

Calculating the Correlation

Now we're ready to compute the correlation value. The formula for the correlation is:

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

Where:

- N = number of pairs of scores
- $\sum xy$ = sum of the products of paired scores
- $\sum x$ = sum of x scores
- $\sum y$ = sum of y scores
- $\sum x^2$ = sum of squared x scores
- $\sum y^2$ = sum of squared y scores

Regression (Best Fit) Line

The best fit line associated with the n points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ has the form

$$y = mx + b$$

where

$$\text{slope} = m = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

$$\text{intercept} = b = \frac{\sum y - m(\sum x)}{n}$$

Here, \sum means "the sum of." Thus

$$\sum xy = \text{sum of products} = x_1y_1 + x_2y_2 + \dots + x_ny_n$$

$$\sum x = \text{sum of x-values} = x_1 + x_2 + \dots + x_n$$

$$\sum y = \text{sum of y-values} = y_1 + y_2 + \dots + y_n$$

$$\sum x^2 = \text{sum of squares of x-values} = x_1^2 + x_2^2 + \dots + x_n^2$$