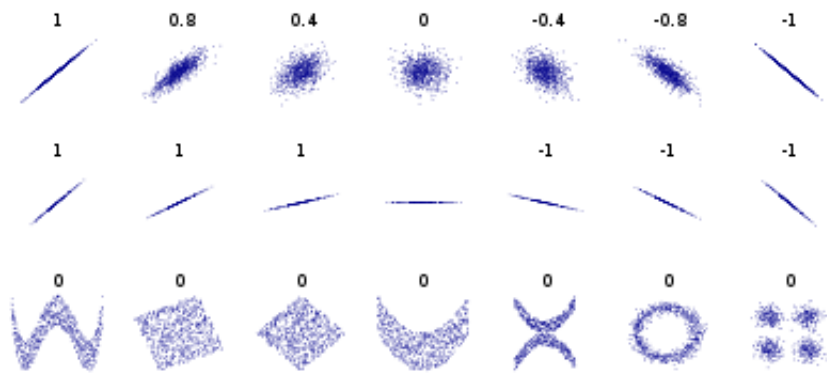


Correlation and dependence

In statistics, **correlation** or **dependence** is any statistical relationship, whether causal or not, between two random variables or bivariate data. In the broadest sense **correlation** is any statistical association, though it commonly refers to the degree to which a pair of variables are linearly related. Familiar examples of dependent phenomena include the correlation between the height of parents and their offspring, and the correlation between the price of a good and the quantity the consumers are willing to purchase, as it is depicted in the so-called demand curve.



Several sets of (x, y) points, with the Pearson correlation coefficient of x and y for each set. The correlation reflects the noisiness and direction of a linear relationship (top row), but not the slope of that relationship (middle), nor many aspects of nonlinear relationships (bottom). N.B.: the figure in the center has a slope of 0 but in that case the correlation coefficient is undefined because the variance of Y is zero.

Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. For example, an electrical utility may produce less power on a mild day based on the correlation between electricity demand and weather. In this example, there is a causal relationship, because extreme weather causes people to use more electricity for heating or cooling. However, in general, the presence of a correlation is not sufficient to infer the presence of a causal relationship (i.e., correlation does not imply causation).

Formally, random variables are *dependent* if they do not satisfy a mathematical property of probabilistic independence. In informal parlance, *correlation* is synonymous with *dependence*. However, when used in a technical sense, correlation refers to any of several specific types of mathematical operations between the tested variables and their respective expected values. Essentially, correlation is the measure of how two or more variables are related to one another. There are several correlation coefficients, often denoted ρ or r , measuring the degree of correlation. The most common of these is the Pearson correlation coefficient, which is sensitive only to a linear relationship between two variables (which may be present even when one variable is a nonlinear function of the other). Other correlation coefficients – such as **Spearman's rank correlation** – have been developed to be more robust than Pearson's, that is, more sensitive to nonlinear relationships.^{[1][2][3]} Mutual information can also be applied to measure dependence between two variables.

Contents

- Pearson's product-moment coefficient
 - Definition
 - Symmetry property

[Correlation and independence](#)
[Sample correlation coefficient](#)

[Example](#)

[Rank correlation coefficients](#)

[Other measures of dependence among random variables](#)

[Sensitivity to the data distribution](#)

[Correlation matrices](#)

[Uncorrelatedness and independence of stochastic processes](#)

[Common misconceptions](#)

[Correlation and causality](#)

[Simple linear correlations](#)

[Bivariate normal distribution](#)

[See also](#)

[References](#)

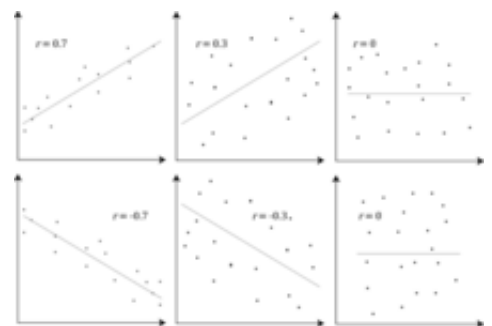
[Further reading](#)

[External links](#)

Pearson's product-moment coefficient

Definition

The most familiar measure of dependence between two quantities is the Pearson product-moment correlation coefficient (PPMCC), or "Pearson's correlation coefficient", commonly called simply "the correlation coefficient". Mathematically, it is defined as the quality of least squares fitting to the original data. It is obtained by taking the ratio of the covariance of the two variables in question of our numerical dataset, normalized to the square root of their variances. Mathematically, one simply divides the covariance of the two variables by the product of their standard deviations. Karl Pearson developed the coefficient from a similar but slightly different idea by Francis Galton.^[4]



Example scatterplots of various datasets with various correlation coefficients.

A Pearson product-moment correlation coefficient attempts to establish a line of best fit through a dataset of two variables by essentially laying out the expected values and the resulting Pearson's correlation coefficient indicates how far away the actual dataset is from the expected values. Depending on the sign of our Pearson's correlation coefficient, we can end up with either a negative or positive correlation if there is any sort of relationship between the variables of our dataset.

The population correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is defined as

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\mathbf{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

where \mathbf{E} is the expected value operator, **cov** means covariance, and **corr** is a widely used alternative notation for the correlation coefficient. The Pearson correlation is defined only if both standard deviations are finite and positive. An alternative formula purely in terms of moments is

$$\rho_{X,Y} = \frac{\mathbf{E}(XY) - \mathbf{E}(X) \mathbf{E}(Y)}{\sqrt{\mathbf{E}(X^2) - \mathbf{E}(X)^2} \cdot \sqrt{\mathbf{E}(Y^2) - \mathbf{E}(Y)^2}}$$

Symmetry property

The correlation coefficient is symmetric: $\text{corr}(X, Y) = \text{corr}(Y, X)$. This is verified by the commutative property of multiplication.

Correlation and independence

It is a corollary of the Cauchy–Schwarz inequality that the absolute value of the Pearson correlation coefficient is not bigger than 1. Therefore, the value of a correlation coefficient ranges between -1 and +1. The correlation coefficient is +1 in the case of a perfect direct (increasing) linear relationship (correlation), -1 in the case of a perfect inverse (decreasing) linear relationship (**anticorrelation**),^[5] and some value in the open interval $(-1, 1)$ in all other cases, indicating the degree of linear dependence between the variables. As it approaches zero there is less of a relationship (closer to uncorrelated). The closer the coefficient is to either -1 or 1, the stronger the correlation between the variables.

If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables.

$$\begin{array}{ll} X, Y \text{ independent} & \Rightarrow \rho_{X,Y} = 0 \quad (X, Y \text{ uncorrelated}) \\ \rho_{X,Y} = 0 \quad (X, Y \text{ uncorrelated}) & \nRightarrow X, Y \text{ independent} \end{array}$$

For example, suppose the random variable X is symmetrically distributed about zero, and $Y = X^2$. Then Y is completely determined by X , so that X and Y are perfectly dependent, but their correlation is zero; they are uncorrelated. However, in the special case when X and Y are jointly normal, uncorrelatedness is equivalent to independence.

Even though uncorrelated data does not necessarily imply independence, one can check if random variables are independent if their mutual information is 0.

Sample correlation coefficient

Given a series of n measurements of the pair (X_i, Y_i) indexed by $i = 1, \dots, n$, the *sample correlation coefficient* can be used to estimate the population Pearson correlation $\rho_{X,Y}$ between X and Y . The sample correlation coefficient is defined as

$$r_{xy} \stackrel{\text{def}}{=} \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

where \bar{x} and \bar{y} are the sample means of X and Y , and s_x and s_y are the corrected sample standard deviations of X and Y .

Equivalent expressions for r_{xy} are

$$\begin{aligned} r_{xy} &= \frac{\sum x_i y_i - n\bar{x}\bar{y}}{n s'_x s'_y} \\ &= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}. \end{aligned}$$

where s'_x and s'_y are the uncorrected sample standard deviations of X and Y .

If x and y are results of measurements that contain measurement error, the realistic limits on the correlation coefficient are not -1 to $+1$ but a smaller range.^[6] For the case of a linear model with a single independent variable, the coefficient of determination (R squared) is the square of r_{xy} , Pearson's product-moment coefficient.

Example

Consider the joint probability distribution of X and Y given in the table below.

$P(X = x, Y = y)$	$y = -1$	$y = 0$	$y = 1$
$x = 0$	0	1/3	0
$x = 1$	1/3	0	1/3

For this joint distribution, the marginal distributions are:

$$P(X = x) = \begin{cases} 1/3 & \text{for } x = 0 \\ 2/3 & \text{for } x = 1 \end{cases}$$

$$P(Y = y) = \begin{cases} 1/3 & \text{for } y = -1 \\ 1/3 & \text{for } y = 0 \\ 1/3 & \text{for } y = 1 \end{cases}$$

This yields the following expectations and variances:

$$\begin{aligned} \mu_X &= 2/3 \\ \mu_Y &= 0 \\ \sigma_X^2 &= 2/9 \\ \sigma_Y^2 &= 2/3 \end{aligned}$$

Therefore:

$$\begin{aligned} \rho_{X,Y} &= \frac{1}{\sigma_X \sigma_Y} \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \frac{1}{\sigma_X \sigma_Y} \sum_{x,y} (x - \mu_X)(y - \mu_Y) P(X = x, Y = y) \\ &= (1 - 2/3)(-1 - 0) \frac{1}{3} + (0 - 2/3)(0 - 0) \frac{1}{3} + (1 - 2/3)(1 - 0) \frac{1}{3} = 0. \end{aligned}$$

Rank correlation coefficients

Rank correlation coefficients, such as Spearman's rank correlation coefficient and Kendall's rank correlation coefficient (τ) measure the extent to which, as one variable increases, the other variable tends to increase, without requiring that increase to be represented by a linear relationship. If, as the one variable increases, the other *decreases*, the rank correlation coefficients will be negative. It is common to regard these rank correlation coefficients as alternatives to Pearson's coefficient, used either to reduce the amount of calculation or to make the coefficient less sensitive to non-normality in distributions. However, this view has little mathematical basis, as rank correlation coefficients measure a different type of relationship than the Pearson product-moment correlation coefficient, and are best seen as measures of a different type of association, rather than as an alternative measure of the population correlation coefficient.^{[7][8]}

To illustrate the nature of rank correlation, and its difference from linear correlation, consider the following four pairs of numbers (x, y):

$$(0, \quad 1), \quad (10, \quad 100), \quad (101, \quad 500), \quad (102, \quad 2000).$$

As we go from each pair to the next pair x increases, and so does y . This relationship is perfect, in the sense that an increase in x is *always* accompanied by an increase in y . This means that we have a perfect rank correlation, and both Spearman's and Kendall's correlation coefficients are 1, whereas in this example Pearson product-moment correlation coefficient is 0.7544, indicating that the points are far from lying on a straight line. In the same way if y always *decreases* when x *increases*, the rank correlation coefficients will be -1 , while the Pearson product-moment correlation coefficient may or may not be close to -1 , depending on how close the points are to a straight line. Although in the extreme cases of perfect rank correlation the

two coefficients are both equal (being both $+1$ or both -1), this is not generally the case, and so values of the two coefficients cannot meaningfully be compared.^[7] For example, for the three pairs (1, 1) (2, 3) (3, 2) Spearman's coefficient is $1/2$, while Kendall's coefficient is $1/3$.

Other measures of dependence among random variables

The information given by a correlation coefficient is not enough to define the dependence structure between random variables.^[9] The correlation coefficient completely defines the dependence structure only in very particular cases, for example when the distribution is a multivariate normal distribution. (See diagram above.) In the case of elliptical distributions it characterizes the (hyper-)ellipses of equal density; however, it does not completely characterize the dependence structure (for example, a multivariate t-distribution's degrees of freedom determine the level of tail dependence).

Distance correlation^{[10][11]} was introduced to address the deficiency of Pearson's correlation that it can be zero for dependent random variables; zero distance correlation implies independence.

The Randomized Dependence Coefficient^[12] is a computationally efficient, copula-based measure of dependence between multivariate random variables. RDC is invariant with respect to non-linear scalings of random variables, is capable of discovering a wide range of functional association patterns and takes value zero at independence.

For two binary variables, the odds ratio measures their dependence, and takes range non-negative numbers, possibly infinity: $[0, +\infty]$. Related statistics such as Yule's Y and Yule's Q normalize this to the correlation-like range $[-1, 1]$. The odds ratio is generalized by the logistic model to model cases where the dependent variables are discrete and there may be one or more independent variables.

The correlation ratio, entropy-based mutual information, total correlation, dual total correlation and polychoric correlation are all also capable of detecting more general dependencies, as is consideration of the copula between them, while the coefficient of determination generalizes the correlation coefficient to multiple regression.

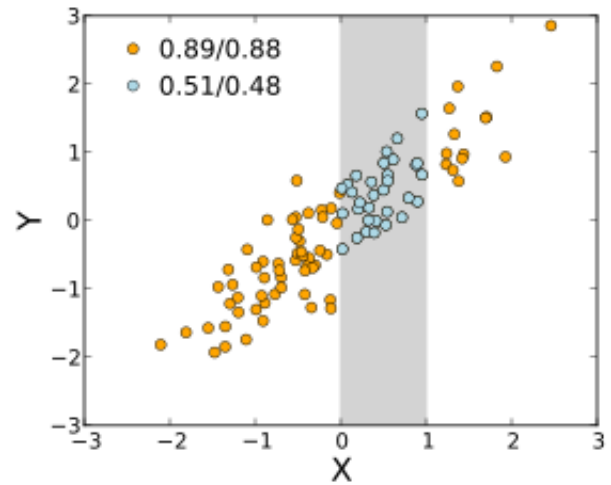
Sensitivity to the data distribution

The degree of dependence between variables X and Y does not depend on the scale on which the variables are expressed. That is, if we are analyzing the relationship between X and Y , most correlation measures are unaffected by transforming X to $a + bX$ and Y to $c + dY$, where a , b , c , and d are constants (b and d being positive). This is true of some correlation statistics as well as their population analogues. Some correlation statistics, such as the rank correlation coefficient, are also invariant to monotone transformations of the marginal distributions of X and/or Y .

Most correlation measures are sensitive to the manner in which X and Y are sampled. Dependencies tend to be stronger if viewed over a wider range of values. Thus, if we consider the correlation coefficient between the heights of fathers and their sons over all adult males, and compare it to the same correlation coefficient calculated when the fathers are selected to be

between 165 cm and 170 cm in height, the correlation will be weaker in the latter case. Several techniques have been developed that attempt to correct for range restriction in one or both variables, and are commonly used in meta-analysis; the most common are Thorndike's case II and case III equations.^[13]

Various correlation measures in use may be undefined for certain joint distributions of X and Y . For example, the Pearson correlation coefficient is defined in terms of moments, and hence will be undefined if the moments are undefined. Measures of dependence based on quantiles are always defined. Sample-based statistics intended to estimate population measures of dependence may or may not have desirable statistical properties such as being unbiased, or asymptotically consistent, based on the spatial structure of the population from which the data were sampled.



Pearson/Spearman correlation coefficients between X and Y are shown when the two variables' ranges are unrestricted, and when the range of X is restricted to the interval $(0,1)$.

Sensitivity to the data distribution can be used to an advantage. For example, scaled correlation is designed to use the sensitivity to the range in order to pick out correlations between fast components of time series.^[14] By reducing the range of values in a controlled manner, the correlations on long time scale are filtered out and only the correlations on short time scales are revealed.

Correlation matrices

The correlation matrix of n random variables X_1, \dots, X_n is the $n \times n$ matrix whose (i, j) entry is $\text{corr}(X_i, X_j)$. If the measures of correlation used are product-moment coefficients, the correlation matrix is the same as the covariance matrix of the standardized random variables $X_i/\sigma(X_i)$ for $i = 1, \dots, n$. This applies both to the matrix of population correlations (in which case σ is the population standard deviation), and to the matrix of sample correlations (in which case σ denotes the sample standard deviation). Consequently, each is necessarily a positive-semidefinite matrix. Moreover, the correlation matrix is strictly positive definite if no variable can have all its values exactly generated as a linear function of the values of the others.

The correlation matrix is symmetric because the correlation between X_i and X_j is the same as the correlation between X_j and X_i .

A correlation matrix appears, for example, in one formula for the coefficient of multiple determination, a measure of goodness of fit in multiple regression.

In statistical modelling, correlation matrices representing the relationships between variables are categorized into different correlation structures, which are distinguished by factors such as the number of parameters required to estimate them. For example, in an exchangeable correlation matrix, all pairs of variables are modeled as having the same correlation, so all non-diagonal elements of the matrix are equal to each other. On the other hand, an autoregressive

matrix is often used when variables represent a time series, since correlations are likely to be greater when measurements are closer in time. Other examples include independent, unstructured, M-dependent, and Toeplitz.

Uncorrelatedness and independence of stochastic processes

Similarly for two stochastic processes $\{X_t\}_{t \in \mathcal{T}}$ and $\{Y_t\}_{t \in \mathcal{T}}$: If they are independent, then they are uncorrelated.^{[15]:p. 151}

Common misconceptions

Correlation and causality

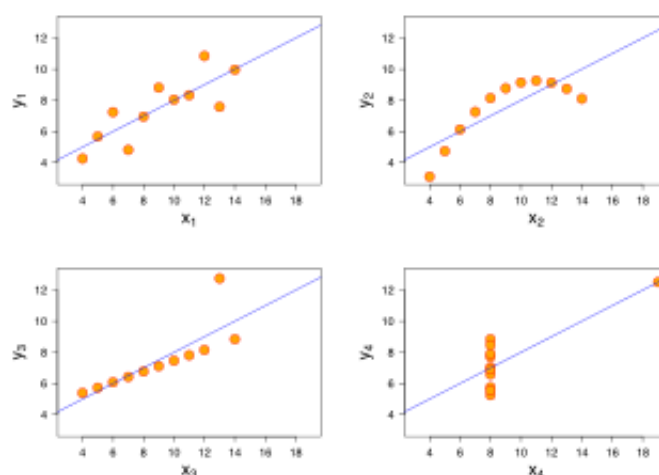
The conventional dictum that "correlation does not imply causation" means that correlation cannot be used by itself to infer a causal relationship between the variables.^[16] This dictum should not be taken to mean that correlations cannot indicate the potential existence of causal relations. However, the causes underlying the correlation, if any, may be indirect and unknown, and high correlations also overlap with identity relations (tautologies), where no causal process exists. Consequently, a correlation between two variables is not a sufficient condition to establish a causal relationship (in either direction).

A correlation between age and height in children is fairly causally transparent, but a correlation between mood and health in people is less so. Does improved mood lead to improved health, or does good health lead to good mood, or both? Or does some other factor underlie both? In other words, a correlation can be taken as evidence for a possible causal relationship, but cannot indicate what the causal relationship, if any, might be.

Simple linear correlations

The Pearson correlation coefficient indicates the strength of a *linear* relationship between two variables, but its value generally does not completely characterize their relationship.^[17] In particular, if the conditional mean of Y given X , denoted $E(Y | X)$, is not linear in X , the correlation coefficient will not fully determine the form of $E(Y | X)$.

The adjacent image shows scatter plots of Anscombe's quartet, a set of four different pairs of variables created by Francis Anscombe.^[18] The four y variables have the same mean (7.5), variance (4.12), correlation (0.816) and regression line ($y = 3 + 0.5x$). However, as can be seen on the plots, the distribution of the variables is very



Four sets of data with the same correlation of 0.816

different. The first one (top left) seems to be distributed normally, and corresponds to what one would expect when considering two variables correlated and following the assumption of normality. The second one (top right) is not distributed normally; while an obvious relationship between the two variables can be observed, it is not linear. In this case the Pearson correlation coefficient does not indicate that there is an exact functional relationship: only the extent to which that relationship can be approximated by a linear relationship. In the third case (bottom left), the linear relationship is perfect, except for one outlier which exerts enough influence to lower the correlation coefficient from 1 to 0.816. Finally, the fourth example (bottom right) shows another example when one outlier is enough to produce a high correlation coefficient, even though the relationship between the two variables is not linear.

These examples indicate that the correlation coefficient, as a summary statistic, cannot replace visual examination of the data. The examples are sometimes said to demonstrate that the Pearson correlation assumes that the data follow a normal distribution, but this is not correct.^[4]

Bivariate normal distribution

If a pair (X, Y) of random variables follows a bivariate normal distribution, the conditional mean $E(X | Y)$ is a linear function of Y , and the conditional mean $E(Y | X)$ is a linear function of X . The correlation coefficient $\rho_{X,Y}$ between X and Y , along with the marginal means and variances of X and Y , determines this linear relationship:

$$E(Y | X) = E(Y) + \rho_{X,Y} \cdot \sigma_Y \frac{X - E(X)}{\sigma_X},$$

where $E(X)$ and $E(Y)$ are the expected values of X and Y , respectively, and σ_X and σ_Y are the standard deviations of X and Y , respectively.

See also

- [Autocorrelation](#)
- [Canonical correlation](#)
- [Coefficient of determination](#)
- [Cointegration](#)
- [Concordance correlation coefficient](#)
- [Cophenetic correlation](#)
- [Correlation function](#)
- [Correlation gap](#)
- [Covariance](#)
- [Covariance and correlation](#)
- [Cross-correlation](#)
- [Ecological correlation](#)
- [Fraction of variance unexplained](#)
- [Genetic correlation](#)
- [Goodman and Kruskal's lambda](#)
- [Illusory correlation](#)
- [Interclass correlation](#)
- [Intraclass correlation](#)
- [Lift \(data mining\)](#)
- [Mean dependence](#)
- [Modifiable areal unit problem](#)
- [Multiple correlation](#)
- [Point-biserial correlation coefficient](#)
- [Quadrant count ratio](#)
- [Spurious correlation](#)
- [Statistical arbitrage](#)
- [Subindependence](#)

References

1. Croxton, Frederick Emory; Cowden, Dudley Johnstone; Klein, Sidney (1968) *Applied General Statistics*, Pitman. ISBN [9780273403159](#) (page 625)
2. Dietrich, Cornelius Frank (1991) *Uncertainty, Calibration and Probability: The Statistics of Scientific and Industrial Measurement* 2nd Edition, A. Higler. ISBN [9780750300605](#) (Page 331)
3. Aitken, Alexander Craig (1957) *Statistical Mathematics* 8th Edition. Oliver & Boyd. ISBN [9780050013007](#) (Page 95)
4. Rodgers, J. L. ; Nicewander, W. A. (1988). "Thirteen ways to look at the correlation coefficient". *The American Statistician*. 42 (1): 59 – 66. doi:[10.1080/00031305.1988.10475524](#) (<https://doi.org/10.1080%2F00031305.1988.10475524>). JSTOR [2685263](#) (<https://www.jstor.org/stable/2685263>).
5. Dowdy, S. and Wearden, S. (1983). "Statistics for Research", Wiley. ISBN [0-471-08602-9](#) pp 230
6. Francis, DP; Coats AJ; Gibson D (1999). "How high can a correlation coefficient be?". *Int J Cardiol*. 69 (2): 185 – 199. doi:[10.1016/S0167-5273\(99\)00028-5](#) ([http://doi.org/10.1016%2FS0167-5273\(99\)00028-5](http://doi.org/10.1016%2FS0167-5273(99)00028-5)).
7. Yule, G.U and Kendall, M.G. (1950), "An Introduction to the Theory of Statistics", 14th Edition (5th Impression 1968). Charles Griffin & Co. pp 258 – 270
8. Kendall, M. G. (1955) "Rank Correlation Methods", Charles Griffin & Co.
9. Mahdavi Damghani B. (2013). "The Non-Misleading Value of Inferred Correlation: An Introduction to the Cointelation Model". *Wilmott Magazine*. 2013 (67): 50 – 61. doi:[10.1002/wilm.10252](#) (<https://doi.org/10.1002%2Fwilm.10252>).
0. Székely, G. J. Rizzo; Bakirov, N. K. (2007). "Measuring and testing independence by correlation of distances". *Annals of Statistics*. 35 (6): 2769 – 2794. arXiv:[0803.4101](#) (<https://arxiv.org/abs/0803.4101>). doi:[10.1214/009053607000000505](#) (<https://doi.org/10.1214%2F009053607000000505>).
1. Székely, G. J. ; Rizzo, M. L. (2009). "Brownian distance covariance" (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2889501>). *Annals of Applied Statistics*. 3 (4): 1233 – 1303. arXiv:[1010.0297](#) (<https://arxiv.org/abs/1010.0297>). doi:[10.1214/09-AOAS312](#) (<https://doi.org/10.1214%2F09-AOAS312>). PMC [2889501](#) (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2889501>). PMID [20574547](#) (<https://pubmed.ncbi.nlm.nih.gov/20574547>).
2. Lopez-Paz D. and Hennig P. and Schölkopf B. (2013). "The Randomized Dependence Coefficient", "Conference on Neural Information Processing Systems" Reprint (<http://papers.nips.cc/paper/5138-the-randomized-dependence-coefficient.pdf>)
3. Thorndike, Robert Ladd (1947). *Research problems and techniques* (Report No. 3). Washington DC: US Govt. print. off.
4. Nikolić, D; Muresan, RC; Feng, W; Singer, W (2012). "Scaled correlation analysis: a better way to compute a cross-correlogram". *European Journal of Neuroscience*. 35 (5): 1 – 21. doi:[10.1111/j.1460-9568.2011.07987.x](#) (<https://doi.org/10.1111%2Fj.1460-9568.2011.07987.x>). PMID [22324876](#) (<https://pubmed.ncbi.nlm.nih.gov/22324876>).
5. Park, Kun Il (2018). *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Springer. ISBN [978-3-319-68074-3](#).
6. Aldrich, John (1995). "Correlations Genuine and Spurious in Pearson and Yule" (<https://doi.org/10.1214%2Fss%2F1177009870>). *Statistical Science*. 10 (4): 364 – 376. doi:[10.1214/ss/1177009870](#) (<https://doi.org/10.1214%2Fss%2F1177009870>). JSTOR [2246135](#) (<https://www.jstor.org/stable/2246135>).
7. Mahdavi Damghani, Babak (2012). "The Misleading Value of Measured Correlation".

Wilmott Magazine. 2012 (1): 64 – 73. doi:10.1002/wilm.10167 (<https://doi.org/10.1002%2Fwilm.10167>).

8. Anscombe, Francis J. (1973). "Graphs in statistical analysis". The American Statistician. 27 (1): 17 – 21. doi:10.2307/2682899 (<https://doi.org/10.2307%2F2682899>). JSTOR 2682899 (<https://www.jstor.org/stable/2682899>).

Further reading

- Cohen, J. ; Cohen P. ; West, S.G. & Aiken, L.S. (2002). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Psychology Press. ISBN 978-0-8058-2223-6.
- "Correlation (in statistics)" ([https://www.encyclopediaofmath.org/index.php?title=Correlation_\(in_statistics\)](https://www.encyclopediaofmath.org/index.php?title=Correlation_(in_statistics))), Encyclopedia of Mathematics, EMS Press, 2001 [1994]
- Oestreicher, J. & D. R. (February 26, 2015). Plague of Equals: A science thriller of international disease, politics and drug discovery. California: Omega Cat Press. p. 408. ISBN 978-0963175540.

External links

- MathWorld page on the (cross-)correlation coefficient/s of a sample (<http://mathworld.wolfram.com/CorrelationCoefficient.html>)
 - Compute significance between two correlations (<http://peaks.informatik.uni-erlangen.de/cgi-bin/usignificance.cgi>), for the comparison of two correlation values.
 - A MATLAB Toolbox for computing Weighted Correlation Coefficients (<http://www.mathworks.com/matlabcentral/fileexchange/20846>)
 - [1] (<https://www.scribd.com/doc/299546673/Proof-that-the-Sample-Bivariate-Correlation-has-limits-plus-or-minus-1>) Proof-that-the-Sample-Bivariate-Correlation-has-limits-plus-or-minus-1
 - Interactive Flash simulation on the correlation of two normally distributed variables (http://nagysandor.eu/AsimovTeka/correlation_en/index.html) by Juha Puranen.
 - Correlation analysis. Biomedical Statistics (<https://web.archive.org/web/20150407112430/http://www.biostat.katerynakon.in.ua/en/association/correlation.html>)
 - R-Psychologist Correlation (<http://rpsychologist.com/d3/correlation/>) visualization of correlation between two numeric variables
-

Retrieved from "https://en.wikipedia.org/w/index.php?title=Correlation_and_dependence&oldid=991370730"

This page was last edited on 29 November 2020, at 18:22 (UTC).

Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.