Computing Regression Line (Line of Best or Least Squares Line) by Hand

The best fit line associated with the **n** points **(x_1, y_1), (x_2, y_2), . . . , (x_n, y_n)** has the form **$y = mx + b$** where

$$\textbf{slope} = \mathbf{m} = \frac{n(S_{xy}) - (S_x)(S_y)}{n(S_{xx}) - (S_x)^2} \quad \text{and} \quad \mathbf{y-intercept} = \mathbf{b} = \frac{(S_y) - m(S_x)}{n}$$

Here we have used the following notation:

$$\mathbf{n} = \textbf{the number of points}$$
$$(S_{xy}) = \textbf{sum of products} = x_1y_1 + x_2y_2 + x_3y_3 + \ldots + x_ny_n$$
$$(S_x) = \textbf{sum of x-values} = x_1 + x_2 + x_3 + \ldots + x_n$$
$$(S_y) = \textbf{sum of y-values} = y_1 + y_2 + y_3 + \ldots + y_n$$
$$(S_{xx}) = \textbf{sum of squares of x-values} = x_1^2 + x_2^2 + x_3^2 + \ldots + x_n^2$$

Example 1. Find the line of best fit (regression line) for the data in the following table.

x	-4	-2	0	2	4
y	1.2	2.8	5.3	6.7	9.1

In order to organize the computations of the sums involved we use the following table construction.

x	y	xy	x^2
-4	1.2		
-2	2.8		
0	5.3		
2	6.7		
4	9.1		
———	———	———	———
S_x=	S_y=	S_{xy}=	S_{xx}=

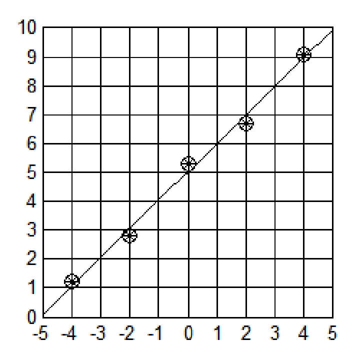
Compute the entries in the ***xy*** and **x^2** columns, then form the sum of the four columns.

Now using the formulas for **m** and **b**:

Compute the slope: m = ______________ Compute the y-intercept: b = ______________

Thus the line of best fit is ________________.

The data and the line of best fit are shown in the accompanying figure. The line of best fit need not go through any of the data points, but comes **closest to the data set in the sense that "the sum of the squares of the vertical deviations is minimized".**



<u>Example 2.</u> Predicting airline passengers.
The table lists the numbers in millions of airline passengers at some of the largest airports in the US during 1999 and 2002.

Airport	1999	2002
Hartsfield Atlanta	78.1	76.9
Chicago O'Hare	72.6	66.5
Los Angeles (LAX)	64.3	56.2
Dallas/Fort Worth	60.0	52.8
Denver	38.0	35.7

Use the 1999 data for x-values and the 2002 data for y-values. Determine the line of best fit.

Construct the table that aids in computing the terms of the formulas for the slope and y-intercept of the line of best fit.

x	y	xy	x^2
78.1	76.9		
72.6	66.5		
64.3	56.2		
60.0	52.8		
38.0	35.7		
—	—	—	—
S_x=	S_y=	S_{xy}=	S_{xx}=

$$\text{slope} = m = \frac{n(S_{xy}) - (S_x)(S_y)}{n(S_{xx}) - (S_x)^2}$$

$$y - \text{intercept} = b = \frac{(S_y) - m(S_x)}{n}$$

Slope = ________________

y-intercept = ____________

Line of best fit has equation ___________________________

Use the equation of the line of best fit to approximate the number of passengers that Miami International Airport had in 2002 given that there were 33.9 million passengers in 1999.

Estimate of number of passengers at Miami International in 2002 is ________________.