$$t = \frac{\hat{\beta}_0}{S_{\hat{\beta}_0}} = \frac{9.75}{9.799} = 0.994$$

Decision: Since the calculated t value (0.994) is less than the table value (2.22), that is:

$$|t_{cal.}| = 0.994 < t(0.025,10) = 2.22$$

So, we accept the null hypothesis and reject the alternative hypothesis, that is, $H_1: \beta_0 = 0$. That is, the linear model that does not contain $\hat{\beta}_0$ is a better fit.
Also, if we compare the value of MSe for each of the two models, we will notice that the model that does not contain the Z parameter has a lower MSe than the second model, even if by a small amount.

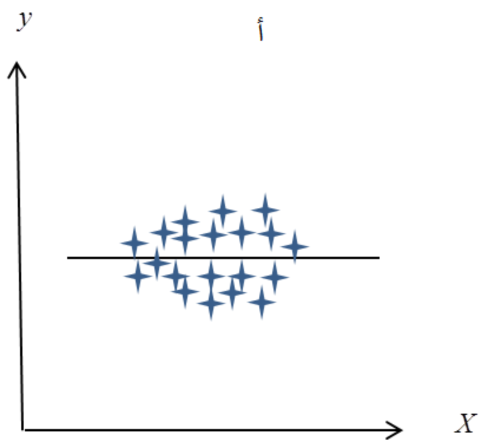
Coefficient of Determination R^2

Definition: The coefficient of determination is defined as what explains the importance of the mathematical model in describing the relationship between X and y to give the percentage of what this relationship explains through the model for the changes occurring in y.

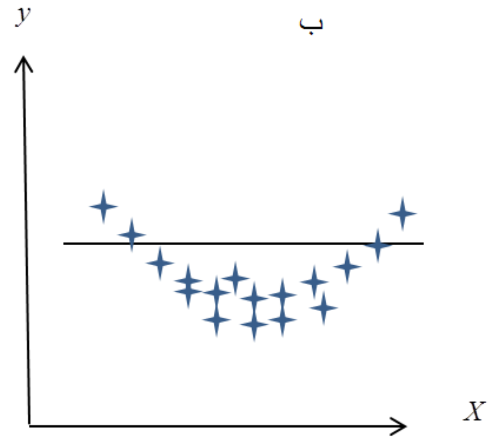
The value of R^2ranges between zero and one, i.e. $0 \leq R^2 \leq 1$.. The formula for the coefficient of determination is as follows:

$$R^2 = \frac{Sum\,of\;Squares\,of\;due\,to\,Regression}{Sum\,of\;Squares\,of\;Total}$$

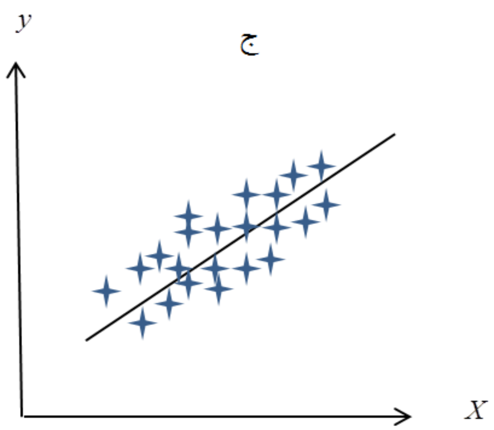
Note that R^2 may not be a good measure of how well the model fits the data. Note the following graphs:



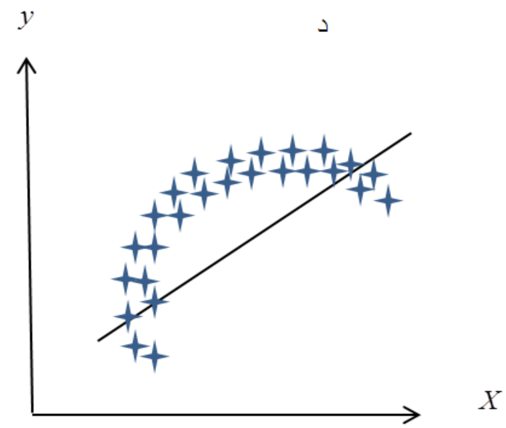
$R^2 = 0$ **while there is no relationship between X and y**



$R^2 = 0$ **while there is a relationship between X and y**



High R^2 value while linear model fits the data



High R^2 value while nonlinear model fits the data

Correlation coefficient and its relationship to simple linear regression

Correlation coefficient

The simple correlation coefficient, symbolized by the symbol r, is a measure of the degree of association and commitment between two independent variables. It can be found using the following formula:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}}$$

The relationship between the regression coefficient and the correlation coefficient is:

$$r_{Xy} = \hat{\beta}_1 \sqrt{\frac{S_{XX}}{S_{yy}}}$$

And also, it can

$$\hat{\beta}_1 = r_{Xy} \sqrt{\frac{S_{yy}}{S_{XX}}}$$

The above two formulas can be proven as follows:

1- Prove that $r_{Xy} = \hat{\beta}_1 \sqrt{\frac{S_{XX}}{S_{yy}}}$

Proof:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}}$$

We multiply and divide the right side by $\sqrt{S_{XX}}$ and we get:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}}\, \frac{\sqrt{S_{XX}}}{\sqrt{S_{XX}}}$$

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}}\, \sqrt{S_{yy}}}\, \frac{\sqrt{S_{XX}}}{\sqrt{S_{XX}}}$$

Multiplying the $\sqrt{S_{XX}}$ in the denominator, we get:

$$r_{Xy} = \frac{S_{Xy}}{S_{XX}}\, \frac{\sqrt{S_{XX}}}{\sqrt{S_{yy}}}$$

Since X we get:

Since $\hat{\beta}_1 = \frac{s_{Xy}}{s_{XX}}$ we get:

$$\therefore r_{Xy} = \hat{\beta}_1 \frac{\sqrt{s_{XX}}}{\sqrt{s_{yy}}}$$

2- Prove that $\hat{\beta}_1 = r_{Xy}\sqrt{\frac{s_{yy}}{s_{XX}}}$

Proof:

as:

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}}$$

We multiply and divide $\sqrt{S_{yy}}$

$$\hat{\beta}_1 = \frac{S_{Xy}}{S_{XX}} \frac{\sqrt{S_{yy}}}{\sqrt{S_{yy}}}$$

We open S_{XX} into two roots, $\sqrt{S_{XX}}$, and thus we get:

$$\hat{\beta}_1 = \frac{S_{Xy}}{\sqrt{S_{XX}}\sqrt{S_{XX}}} \frac{\sqrt{S_{yy}}}{\sqrt{S_{yy}}}$$

By arranging the roots to obtain the formula for r_{Xy}, we get:

$$\hat{\beta}_1 = \frac{S_{Xy}}{\sqrt{S_{XX}}\sqrt{S_{yy}}} \frac{\sqrt{S_{yy}}}{\sqrt{S_{XX}}}$$

Thus we arrive at:

$$\therefore \quad \hat{\beta}_1 = r_{Xy} \frac{\sqrt{S_{yy}}}{\sqrt{S_{XX}}}$$

Finding a table of variance analysis for regression using the correlation coefficient

The table of variance analysis for regression using the correlation coefficient can be organized as follows:

S.O.V	d.f	S.S
R(X1)	1	$r_{Xy}^2 S_{yy}$
Error	n-2	$(1-r_{Xy}^2)S_{yy}$
Total	n-1	

Testing Hypotheses about the Correlation Coefficient

In order to conduct a test about the correlation coefficient, the following hypothesis must be tested:

$H_0 : \rho = r_0$

$H_1 : \rho \neq r_0$

where $r_0$is any value.

This is done using the t-test, which takes the following form:

$$t = \frac{r-r_0}{\sqrt{\frac{1-r^2}{n-2}}}$$

The value of the table t is $t_{tab.} = t(\frac{\alpha}{2}, n-2)$. If it is $|t_{cal.}| > t_{tab.}$, this means rejecting the null hypothesis and accepting the alternative hypothesis, i.e. the correlation coefficient is not equal to r_0.

Example: From the data for the example regarding blood pressure y and age X, find the following:

y	112	128	130	138	158	162	140	175	125	142
X	35	40	38	44	67	64	59	69	25	50

1- Correlation coefficient between y and X.
2- Create an analysis of variance table using the correlation coefficient.
3- Find the coefficient of determination.
4- Test the following hypothesis $\begin{matrix} H_0 : \rho = 0 \\ H_1 : \rho \neq 0 \end{matrix}$:

Solution:

1- The correlation coefficient between X and y.
At first, we find the correlation coefficient as follows:

$$r_{Xy} = \frac{S_{Xy}}{\sqrt{S_{XX}\, S_{yy}}}$$

$$S_{Xy} = \sum X_i y_i - \frac{(\sum X_i)(\sum y_i)}{n}$$

$$S_{XX} = \sum X_i^2 - \frac{(\sum X_i)^2}{n}$$

$$S_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n}$$

We make a table to find the sums of squares, as follows: