

Least Square Method

A line of best fit is a straight line that is the best approximation of the given set of data. It is used to study the nature of the relation between two variables.

A line of best fit can be roughly determined using an eyeball method by drawing a straight line on a scatter plot so that the number of points above the line and below the line is about equal (and the line passes through as many points as possible).

A more accurate way of finding the line of best fit is the least square method .

Use the following steps to find the equation of line of best fit for a set of ordered pairs $(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)$

Step 1: Calculate the mean of the x-values and the mean of the y-values.

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\bar{Y} = \frac{\sum_{i=1}^n y_i}{n}$$

Step 2: The following formula gives the slope of the line of best fit:

$$m = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2}$$

Step 3: Compute the y-intercept of the line by using the formula:

$$b = \bar{Y} - m\bar{X}$$

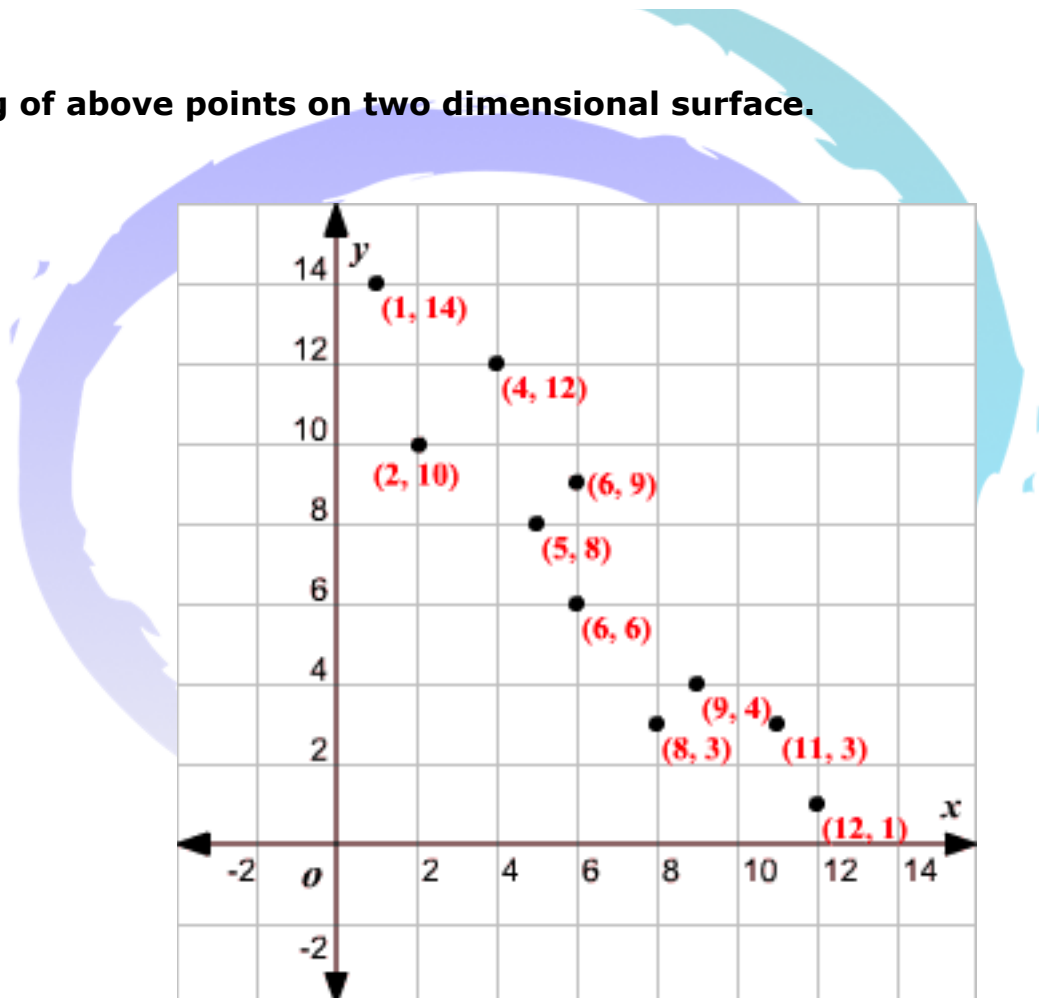
Step 4: Use the slope m and the y-intercept b to form the equation of the line.

Example :

Consider below data set.

x	8	2	11	6	5	4	12	9	6	1
y	3	10	3	6	8	12	1	4	9	14

Plotting of above points on two dimensional surface.



Calculate the means of the x-values and the y-values.

$$\bar{X} = \frac{8+2+11+6+5+4+12+9+6+1}{10} = 6.4$$

$$\bar{Y} = \frac{3+10+3+6+8+12+1+4+9+14}{10} = 7$$

Now calculate $x_i - \bar{X}$, $y_i - \bar{Y}$, $(x_i - \bar{X}) (y_i - \bar{Y})$, and $(x_i - \bar{X})^2$ for each i .

i	x_i	y_i	$x_i - \bar{X}$	$y_i - \bar{Y}$	$(x_i - \bar{X}) (y_i - \bar{Y})$	$(x_i - \bar{X})^2$
1	8	3	1.6	-4	-6.4	2.56
2	2	10	-4.4	3	-13.2	19.36
3	11	3	4.6	-4	-18.4	21.16
4	6	6	-0.4	-1	0.4	0.16
5	5	8	-1.4	1	-1.4	1.96
6	4	12	-2.4	5	-12	5.76
7	12	1	5.6	-6	-33.6	31.36
8	9	4	2.6	-3	-7.8	6.76
9	6	9	-0.4	2	-0.8	0.16
10	1	14	-5.4	7	-37.8	29.16
					$\sum_{i=1}^n (x_i - \bar{X}) (y_i - \bar{Y})$ = -131	$\sum_{i=1}^n (x_i - \bar{X})^2$ = 118.4

Now calculate the slope

$$m = \frac{\sum_{i=1}^n (x_i - \bar{X}) (y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{X})^2} = \frac{-131}{118.4} \approx -1.1$$

Calculate the y -intercept.

Use the formula to compute the y-intercept.

$$\begin{aligned} b &= \bar{Y} - m\bar{X} \\ &= 7 - (-1.1 \times 6.4) \\ &= 7 + 7.04 \\ &\approx 14.0 \end{aligned}$$

Use the slope and y-intercept to form the equation of the line of best fit. The slope of the line is -1.1 and the y-intercept is 14.0 . Therefore, the equation is

$$y = -1.1x + 14.0$$

Draw the line on the scatter plot.

