Answer on Question #82543 – Math – Statistics and Probability

Question

Consider a set of 10 data points:
x: 1 2 3 4 4 5 5 6 6 7
y: 7 8 9 8 9 11 10 13 14 13
Derive the simple linear regression model.

Solution

$$mean: \bar{x} = \frac{\sum x_i}{n}, \bar{y} = \frac{\sum y_i}{n}$$

$$n = 10$$

$$\sum_{i=1}^{10} x_i = 43, \ \bar{x} = 4.3$$

$$\sum_{i=1}^{10} y_i = 102, \ \bar{y} = 10.2$$

$$\sum_{i=1}^{10} x_i^2 = 1 + 4 + 9 + 16 + 16 + 25 + 25 + 36 + 36 + 49 = 217$$

$$\sum_{i=1}^{10} y_i^2 = 49 + 64 + 81 + 64 + 81 + 121 + 100 + 169 + 196 + 169 = 1094$$

$$\sum_{i=1}^{10} x_i y_i = 7 + 16 + 27 + 32 + 36 + 55 + 50 + 78 + 84 + 91 = 476$$

$$S_{xx} = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} - \bar{x}^2$$

$$S_{xx} = \frac{217}{10} - (4.3)^2 = 3.21$$

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

$$S_{yy} = \frac{1094}{10} - (10.2)^2 = 5.36$$

$$S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x}\,\bar{y}$$

$$S_{xy} = \frac{476}{10} - 4.3(10.2) = 3.74$$

$$trend\ line: y = A + Bx, B = \frac{S_{xy}}{S_{xx}}, A = \bar{y} - B\bar{x}$$

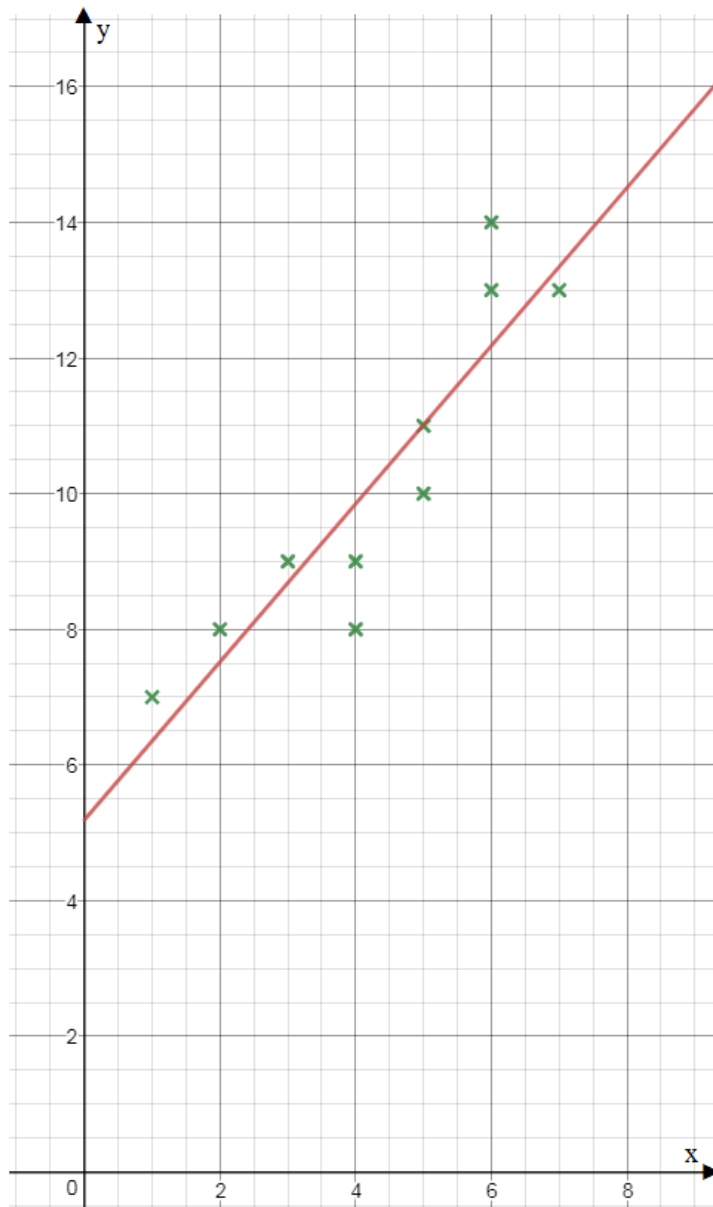$$B = \frac{3.74}{3.21} = 1.165109034$$

$$A = 10.2 - \frac{3.74}{3.21}(4.3) = 5.190031153$$
$$y = 5.190031153 + 1.165109034x$$

correlation coefficient:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

$$r = \frac{3.74}{\sqrt{3.21}\sqrt{5.36}} = 0.9016473096, strong\ correlation$$

A regression line is
$$y = 5.190031153 + 1.165109034x$$

Answer: $y = 5.190031153 + 1.165109034x$