Least Squares Regression Line Equation

Example Data:

$$x_i, y_i = (1,1), (2,2), (2,3), (3,6) \qquad \Rightarrow \qquad \bar{x} = 2,\ s_x = 0.816, \text{ and } \bar{y} = 3,\ s_y = 2.160$$

0.1 Review: Find the Correlation Coefficient

Formula:

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right) \tag{1}$$

Evaluate:

$$r = \frac{1}{3} \sum \quad \left(\frac{1-2}{0.816} \right) \left(\frac{1-3}{2.16} \right) \quad + \quad \left(\frac{2-2}{0.816} \right) \left(\frac{2-3}{2.16} \right) \quad + \quad \left(\frac{2-2}{0.816} \right) \left(\frac{3-3}{2.16} \right) \quad + \quad \left(\frac{3-2}{0.816} \right) \left(\frac{6-3}{2.16} \right)$$

Solution:

$$\boxed{r \approx 0.946}$$

0.2 Find Least Squares Regression Equation

Definition: A regression line is simply a single line that best fits the data (in terms of having the smallest overall distance from the line to the points). Statisticians call this technique for finding the best-fitting line a *simple linear regression analysis using the least squares method.*

Formula:

$$\hat{y} = a + bx \tag{2}$$

Evaluate:

$$b = r \left(\frac{S_y}{S_x} \right) \qquad\qquad a = \bar{y} - b\bar{x}$$

$$b = 0.946 \left(\frac{2.16}{0.816} \right) \qquad\qquad a = 3 - 2.5(2)$$

$$b \approx 2.5 \qquad\qquad a = -2$$

Solution:

$$\boxed{\hat{y} = -2 - 2.5x}$$