in Kg) |
| 1 | 52.5 |
| 2 | 58.7 |
| 3 | 65.0 |
| 4 | 70.2 |
| 5 | 75.4 |
| 6 | 81.1 |
| 7 | 87.2 |
| 8 | 95.5 |
| 9 | 102.2 |
| 10 | 108.4 |

Plot the data and see whether the data is linear. Find a straight line of best fit and the mean squared error. Estimate the weight of the calf in the 11th week.

**Ex. 3:** The latent heat of vaporisation of steam r, is given in the following table at different temperatures t:

|  |  |
| --- | --- |
| t | R |
| 40 | 1069.1 |
| 50 | 1063.6 |
| 60 | 1058.2 |
| 70 | 1052.7 |
| 80 | - |
| 90 | 1041.8 |
| 100 | 1036.3 |
| 110 | 1030.8 |

Plot the data and see whether the data is linear. Find the equation of the line of best fit and the mean squared error. Also estimate the missing data.