Question

Calculate the regression equation of X on Y
X 1 2 3 4 5
Y 2 5 3 8 7

Solution

Linear regression

$$mean: \bar{x} = \frac{\sum x_i}{n} = \frac{1+2+3+4+5}{5} = 3, \bar{y} = \frac{\sum y_i}{n} = \frac{2+5+3+8+7}{5} = 5$$

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

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

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

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

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

$$S_{xx} = \frac{(1)^2 + (2)^2 + (3)^2 + (4)^2 + (5)^2}{5} - (3)^2 = 2$$

$$S_{yy} = \frac{(2)^2 + (5)^2 + (3)^2 + (8)^2 + (7)^2}{5} - (5)^2 = 5.2$$

$$S_{xy} = \frac{1(2) + 2(5) + 3(3) + 4(8) + 5(7)}{5} - 3(5) = 2.6$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} = \frac{2.6}{\sqrt{2}\sqrt{5.2}} = \frac{\sqrt{2.6}}{2} \approx 0.80622577$$

$0.7 < |r| \leq 1$, strong correlation

$$B = \frac{S_{xy}}{S_{yy}} = \frac{2.6}{5.2} = 0.5$$

$A = \bar{x} - B\bar{y} = 3 - 0.5(5) = 0.5$
Linear regression equation is
$x = 0.5y + 0.5$