Linear Regressions By Hand
Lines of Best Fit

The best fit line associated with the n points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ has the form

$$y = mx + b$$

where

$$\text{slope} = m = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$\text{intercept} = b = \frac{\Sigma y - m(\Sigma x)}{n}$$

Σ means "the sum of." Thus

Σxy = sum of products = $x_1y_1 + x_2y_2 + \ldots + x_ny_n$
Σx = sum of x-values = $x_1 + x_2 + \ldots + x_n$
Σy = sum of y-values = $y_1 + y_2 + \ldots + y_n$
Σx^2 = sum of squares of x-values = $x_1^2 + x_2^2 + \ldots + x_n^2$

Example 1:

Find the least squares line associated with the following data:

x	1	2	3	4
y	1.5	1.6	2.1	3.0

Solution:

In order to apply the formula, it is best to organize the data in a table as shown.

Fill in the values for xy and x^2 in each row.

x	**y**	**xy**	$\mathbf{x^2}$
1	1.5		
2	1.6		
3	2.1		
4	3.0		
$\Sigma x = 10$	$\Sigma y = 8.2$	$\Sigma xy =$	$\Sigma x^2 =$

Substituting the correct values from the above table into the formula gives

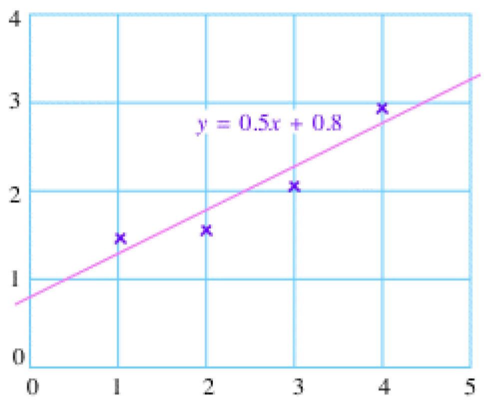
$$\text{slope} = m = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2} = \frac{4(23) - (10)(8.2)}{4(30) - 10^2} = 0.5$$

$$\text{intercept} = b = \frac{\Sigma y - m(\Sigma x)}{n} = \frac{8.2 - (0.5)(10)}{4} = 0.8$$

Thus our least squares line is

$$y = 0.5x + 0.8$$

Here is a plot the data points and the least squares line.



Notice that the line doesn't pass through even one of the original points, and yet it is the straight line that best approximates them.