Calculate the Equation of a Regression Line

0.1 Review: Find the Correlation Coefficient

Multiply the sum, of the products of each $(x_i * y_i)$, by $(\frac{1}{n-1})$ to find (r):

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

Sample Data:

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

Plugin Values and Evaluate:

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

Result:

$$r \approx 0.946 \tag{4}$$

0.2 Find the Regression Line

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.*

The formula for a regression line is denoted with a hat:

$$\hat{y} = mx + b \tag{5}$$

Find the slope using Correlation Coefficient and Deviations:

$$m = r \left(\frac{S_y}{S_x} \right) \tag{6}$$

$$m = 0.946 \left(\frac{2.160}{0.816} \right) \approx \boxed{2.50} \tag{7}$$

Find the y-intercept using m and the x, y values and means:

$$b = \bar{y} - m\bar{x} \tag{8}$$

$$b = 3 - 2.5\,(2) = \boxed{-2} \tag{9}$$

So the explicit formula for the regression line is:

$$\hat{y} = 2.5(x) - 2 \tag{10}$$