

Regression and Correlation Summary of Steps

1- You are given data for X_i (*independent variable*) and Y_i (*dependent variable*).
Calculate: ΣY_i ΣX_i $\Sigma X_i Y_i$ ΣX_i^2 ΣY_i^2

2- Calculate the *correlation coefficient*, r :

$$r = \frac{n\Sigma X_i Y_i - (\Sigma X_i)(\Sigma Y_i)}{\sqrt{[n\Sigma X_i^2 - (\Sigma X_i)^2][n\Sigma Y_i^2 - (\Sigma Y_i)^2]}} \quad -1 \leq r \leq 1$$

[This can be tested for significance. $H_0: \rho=0$. If the correlation is not significant, then X and Y are not related. You really should not be doing this regression!]

3- Calculate the *coefficient of determination*: $r^2 = \frac{\text{Explained Variation}}{\text{Total Variation}} = (r)^2$

$$0 \leq r^2 \leq 1$$

This is the proportion of the variation in the dependent variable (Y_i) explained by the independent variable (X_i)

4- Calculate the *regression coefficient* b_1 (the slope):

$$b_1 = \frac{n\Sigma X_i Y_i - (\Sigma X_i)(\Sigma Y_i)}{n\Sigma X_i^2 - (\Sigma X_i)^2}$$

Note that you have already calculated the numerator and the denominator for parts of r . Other than a single division operation, no new calculations are required.

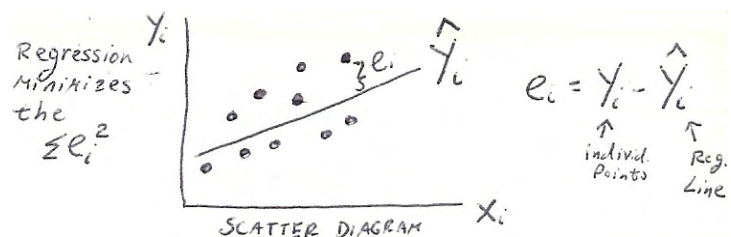
BTW, r and b_1 are related. If a correlation is negative, the slope term must be negative; a positive slope means a positive correlation.

5- Calculate the regression coefficient b_0 (the Y-intercept, or constant):

$$b_0 = \bar{Y} - b_1 \bar{X}$$

6- The regression equation (a straight line) is:

$$\hat{Y}_i = b_0 + b_1 X_i$$



EXAMPLE:

n = 5 pairs of X,Y observations

Independent variable is amount spent on advertising; Dependent variable is sales.

Sales Y_i	Adv X_i	$X_i Y_i$	X_i^2	Y_i^2
2	0	0	0	4
4	1	4	1	16
6	3	18	9	36
7	4	28	16	49
9	5	45	25	81
28	13	95	51	186

Step 1-

$$\begin{aligned}\Sigma Y_i &= 28 \\ \Sigma X_i &= 13 \\ \Sigma X_i Y_i &= 95 \\ \Sigma X_i^2 &= 51 \\ \Sigma Y_i^2 &= 186\end{aligned}$$

Step 2-
$$r = \frac{(5)(95) - (28)(13)}{\sqrt{((5)(51) - (13)^2)((5)(186) - (28)^2)}} = \frac{111}{\sqrt{(86)(146)}} = .9906$$

Step 3-
$$r^2 = (.9906)^2 = 98.13\%$$

Step 4-
$$b_1 = \frac{111}{86} = 1.29$$

Note that b_1 is positive. There is a positive relationship between advertising and sales.

Step 5-
$$b_0 = \left(\frac{28}{5}\right) - 1.29\left(\frac{13}{5}\right) = 2.25$$

Step 6- Thus,
$$\hat{Y}_i = 2.25 + 1.29X_i$$