Answer on Question #79272 – Math – Statistics and Probability

Question

Given the data:
x 1 2 3 4 5 6 7 8 9
y 9 8 10 12 11 13 14 16 15
(a) Calculate the coefficient of correlation
(b) Obtain the line of regression
(c) Estimate the value of y which should correspond to x = 6.2

Solution

(a)

A correlation coefficient is given by a formula:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2 \sum_{i=1}^{N}(y_i - \bar{y})^2}},$$

where $\bar{x}$ and $\bar{y}$ are the sample means of x and y

Calculate the sample means (N=9):

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = 5$$

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i = 12$$

Thus, the correlation coefficient $r_{xy} = 0.95$.

(b)

Consider the simple linear regression model and find the linear function y=Ax+B that best fits the data. Use the method of least squares to find optimal values of parameters A and B.

It is necessary to minimize the sum:

$$S = \sum_{i=1}^{N}(y_i - (Ax_i + B))^2$$

The minimum of the sum can be found by setting the partial derivatives to zero.

$$\frac{\partial S}{\partial A} = -2\sum_{i=1}^{N} x_i(y_i - Ax_i - B) = -2N(\overline{xy} - A\overline{x^2} - B\bar{x}) = 0$$

$$\frac{\partial S}{\partial B} = -2\sum_{i=1}^{N}(y_i - Ax_i - B) = -2N(\bar{y} - A\bar{x} - B) = 0,$$

where $\overline{xy} = \frac{1}{N}\sum_{i=1}^{N} x_i\, y_i = \frac{199}{3}$, $\overline{x^2} = \frac{1}{N}\sum_{i=1}^{N} x_i^2 = \frac{95}{3}$.

Thus, we receive the system of equations:

$$\begin{cases} -A\overline{x^2} - B\bar{x} + \overline{xy} = 0 \\ -A\bar{x} - B + \bar{y} = 0 \end{cases}$$

Solving this system, we obtain the values of the parameters A and B:

$$A = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - \bar{x}} = 0.95$$

$$B = \frac{\overline{x^2}\,\bar{y} - \bar{x}\,\overline{xy}}{\overline{x^2} - \bar{x}} = 7{,}25$$

The equation of the regression line is $y = 0.95x + 7.25$

(c)

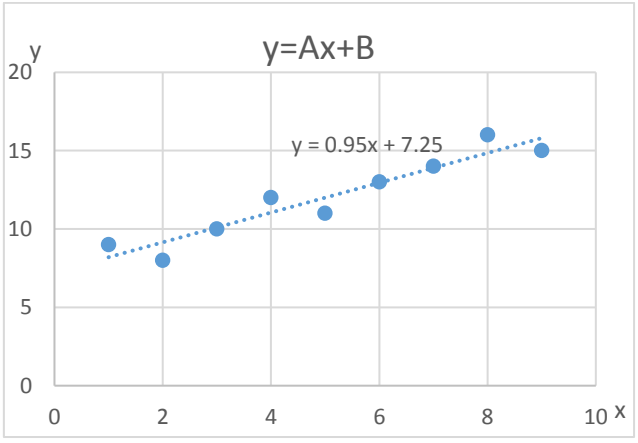
To estimate the value of y, which correspond to x=6.5 we should substitute the value of x into the equation of the line.

$$y = 0.95 \times 6.5 + 7.25 = 13.425$$

Comment

The data used in the calculations are given in table.

x_i	x_i^2	$x_i - \bar{x}$	$(x_i-\bar{x})^2$	y_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	xy
1	1	-4	16	9	-3	9	12	9
2	4	-3	9	8	-4	16	12	16
3	9	-2	4	10	-2	4	4	30
4	16	-1	1	12	0	0	0	48
5	25	0	0	11	-1	1	0	55
6	36	1	1	13	1	1	1	78
7	49	2	4	14	2	4	4	98
8	64	3	9	16	4	16	12	128
9	81	4	16	15	3	9	12	135
$\bar{x}$	$\overline{x^2}$		$\sum(x_i - \bar{x})^2$	$\bar{y}$		$\sum(y_i - \bar{y})^2$	$\sum(x_i - \bar{x})(y_i - \bar{y})$	$\overline{xy}$
5	31.67		60	12		60	57	66.33



Answer:

(a) $r_{xy} = 0.95$

(b) $\mathrm{y} = 0.95\mathrm{x} + 7.25$

(c) $x = 6.5,\ \mathrm{y} = 13.425$