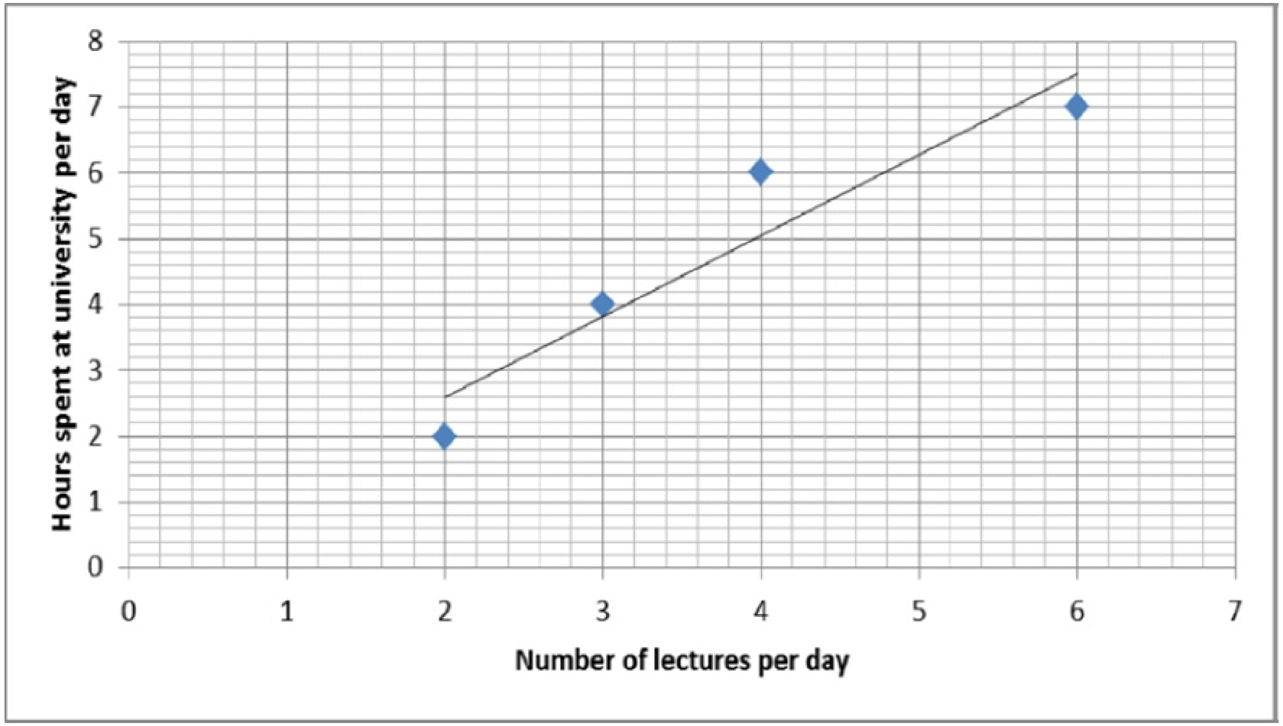
Exercise

Below is a graph showing how the number lectures per day affects the number of hours spent at university per day. The equation of the regression line is drawn on the graph and it has equation $\hat{y} = 0.143 + 1.229x$.



Calculate the R^2

Solution

From the plot, the observed data points are:

$$(x_1, y_1) = (2, 2)$$
$$(x_2, y_2) = (3, 4)$$
$$(x_3, y_3) = (4, 6)$$
$$(x_4, y_4) = (6, 7)$$

The estimated regression line is:

$$\hat{y} = 0.143 + 1.229x$$

The residual associated with the i-th observation is defined as:

$$u_i = y_i - \hat{y}_i,$$

where $\hat{y}_i$ denotes the fitted (theoretical) value obtained from the regression line.
Using the estimated regression line: $\hat{y} = 0.143 + 1.229x$, the fitted (theoretical) values $\hat{y}_i$ are obtained by substituting each observed value x_i into the regression equation.
For $x_1 = 2$, the fitted value is

$$\hat{y}_1 = 0.143 + 1.229 \cdot 2 = 2.601.$$

For $x_2 = 3$, the fitted value is

$$\hat{y}_2 = 0.143 + 1.229 \cdot 3 = 3.830.$$

For $x_3 = 4$, the fitted value is

$$\hat{y}_3 = 0.143 + 1.229 \cdot 4 = 5.059.$$

For $x_4 = 6$, the fitted value is

$$\hat{y}_4 = 0.143 + 1.229 \cdot 6 = 7.517.$$

x_i	y_i	$\hat{y}_i$	u_i	u_i^2
2	2	2.601	-0.601	0.361201
3	4	3.830	0.170	0.028900
4	6	5.059	0.941	0.885481
6	7	7.517	-0.517	0.267289
			RSS	**1.542871**

The **Residual Sum of Squares (RSS)** is therefore:

$$\text{RSS} = \sum_{i=1}^{4} u_i^2 = 1.542871$$

The sample mean of the observed values is:

$$\bar{y} = \frac{2+4+6+7}{4} = 4.75$$

The **Total Sum of Squares (TSS)** measures the total variability of the observed values of the dependent variable around their sample mean and is defined as

$$\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

The deviations from the mean are therefore -2.75, -0.75, 1.25, and 2.25, and squaring and summing these deviations yields

$$\text{TSS} = (-2.75)^2 + (-0.75)^2 + (1.25)^2 + (2.25)^2 = 14.75$$

The coefficient of determination R^2 is given by:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}}$$

Substituting the computed values:

$$R^2 = 1 - \frac{1.542871}{14.75} = 0.895$$

The regression model therefore explains approximately **89.5%** of the total variability of the dependent variable.