Correlation Coefficient

Definition

A correlation coefficient is a numerical measure of some type of correlation, meaning a statistical relationship between two variables. The variables may be two columns of a given data set of observations, often called a sample, or two components of a multivariate random variable with a known distribution.

Several types of correlation coefficient exist, each with their own definition and own range of usability and characteristics. They all assume values in the range from 1 to +1, where 1 indicates the strongest possible agreement and 0 the strongest possible disagreement. As tools of analysis, correlation coefficients present certain problems, including the propensity of some types to be distorted by outliers and the possibility of incorrectly being used to infer a causal relationship between the variables.
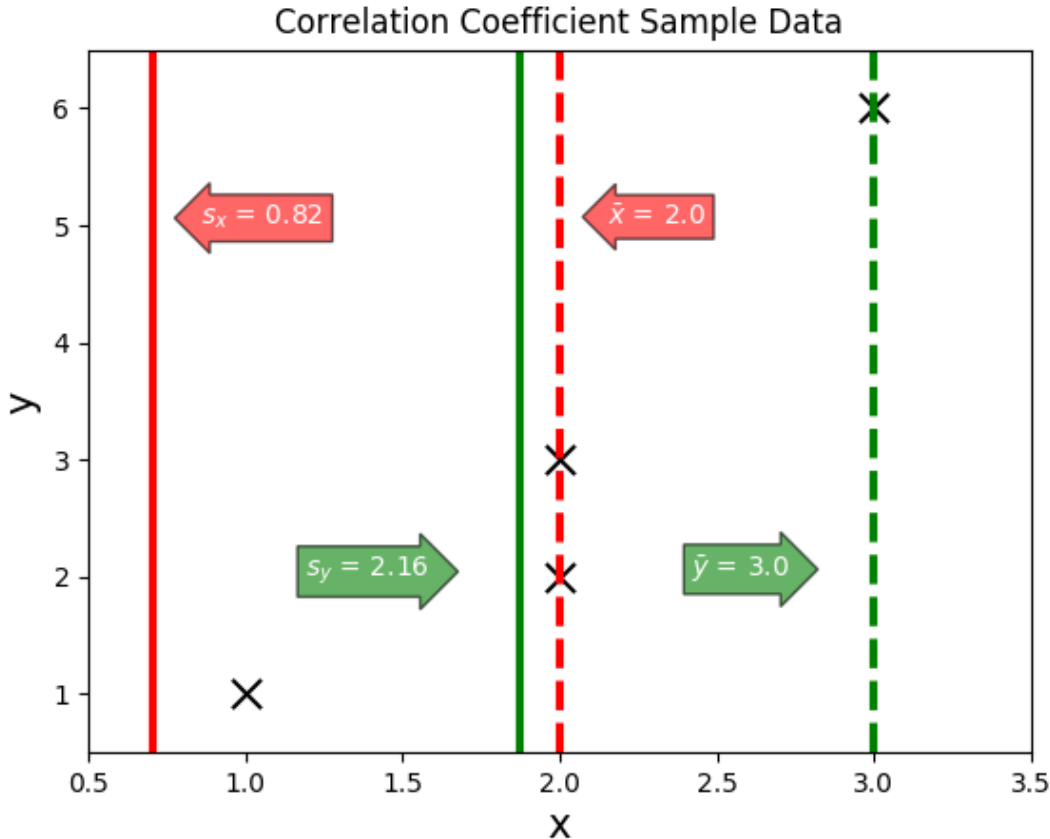
Formula

For each corresponding x and y find the z-score for x:

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

Example

Sample Data $x_i, y_i = (1,1), (2,2), (2,3), (3,6)$, so $\bar{x} = 2$, and $s_x = 0.816$, $\bar{y} = 3$, $s_y = 2.160$

Correlation Coefficient Sample Data

Calculate the sum of the product of the z-scores:

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

Result

$$r \approx 0.946$$

Summary The closer the result is to positive one, the stronger positive correlation. The closer the result is to negative one, the stronger the negative correlation. The closer the result is to zero the weaker correlation. This result has a strong positive correlation. This can be visualized by drawing a line.

Correlation Coefficient Sample Data