Pearson correlation coefficient[1]

In statistics, the Pearson correlation coefficient, is a measure of the linear correlation between two variables X and Y. It has a value between +1 and -1, where 1 is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation.

For a population

Pearson's correlation coefficient when applied to a population is commonly represented by the Greek letter ρ (rho) and may be referred to as the population correlation coefficient or the population Pearson correlation coefficient. The formula for ρ is $\rho_{xy}=\frac{E[(X-\mu_x)(Y-\mu_y)]}{\sigma_x\sigma_y}$ where, μ_x is the mean of X, μ_y is the mean of Y, E is the expectation.

For a sample

Pearson's correlation coefficient when applied to a sample is commonly represented by the letter r and may be referred to as the sample correlation coefficient or the sample Pearson correlation coefficient. The formula for r_{xy} is $r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\cdot\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}$, where n is the sample size, x_i e y_i são the individual sample points indexed with i, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ and analogously for $\bar{y}$.

Rule of Thumb for Interpreting the Size of a Correlation Coefficient[7]

Size of Correlation	Interpretation
.90 to 1.00 (−.90 to −1.00)	Very high positive (negative) correlation
.70 to .90 (−.70 to −.90)	High positive (negative) correlation
.50 to .70 (−.50 to −.70)	Moderate positive (negative) correlation
.30 to .50 (−.30 to −.50)	Low positive (negative) correlation
.00 to .30 (.00 to −.30)	negligible correlation

Expected value[2]

In probability theory, the expected value of a random variable, intuitively, is the long-run average value of repetitions of the experiment it represents.

Total sum of squares[3]

In statistical data analysis the total sum of squares (TSS or SST) is a quantity that appears as part of a standard way of presenting results of such analyses. It is defined as being the sum, over all observations, of the squared differences of each observation from the overall mean.[4]

In statistical linear models, (particularly in standard regression models), the TSS is the sum of the squares of the difference of the dependent variable and its mean $TSS=\sum_{i=1}^{n}(y_i-\bar{y})^2$ where $\bar{y}$ is the mean.

Deviation[6]

In mathematics and statistics, deviation is a measure of difference between the observed value of a variable and some other value, often that variable's mean. The formula for deviation is $Deviation_i=(x_i-\bar{x})$.

Cross-product Deviations [6]

a) the product of the deviation of two variables from their means.
b) the contribution of each observation to the direction and strength of the association.

$$CP=(x-\bar{x})(y-\bar{y})$$

Covariation (Sum of Cross-Product Deviations) [6]

a) the sum of the product of the joint deviations of the individual observations of *X* and *Y* from their respective means.
b)the aggregate association of the association between x and y.
c) if there is no association between x and y the Covariance will be zero, if the association is positive the Covariation will be positive and if the association is negative the Covariation will be negative, although the strength of the association cannot be evaluated by the Covariation.

$$SCP(xy)=\sum(x-\bar{x})(y-\bar{y})$$

For a particular observation if both X and Y are greater than their respective means then the SCP is greater than zero.

$$if\ x>\bar{x}\wedge y>\bar{y}\ then\sum(x-\bar{x})(y-\bar{y})>0$$

The same holds true if both X and Y are less than their respective means.

$$if\ x<\bar{x}\wedge y<\bar{y}\ then\sum(x-\bar{x})(y-\bar{y})>0$$

But if X is greater than its mean, while Y is less than its mean, or vice versa, the SCP is less than zero.

$$if\ x>\bar{x}\wedge y<\bar{y}\ then\sum(x-\bar{x})(y-\bar{y})<0$$

$$if\ x<\bar{x}\wedge y>\bar{y}\ then\sum(x-\bar{x})(y-\bar{y})<0$$

References

[1]Pearson correlation coefficient, wikipedia, https://en.wikipedia.org/wiki/Pearson_correlation_coefficient

[2]Expected value, wikipedia, https://en.wikipedia.org/wiki/Expected_value

[3]Total sum of square, wikipedia, https://en.wikipedia.org/wiki/Total_sum_of_squares

[4]Everitt, B.S. (2002) The Cambridge Dictionary of Statistics, CUP, ISBN 0-521-81099-X

[5]Deviation, wikipedia, https://en.wikipedia.org/wiki/Deviation_(statistics)

[6]Correlation and Simple Regression, Oregon State University, people.oregonstate.edu/~hammerr/Motulsky/Lecture_7_Correlation.doc

[7] Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 5th ed. Boston: Houghton Mifflin; 2003.