Question #65454, Math / Statistics and Probability

Compute the appropriate regression equation for the following data:

X	Y
2	18
4	12
5	10
6	8
8	7
11	5

Where X is the independent variable and Y is the dependent variable.
Also find the correlation coefficient between X and Y and infer about the
relationship between X and Y.

Answer.

$$\sum X = 36, \ \sum Y = 60, \ \sum X^2 = 266, \ \sum Y^2 = 706, \ \sum XY = 293.$$

Regression equation $y = a + bx.$

Where $a = \dfrac{\sum Y \sum X^2 - \sum X \sum XY}{n \sum X^2 - (\sum X)^2} = \dfrac{60*266 - 36*293}{6*266 - 36^2} \approx 18.04,$

$b = \dfrac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2} = \dfrac{6*293 - 36*60}{6*266 - 36^2} \approx -1.34.$

So $y = 18.04 - 1.34x.$

Correlation coefficient

$$r = \frac{n \sum XY - \sum X \sum Y}{\sqrt{n \sum X^2 - (\sum X)^2}\sqrt{n \sum Y^2 - (\sum Y)^2}} = \frac{6*293 - 36*60}{\sqrt{6*266 - 36^2}\sqrt{6*706 - 60^2}} \approx -0.9203.$$

There is a strong negative linear correlation between X and Y.