Example - Linear Regression

In the table below, the x_i column shows scores on the aptitude test. Similarly, the y_i column shows statistics grades. If a student scores 90 marks in aptitude, find his score in statistics using linear regression.

Student	**x_i**	**y_i**
1	95	85
2	85	95
3	80	70
4	70	65
5	60	70

In the table below, the x_i column shows scores on the aptitude test. Similarly, the y_i column shows statistics grades. The last two columns show deviations scores - the difference between the student's score and the average score on each test. The last two rows show sums and mean scores that we will use to conduct the regression analysis

Student	**x_i**	**y_i**	**(x_i-x)**	**(y_i-y)**
1	95	85	17	8
2	85	95	7	18
3	80	70	2	-7
4	70	65	-8	-12
5	60	70	-18	-7
Sum	390	385		
Mean	78	77		

And for each student, we also need to compute the squares of the deviation scores (the last two columns in the table below).

Student	**x_i**	**y_i**	**$(x_i-x)^2$**	**$(y_i-y)^2$**
1	95	85	289	64

2	85	95	49	324
3	80	70	4	49
4	70	65	64	144
5	60	70	324	49
Sum	390	385	730	630
Mean	78	77		

And finally, for each student, we need to compute the product of the deviation scores.

Student	**x_i**	**y_i**	**$(x_i$-x)(y_i-y)**
1	95	85	136
2	85	95	126
3	80	70	-14
4	70	65	96
5	60	70	126
Sum	390	385	470
Mean	78	77	

The regression equation is a linear equation of the form: $\hat{y} = b_0 + b_1 x$. To conduct a regression analysis, we need to solve for b_0 and b_1. Computations are shown below. Notice that all of our inputs for the regression analysis come from the above three tables.

First, we solve for the regression coefficient (b_1):

$$b_1 = \Sigma \left[(x_i - x)(y_i - y) \right] / \Sigma \left[(x_i - x)^2 \right]$$

$$b_1 = 470/730$$

$$b_1 = 0.644$$

Once we know the value of the regression coefficient (b_1), we can solve for the regression slope (b_0):

$$b_0 = y - b_1 * x$$

$$b_0 = 77 - (0.644)(78)$$

$$b_0 = 26.768$$

Therefore, the regression equation is: $\hat{y} = 26.768 + 0.644x$.

$\hat{y} = 84.728$