

Correlação

Origem: Wikipédia, a enciclopédia livre.

Este artigo abrange a correlação entre duas variáveis. O termo correlação pode também significar a relação invertida de duas funções ou a correlação eletrônica em sistemas moleculares.

Em probabilidade e estatística, **correlação**, dependência ou associação é qualquer relação estatística (causal ou não causal) entre duas variáveis.^[1] e correlação é qualquer relação dentro de uma ampla classe de relações estatísticas que envolva dependência entre duas variáveis.^[2] Por exemplo, a correlação entre a estatura dos pais e a estatura dos pais e dos filhos. Embora seja comumente denotada como a medida de relação entre duas variáveis aleatórias,^{[2][3]} correlação não implica causalidade.^[4] Em alguns casos, correlação não identifica dependência entre as variáveis. Em geral, há pares de variáveis que apresentam forte dependência estatística, mas que possuem correlação nula. Para este casos, são utilizadas outras medidas de dependência.^[5]

Informalmente correlação é sinônimo de dependência. Formalmente variáveis são dependentes se não satisfizerem a propriedade matemática da independência probabilística. Em termos técnicos, correlação refere-se a qualquer um dos vários tipos específicos de relação entre os valores médios. Existem diferentes coeficientes de correlação (***ρ*** ou ***r***) para medir o grau de correlação. Um dos coeficientes de correlação mais conhecidos é o coeficiente de correlação de Pearson, obtido pela divisão da covariância de duas variáveis pelo produto dos seus desvios padrão^[6] e sensível a uma relação linear entre duas variáveis.^[7] Entretanto, há outros coeficientes de correlação mais robustos que o coeficiente de correlação de Pearson. Isto é, mais sensíveis às relações não lineares.^{[8][9][10]}

Índice

Histórico

- Weldon e Pearson
- Edgeworth

Correlação linear

- Correlação populacional
- Correlação amostral
 - Coeficiente de correlação de Pearson
 - Propriedades de ***r*** para variáveis ***x*** e ***y***
 - Interpretação geométrica

Correlação não linear

- Coeficiente de correlação ***η*** (ETA)
- Correlação de postos de Spearman
- Correlação de postos de Kendall

Classificação dos coeficientes de correlação

Regressão linear

- Regressão linear simples
- Método dos mínimos quadrados

Exemplos

- Área comercial (correlação linear)
- Área da saúde (correlação não linear)

Outras medidas de dependência entre variáveis

Sensibilidade à distribuição dos dados

Matrizes de correlação

- Interpretação
- Dependência

Precauções

- Correlação e causalidade
- Correlação e linearidade

Ver também
Referências
Leitura adicional
Ligações externas

Histórico

Os conceitos de correlação e de regressão originaram—se nos anos de 1880 com Sir Francis Galton (1822 – 1911), de acordo com sua autobiografia *Memories of My Life*, publicada em 1890.^[11] Em seus estudos sobre o diâmetro das sementes das ervilhas—de—cheiro, Galton chegou à equação de reversão à média:

$$y - \bar{y} = r(x - \bar{x}),$$

em que *x* são o diâmetro das sementes das plantas filhas, *y* são o diâmetro das sementes das plantas pais e *r* é a reversão (posteriormente, regressão). Em suas pesquisas sobre a comparação entre estaturas de pais e filhos, Galton usou o termo regressão pela primeira vez para denotar a regressão à média da população observada. Com base na análise da tábua de frequência bidimensional representada pelas alturas dos filhos adultos e meio—pais (média do pai e da mãe), Galton chegou à fórmula da superfície da representação gráfica (atualmente, a fórmula da função normal bidimensional) e à expressão da co—relação (atualmente, correlação) como consequência das variações devidas a causas comuns.^[12]

O desenvolvimento da correlação, com a publicação de *Natural Inheritance* em 1889, marca o início do período moderno da estatística. A correlação prometia ser útil em muitos campos científicos, especialmente diante da necessidade de estabelecer diretrizes de causalidade. Entretanto, na época de Galton a disciplina de teoria estatística era não existia na Inglaterra (não havia curso universitário revistas especializadas sobre o assunto). Então, Galton percebeu que regressão e correlação poderiam ser o ponto de partida de uma nova disciplina, embora fosse preciso organizar as contribuições dispersas e estruturar o conhecimento de modo sistemático. Por volta de 1890, este contexto começou a chamar atenção de estatísticos e jovens cientistas como Edgeworth, Weldon e Pearson.^[13]

Weldon e Pearson

A expressão de correlação foi modificada pelo professor de Zoologia em Cambridge Walter Frank Raphael Weldon (1860 – 1906), que incorporou o sinal positivo e o sinal negativo a partir da análise das dimensões de 22 pares de órgãos de várias espécies de camarões. Entretanto, a atual fórmula do coeficiente de correlação foi determinada em 1896 por Karl Pearson (1857 – 1936).^[12] Em *Mathematical Contributions to the Theory of Evolution*, Pearson publicou entre 1893 e 1912 suas contribuições para a análise de regressão, o coeficiente de correlação e o teste qui—quadrado de significância estatística. Especialmente, o coeficiente de correlação produto—momento de Pearson com foi desenvolvido com base em um conjunto de dados sobre estatura de pais e filhos (inicialmente coletados por Galton), tornando—se a primeira medida de força de associação em estatística.^[14]

Edgeworth

Em 1892, Francis Ysidro Edgeworth (1845 – 1926) publicou parte do seu trabalho sobre correlação em *Philosophical Magazine* com base nas contribuições prévias de Galton. Nos artigos *Correlated Averages* e *The Law of Error and Correlated Averages*, Edgeworth introduziu o termo coeficiente de correlação, forneceu a primeira análise matemática de correlação e mostrou sua relação com a distribuição multivariada normal. Entre 1893 e 1895, Edgeworth publicou a continuação do seu trabalho sobre correlação na *Philosophical Magazine* e na *Journal of the Royal Statistical Society*. Por exemplo, o resumo de *Recent Contributions to the Theory of Statistics*, que culminou em quatro importantes trabalhos *Statistics on Unprogressive Communities*, *The Assymetrical Probability Curve*, *The Compound Law of Error* e *Supplementary Notes on Statistics*.^[15]

Correlação linear

Os diferentes setores da sociedade utilizam amostras *θ* de populações *α*, em vez de populações de dados *α*, para auxiliar nas análises dos objetivos das suas áreas de conhecimento. Isto ocorre porque um conjunto reduzido de dados (*θ* < *α*) torna o processo mais econômico e mais rápido do ponto de vista matemático. Isto deve—se aos erros de arredondamentos dos cálculos que são feitos em estatística, considerando que uma grande quantidade de dados envolve mais cálculos que uma pequena quantidade de dados.^[16]

Correlação populacional

A correlação populacional trata da medida da direção e do grau com que as variáveis \mathbf{X} e \mathbf{Y} se associam linearmente em uma população. O coeficiente de correlação populacional $\rho_{X,Y}$ entre duas variáveis \mathbf{X} e \mathbf{Y} com valores esperados μ_X e μ_Y e desvios padrão σ_X e σ_Y é definido como:

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

em que E é o operador do valor esperado, cov é a covariância e corr é uma notação alternativa para coeficiente de correlação.^[17]

O coeficiente de correlação é simétrico: $\text{corr}(X, Y) = \text{corr}(Y, X)$.^[18]

O coeficiente de correlação de Pearson é +1 no caso de uma relação linear (correlação) direta perfeita (crescente), -1 no caso de uma relação linear (anticorrelação) decrescente perfeita (inversa) e qualquer valor no intervalo aberto (-1, 1) nos outros casos indicando o grau de dependência linear entre as variáveis.^[19] É um corolário da desigualdade de Cauchy–Schwarz, em que a correlação não pode exceder 1 em valor absoluto. Quanto mais próximo de 0, mais fraca a correlação entre as variáveis (mais próximas de não correlacionados). Quando mais próximo de -1 ou +1, mais forte a correlação entre as variáveis.^[20]

Se as variáveis forem independentes, o coeficiente de correlação de Pearson é 0. Entretanto, o contrário não é verdadeiro porque o coeficiente de correlação detecta apenas dependências lineares entre duas variáveis. Por exemplo, suponha-se que a variável aleatória \mathbf{X} é simetricamente distribuída em torno de 0 e que a variável aleatória $\mathbf{Y} = \mathbf{X}^2$. Então, \mathbf{Y} é completamente determinada por \mathbf{X} , de modo que \mathbf{X} e \mathbf{Y} são perfeitamente dependentes, mas a correlação entre elas é 0. Em outras palavras, as variáveis não são correlacionadas. Entretanto, no caso especial em que \mathbf{X} e \mathbf{Y} são conjuntamente normais, não correlação é equivalente à independência.^[17]

Correlação amostral

Coeficiente de correlação de Pearson

A correlação amostral trata da medida da direção e do grau com que as variáveis \mathbf{X} e \mathbf{Y} se associam linearmente em uma amostra. Karl Pearson desenvolveu o coeficiente amostral a partir de uma ideia semelhante, porém ligeiramente diferente da de Francis Galton. Então, o coeficiente amostral pode ser chamado de coeficiente produto–momento de Pearson, coeficiente de correlação de Pearson ou simplesmente coeficiente de correlação, que é a medida mais conhecida de dependência entre duas variáveis quantitativa.^[7]

Para uma série de n medições de \mathbf{X} e \mathbf{Y} , x_i e y_i para $i = 1, 2, \dots, n$, o coeficiente de correlação da amostra pode ser usado para estimar o coeficiente de correlação de Pearson da população r entre \mathbf{X} e \mathbf{Y} . Então, o coeficiente de correlação da amostra é escrito como:

$$r_{xy} = \frac{\text{Cov}(X, Y)}{S_x S_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

em que \bar{x} e \bar{y} são as médias amostrais de \mathbf{X} e \mathbf{Y} .^[6]

O coeficiente de correlação de Pearson r_{xy} é definido apenas se os desvios padrão S_x e S_y forem finitos e diferentes de zero. Isto também pode ser escrito como:

$$r_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n s_x s_y} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}. \quad [6]$$

Se x e y são resultados de medições que contêm erros de medições, os limites realistas no coeficiente de correlação não são -1 para +1, mas um intervalo menor.^[21]

No caso de um modelo linear com uma única variável independente, o coeficiente de determinação R^2 é o quadrado do coeficiente de correlação de Pearson r .^[22]

Propriedades de r para variáveis x e y

1. $-1 < r < 1$

2. $r = 1$: relação linear perfeita e positiva
3. $r = 0$: inexistência linear
4. $r = -1$: relação linear perfeita e negativa
5. $r > 0$: relação linear positiva
6. $r < 0$: relação linear negativa
7. Se os valores das variáveis x e y são convertidos para uma escala diferente, o valor de r não é alterado
8. O valor de r não é afetado pela escolha das variáveis x ou y
9. Permutando x e y , o valor de r não é alterado
10. Mede direção e o grau (intensidade) com que as variáveis x e y as quais se associam linearmente
11. Não é útil para casos não-lineares^[23]

Interpretação geométrica

As duas séries de valores $\mathbf{X} = (x_1, \dots, x_n)$ e $\mathbf{Y} = (y_1, \dots, y_n)$ podem ser consideradas vetores em um espaço de dimensão n . Com a substituição por vetores centrados nas médias, têm-se $\mathbf{X} = (x_1 - \bar{x}, \dots, x_n - \bar{x})$ e $\mathbf{Y} = (y_1 - \bar{y}, \dots, y_n - \bar{y})$.

O cosseno do ângulo α entre os vetores \mathbf{X} e \mathbf{Y} , usando a norma euclidiana e o produto escalar normalizado, é dado pela fórmula:^[24]

$$\cos(\alpha) = \frac{\sum_{i=1}^N (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}},$$

Portanto, $\cos(\alpha) = r_{xy}$, em que r_{xy} está sempre -1 e 1.

O coeficiente de correlação é o cosseno do ângulo α entre os dois vetores centrados:^[24]

- Se $\alpha = 0^\circ$, então $r = 1$ e os dois vetores são colineares (paralelos);
- Se $\alpha = 90^\circ$, então $r = 0$ e os dois vetores são ortogonais;
- Se $\alpha = 180^\circ$, então $r = -1$ e os dois vetores são colineares, mas em direções opostas.

Mais genericamente, $\alpha = \arccos(r)$, em que \arccos é a função inversa do cosseno. α representa, do ponto de vista geométrico, a intensidade da correlação entre os dois vetores aleatórios \mathbf{X} e \mathbf{Y} , o que não pode ser medido em um teste de significância.^[24]

Mais genericamente, $\alpha = \arccos(r)$, em que \arccos é o inverso da função cosseno. Do ponto de vista geométrico, não se fala em *correlação linear*. O coeficiente de correlação sempre tem um significado, independentemente do valor entre -1 e 1.

Correlação não linear

Coeficiente de correlação η (ETA)

O coeficiente de correlação η (ETA) ou razão de correlação mede a intensidade de associação entre variável dependente e independente para casos não-lineares. Isto é, trata a relação de uma variável quantitativa y e outra variável categórica ou nominal x . Expressa-se em percentagem. Assume valores entre 0 e 1.^[25]

O coeficiente de correlação η é dado pela expressão matemática:

$$\eta^2(G|X) = \frac{\text{Var}(\mathbb{E}(G|X))}{\text{Var}(G)}$$

Onde, $G = G(x_1, x_2, x_3, \dots, x_n)$ é uma função. Ou seja, $Y = G$.^[26]

Correlação de postos de Spearman

A correlação de postos de Spearman é utilizada para a relação entre dados não-lineares. Quando os dados de uma amostra são ordenados de forma crescente, obtém-se uma condição de ordem para que cada elemento ordenado seja um posto.

A correlação de postos de Spearman é dada pela expressão matemática:

$$r_s = 1 - \frac{6 \cdot \sum_{i=1}^n d_i^2}{n^3 - n},$$

em que $d_i = r_{x_i} - r_{y_i}$, em que r_{x_i} e r_{y_i} variando de 1 a n , sendo n a quantidade de elementos da amostra.^[27]

Correlação de postos de Kendall

A correlação de postos de Kendall, também conhecido como τ de Kendall, é uma medida a partir de uma amostra de dados ordenados como a correlação de postos de Spearman. A vantagem de τ de Kendall é a generalização para um coeficiente de correlação parcial.

A correlação τ de Kendall é dada pela expressão matemática:

$$\tau = \frac{n_c - n_d}{\frac{1}{2}n(n-1)},$$

em que n_c são os números de pares concordantes, n_d são os números de pares discordantes e n é a quantidade de elementos da amostra.^[28]

Classificação dos coeficientes de correlação

Coefficientes de correlação como correlação de postos de Spearman e τ de Kendall medem a como uma variável aumenta enquanto outra variável também aumenta, sem exigir que o aumento seja representado por uma relação linear. Se, à medida que uma variável aumenta, uma outra variável diminui, os coeficientes de correlação de postos serão negativos. É comum considerar estes coeficientes de correlação de postos como alternativas ao coeficiente de correlação de Pearson, usado tanto para reduzir a quantidade de cálculo quanto para tornar o coeficiente menos sensível para distribuições não normais. Entretanto, esta visão tem pouca base matemática, na medida em que coeficientes de correlação de postos medem um tipo de relação diferente do coeficiente de correlação de Pearson e são melhor vistos como medidas de um tipo de associação diferente em vez de medidas alternativas do coeficiente de correlação da população.^{[29][30]}

Para ilustrar a natureza dos coeficientes de correlação de postos, sejam os seguintes pares de números $(x, y) : (0, 1), (10, 100), (101, 500), (102, 2000)$. De um par para outro, x e y aumentam. A relação é perfeita no sentido que o aumento de x sempre é acompanhado do aumento de y . Isto corresponde a uma correlação de posto perfeita e tanto a correlação de postos de Spearman quanto o τ de Kendall são iguais a 1, enquanto que o coeficiente de correlação de Pearson é igual a 0,7544, indicando que os pontos estão longe de cair em uma linha reta. Da mesma maneira, se y sempre diminuir quando x aumentar, o coeficiente de correlação de postos será -1 enquanto o coeficiente de correlação de Pearson poderá ou não poderá ser próximo de -1, dependendo do quão próximo os pontos estiverem de uma linha reta. Embora em casos extremos de correlação de posto perfeita ambos os coeficientes seja iguais (+1 ou -1), este geralmente não é o caso. Então, valores de dois coeficientes não podem ser comparados. Por exemplo, para os três pares $(1, 1)(2, 3)(3, 2)$ a correlação de postos de Spearman é $\frac{1}{2}$ e o τ de Kendall é $\frac{1}{3}$.^[29]

Regressão linear

A regressão linear estuda a relação entre muitas ou poucas variáveis. A relação entre duas variáveis é chamada de regressão linear simples. A relação entre mais de duas variáveis é chamada de regressão linear múltipla.^[31]

Regressão linear simples

A regressão linear simples estuda a relação entre uma variável independente x e outra variável dependente y , ambas quantitativas. A regressão linear resulta em uma equação sobre o comportamento das variáveis x e y :

$$y_i = \alpha + \beta x_i + \varepsilon_i,$$

em que y_i é a variável resposta associada a i -ésima observação de y , x_i é a i -ésima observação do valor independente e fixado para a variável independente x , ε_i é o erro aleatório para i -ésima observação e, α e β são os parâmetros que precisam ser estimados.^[32]

Por meio da correlação linear, encontra-se o gráfico de uma reta que possui as características de coeficiente angular e de coeficiente linear. Em termos matemáticos, a equação geral de uma reta é dada por $y = ax + b$, em que a é o coeficiente angular e b é o coeficiente linear. Como a regressão linear é uma equação de uma reta, em que α é o coeficiente linear e β o coeficiente angular para o mesmo grupo de dados amostrais, admite-se $\alpha = b$ e $\beta = a$. Portanto, encontrando-se uma correlação linear de uma amostra de dados, obtém-se a estimação dos parâmetros α e β para a regressão linear.^[32]

Uma regressão linear é construída a partir de uma reta r_1 , determinada por uma correlação linear. Então, r_1 possui pontos cartesianos para determinar o coeficiente linear e o coeficiente angular para estabelecer os parâmetros α e β da regressão linear. Como x_i e y_i são observações amostrais e ϵ_i é o erro amostral, encontra-se o valor y_i . Uma reta de regressão pode ser determinada somente com os valores α e β .^[32]

Método dos mínimos quadrados

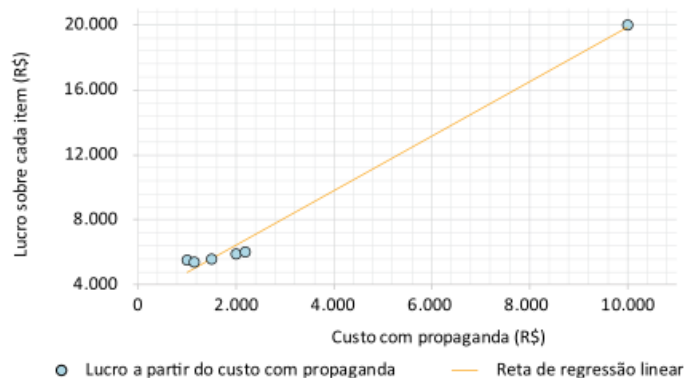
O cálculo do coeficiente de correlação entre duas variáveis é dado pelo ajuste linear. Para o cálculo das características da linha, o erro cometido ao representar a relação entre as variáveis precisa ser o menor possível. Embora não seja o único critério possível, na maioria das vezes minimiza-se a soma de todos os erros cometidos ao quadrado. Isto é chamado de ajuste para o método dos mínimos quadrados, que resulta na linha de regressão (quanto melhor a qualidade de representação da relação entre as variáveis pela linha de regressão, maior a associação do coeficiente de correlação linear).^[33] O método dos mínimos quadrados é uma forma ágil para determinar os parâmetros desconhecidos para uma equação resultante de uma regressão linear, uma vez que minimiza a soma dos quadrados dos resíduos. O método dos mínimos quadrados é útil se os resíduos tiverem distribuição normal.^[34]

Exemplos

Área comercial (correlação linear)

O setor comercial procura investir em propagandas para aumentar as vendas. Um gráfico de dispersão mostra se há uma correlação linear entre vendas e custo de propaganda. Supondo que uma empresa **A** tem o cenário conforme a tabela abaixo.^[35]

Tabela A	
Custo com propaganda	Lucro sobre cada item
R\$ 1.000,00	R\$ 5.500,00
R\$ 1.150,00	R\$ 5.400,00
R\$ 1.500,00	R\$ 5.600,00
R\$ 2.000,00	R\$ 5.900,00
R\$ 2.200,00	R\$ 6.000,00
R\$ 10.000,00	R\$ 20.000,00



A partir do gráfico de dispersão da tabela acima, encontra-se uma correlação linear que mostra uma tendência entre investimento em propaganda e retorno positivo em vendas. É possível estimar um retorno de vendas com um aumento no custo em propaganda pelo cálculo matricial por meio da equação geral da reta $ax + b = c$, em que x são os valores independentes que correspondem aos custos de propaganda e c são os valores de retorno tratados para o cálculo do vetor chamado na tabela A de lucro sobre cada item. Esse cálculo é uma generalização método dos mínimos quadrados.^[35]

Então,

$$ax + b = c \Rightarrow \begin{matrix} 1,00a + b = 5,5 \\ 1,15a + b = 5,4 \\ 1,50a + b = 5,6 \\ 2,00a + b = 5,9 \\ 2,20a + b = 6,0 \\ 10,00a + b = 20,0 \end{matrix} \Rightarrow A = \begin{bmatrix} 1,00 & 1,0 \\ 1,15 & 1,0 \\ 1,50 & 1,0 \\ 2,00 & 1,0 \\ 2,20 & 1,0 \\ 10,00 & 1,0 \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} 1,0 & 1,15 & 1,5 & 2,0 & 2,2 & 10,0 \\ 1,0 & 1,0 & 1,0 & 1,0 & 1,0 & 1,0 \end{bmatrix}$$

$$c = \begin{bmatrix} 5,5 \\ 5,4 \\ 5,6 \\ 5,9 \\ 6,0 \\ 20,0 \end{bmatrix}$$

A relação entre as variáveis é:

$$(A^T \cdot A)x = A^T c,$$

em que $\mathbf{x} = \begin{bmatrix} a \\ b \end{bmatrix}$.

Realizando-se o cálculo entre as matrizes, obtém-se:

$$(A^T \cdot A)\mathbf{x} = A^T \mathbf{c} \Rightarrow \begin{bmatrix} 113,41 & 17,85 \\ 17,85 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 245,11 \\ 48,4 \end{bmatrix} \Rightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 113,41 & 17,85 \\ 17,85 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 245,11 \\ 48,4 \end{bmatrix} \Rightarrow \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1,677 \\ 3,078 \end{bmatrix}$$

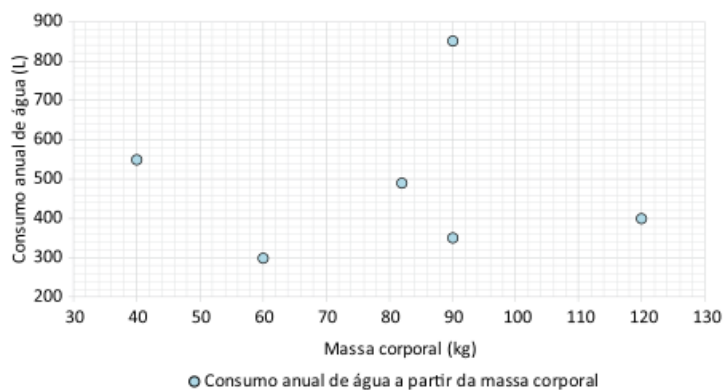
Portanto, com um investimento de 12 mil reais em propaganda no próximo mês, o lucro será de:

$$1,677x + 3,078 = c \Rightarrow 1,677(12) + 3,078 = c \Rightarrow c = 23,202 \text{ (23 mil e 202 reais)}$$

Área da saúde (correlação não linear)

Para se manter saudável, o consumo de água é indispensável. Uma gestora pública solicita à uma consultoria para verificar se há relação entre o consumo de água e a massa corporal de uma pessoa. A tabela B mostra indivíduos que fazem o consumo de água.^[36]

Tabela B		
Indivíduo	X: Massa corporal (kg)	Y: Consumo anual de água (L)
1	90	850
2	120	400
3	60	300
4	40	550
5	82	490
6	90	350



Com a elaboração do diagrama de dispersão sobre a Tabela B, torna-se intuitivo pela plotagem (visualização) a não linearização entre massa e consumo de água. Embora necessárias, a intuição e a visualização não são suficientes para provar que não há relação linear. Dessa forma, o cálculo pelo coeficiente de Pearson facilita a demonstração da ausência de linearização para o objetivo da gestora pública. Para o cálculo é utilizado neste exemplo a segunda versão do coeficiente de Pearson, embora visualmente a fórmula seja mais extensa, ela é mais direta. Então,

$$r_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad [6]$$

Análise rápida sobre o cálculo

Neste exemplo, o coeficiente de Pearson tem o numerador $n \sum x_i y_i - \sum x_i \sum y_i$ igual a zero. Portanto, a razão é zero, o que significa inexistência linear. Em cálculo,

$$n \sum x_i y_i - \sum x_i \sum y_i = (6 \cdot 236.180) - (482 \cdot 2.940) = 1.417.080 - 1.417.080 = 0 \quad [6]$$

A consultoria retorna a demonstração e a plotagem dos dados para a gestora pública que determina os planos de ações sobre o controle da reserva de água.^[36]

Outras medidas de dependência entre variáveis

A informação fornecida por um coeficiente de correlação não é suficiente para definir a estrutura de dependência entre variáveis.^[37] O coeficiente de correlação define completamente a estrutura de dependência apenas em casos particulares como uma distribuição normal multivariada. No caso de distribuições elípticas, caracteriza as (hiper)elipses de igual densidade. Entretanto, não caracteriza completamente a estrutura de dependência. Por exemplo, os graus de liberdade de uma distribuição t multivariada determinam o nível de dependência de cauda.^[38]

Correlação da distância — Foi introduzida para endereçar a deficiência da correlação de Pearson que pode ser 0 para variáveis dependentes. Uma correlação de distância igual a 0 implica independência.^{[39][40]}

Coefficiente de dependência randomizado — É uma medida baseada em cópula computacionalmente eficiente da dependência entre variáveis multivariadas. O coeficiente de dependência randomizado é invariante em relação a escalas não lineares de variáveis e é capaz de descobrir uma ampla variedade de padrões de associação funcionais e assume o valor 0 na independência.^[41]

Proporção de correlação — É capaz de detectar praticamente qualquer dependência funcional.^{[42][43]}

Informação mútua baseada na entropia / Correlação total / Correlação total dupla — São capazes de detectar ainda mais dependências gerais. Estas medidas de dependência às vezes são referidas como medidas de correlação de múltiplos momentos em comparação com aquelas medidas de dependência que consideram apenas a dependência do segundo momento (pairwise ou quadrática).^[43]

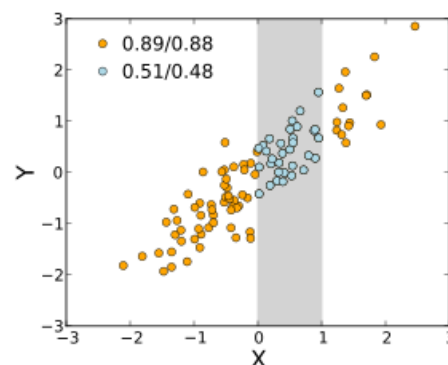
Correlação policórica — É outra correlação aplicada a dados ordinais que buscam estimar a correlação entre as variáveis latentes teorizadas. Uma forma de capturar uma visão mais completa da estrutura de dependência é considerar uma cópula entre elas.^[43]

Coefficiente de determinação — Generaliza o coeficiente de correlação para relações além da regressão linear simples.^[44]

Sensibilidade à distribuição dos dados

O grau de dependência entre as variáveis X e Y não depende da escala, na qual as variáveis são expressas. Isto é, quando se analisa a relação entre X e Y , a maior parte das medidas de correlação não são afetadas pela transformação de X em $a + bX$ e de Y em $c + dY$, em que a, b, c, d são constantes (sendo b, d positivos). Isto é verdade para algumas estatísticas de correlação, assim como para os análogos da sua população. Algumas estatísticas de correlação como o coeficiente de correlação de postos também são invariantes para transformações monótonas de distribuições marginais de X e/ou Y .^[45]

A maioria das medidas de correlação são sensíveis à maneira pela qual X e Y são amostradas. As dependências tendem a ser mais fortes se vistas sobre uma variedade mais ampla de valores. Portanto, se for considerado o coeficiente de correlação entre as alturas dos pais e filhos entre todos os homens adultos e isto for comparado com o mesmo coeficiente de correlação calculado quando os pais selecionado têm entre 1,65m e 1,70m de altura, a correlação será mais fraca no último caso. Várias técnicas que foram desenvolvidas que tentam corrigir a restrição do intervalo em uma ou em ambas as variáveis são comumente usadas em metanálise. As equações do caso III e do caso II de Thorndike são as mais comuns.^[46]



Os coeficientes de correlação de Pearson / Spearman entre x e y são mostrados quando os intervalos das duas variáveis são irrestritos e quando o intervalo de x é restrito ao intervalo (0,1).

Várias medidas de correlação em uso podem ser indefinidas para certas distribuições conjuntas de X e Y . Por exemplo, o coeficiente de correlação de Pearson é definido em termos de momentos. Portanto, será indefinido se os momentos também forem indefinidos. As medidas de dependência baseadas em quantis são sempre definidas. As estatísticas baseadas em amostras que estimam as medidas populacionais de dependência podem ou não podem ter propriedades estatísticas desejáveis, como não ser viesada ou assintoticamente consistente, com base na estrutura espacial da população a partir da qual os dados foram amostrados.^[carece de fontes?]

A sensibilidade à distribuição dos dados tem uma vantagem. Por exemplo, a correlação em escala é designada para usar a sensibilidade ao intervalo para identificar correlações entre componentes rápidos de séries temporais. Reduzindo o intervalo de valores de maneira controlada, as correlações em uma longa escala de tempo são filtradas e apenas as correlações em uma curta escala de tempo são reveladas.^[47]

Matrizes de correlação

A matriz de correlação de n variáveis X_1, \dots, X_n é a matriz $n \times n$, cujo i, j -ésimo elemento da matriz é $\text{corr}(X_i, X_j)$. Se as medidas de correlação usadas são coeficientes de produto—momento, a matriz de correlação é igual a matriz de covariância das variáveis padronizadas $\frac{X_i}{\sigma(X_i)}$ para $i = 1, \dots, n$. Isto é aplicado tanto para matriz de correlações populacionais (caso em que σ

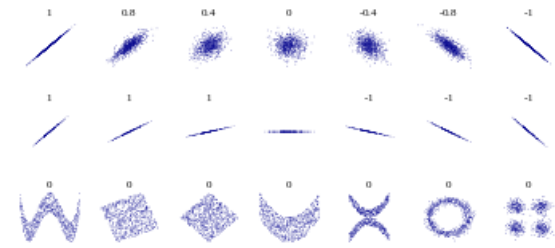
denota o desvio padrão populacional) quanto para matriz de correlações amostrais (caso em que σ denota o desvio padrão amostral). Consequentemente, cada uma é necessariamente uma matriz semidefinida—positiva. Além disso, a matriz de correlação é estritamente definida positiva se nenhuma variável puder ter todos os seus valores exatamente gerados como uma combinação linear dos outros. A matriz de correlação é simétrica porque a correlação entre X_i e Y_j é igual a correlação entre X_j e Y_i . Por exemplo,

uma matriz de correlação aparece, por exemplo, na fórmula para o coeficiente de correlação múltipla (**R**), uma medida de qualidade de ajuste em regressão múltipla: $R^2 = c^T R_{XX}^{-1} c$, onde **c** é o vetor de correlações entre as variáveis independentes **X_n** e a variável dependente **Y_i** e **R_{XX}** é a matriz de correlação.

Interpretação

É igual a 1, se uma das variáveis é uma função linear crescente da outra variável. É igual a -1, se uma variável é uma função decrescente. Os valores intermédios fornecem informações sobre o grau de dependência linear entre as duas variáveis.^[19] Quanto mais próximo o coeficiente estiver dos valores extremos -1 e 1, mais forte é a correlação linear entre as variáveis.^[20] Uma correlação igual a 0 significa que as variáveis não estão linearmente correlacionadas, mas podem estar correlacionado de forma não linear.^[17]

Correlação	Negativa	Positiva
Baixa	de -0,5 à 0	de 0 à 0,5
Alta	de -1 à -0,5	de 0,5 à 1



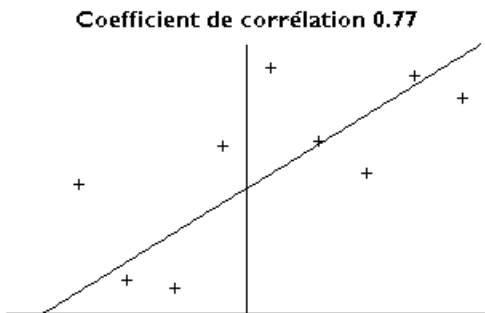
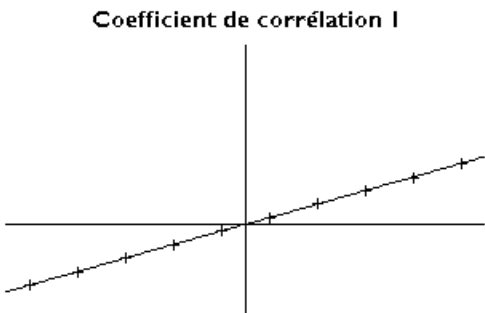
Exemplos de coeficientes de correlação. Nas duas primeiras linhas, ocorre correlação linear. Na terceira linha, ocorre correlação não linear.

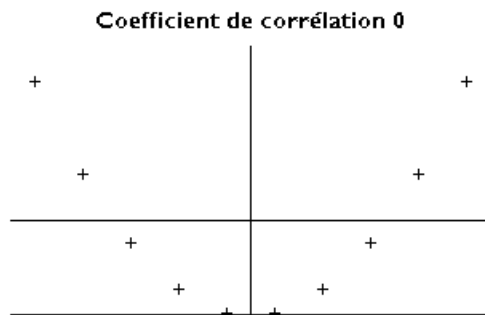
O coeficiente de correlação não é sensível às unidades de cada variável. Por exemplo, o coeficiente de correlação linear entre a idade e o peso de uma pessoa será a mesma idade medida em semanas, meses ou anos. Entretanto, este coeficiente de correlação é extremamente sensível à presença de outliers no conjunto de dados (valores muito discrepantes, que podem ser considerados exceções).^[48]

Vários autores propuseram diretrizes para a interpretação de um coeficiente de correlação. No entanto, Jacob Cohen observa que estes critérios são de certa forma arbitrários e não devem ser muito estritamente observados. A interpretação de um coeficiente de correlação depende do contexto e dos objetivos. Por exemplo, uma correlação de 0,9 pode ser considerada muito baixa se uma lei física for verificada utilizando ferramentas de qualidade, mas pode ser considerada muito alto nas ciências sociais em que pode haver uma maior presença de fatores complicadores.^[49]

Dependência

É sempre possível calcular um coeficiente de correlação para variáveis (exceto para casos muito especiais), mas este coeficiente de correlação nem sempre é capaz de explicar a relação entre as variáveis.^[50] Em uma tentativa de julgar a existência de uma relação linear entre duas variáveis, não é adequado julgar as correlações quando as relações entre elas não são lineares ou quando os dados são muito heterogêneos, uma vez que não há uma relação de causa e efeito entre as variáveis.^[50] Se as duas variáveis aleatórias forem completamente independentes, então a correlação entre elas é nula. Entretanto, o inverso não necessariamente é verdadeiro, isto é, coeficiente de correlação nulo não necessariamente implica independência entre as variáveis, pois se as variáveis apresentarem uma relação não linear ($Y = X^2$, por exemplo), o coeficiente de correlação será nulo, mesmo sendo relacionadas.^[51] Estas considerações são ilustradas pelos seguintes exemplos no domínio de estatísticas:





Por exemplo, em antropometria são medidos para um certo número de indivíduos a estatura, a altura do busto e o comprimento do membro superior. Quanto maiores as medições, mais representativas são as correlações. Então, calculando por leis estatísticas a influência de algumas variáveis sobre outras, obtém-se a matriz ao lado (com valores fictícios).

MATRIZ DE CORRELAÇÃO			
	Estatura	Busto	Membro Superior
Estatura	1		
Busto	0,85	1	
Membro Superior	0,55	0,63	1

- O valor 1 significa que as duas variáveis são exatamente correlacionadas. É o caso de uma relação linear entre as duas variáveis;
- O valor 0,85 significa que a estatura responde por 72,25% ($0,85 \times 0,85$) do valor da altura do busto e assim por diante;
- A outra metade em falta da matriz pode ser completado por uma simetria ao longo da diagonal se as correlações forem reversíveis.

Precauções

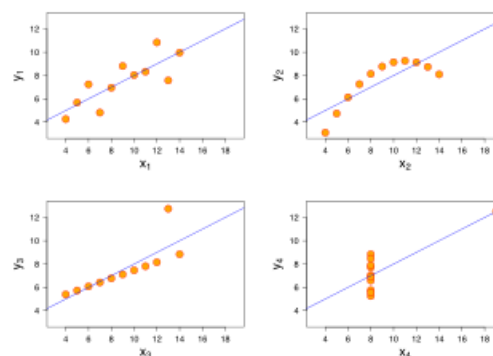
Em geral, o estudo da relação entre as variáveis deve ser acompanhado de gráficos descritivos para a apreensão dos dados à disposição para evitar incorrer os limites estritamente técnicos dos cálculos utilizados.^[52] Entretanto, quando se trata de se concentrar nas ligações entre diversas variáveis, as representações gráficas podem não ser possíveis ou podem não ter os dados dispersos de uma forma que seja imediatamente possível identificar a correlação. Os cálculos podem então ajudar a simplificar as interpretações possíveis sobre as ligações entre as variáveis e/ou indicarem as principais hipóteses necessárias para boas leituras.

Correlação e causalidade

A expressão *correlação não implica causalidade* significa que correlação não pode ser usada para a relação causal entre as variáveis. Por exemplo, a quantidade de queimaduras de sol pode estar fortemente correlacionada ao número de óculos de sol vendidos em uma cidade litorânea, mas nenhum fenômeno é provavelmente a causa do outro. Entretanto, por mais que a expressão não signifique que as correlações não podem indicar a potencial existência de relações causais, as causas subjacentes à correlação (se houver) podem ser indiretas e desconhecidas. Por exemplo, a correlação entre idade e altura em crianças é clara. No entanto, a correlação entre humor e saúde nas pessoas é menos óbvia. A afirmação correta é a melhora do humor leva à melhora da saúde, a boa saúde leva ao bom humor ou ambas? Em outras palavras, uma correlação pode ser tomada como evidência de uma possível relação causal, mas não pode indicar qual é a relação causal.^[53]

Correlação e linearidade

O coeficiente de correlação de Pearson indica a força de uma relação linear entre duas variáveis, mas seu valor geralmente não caracteriza completamente sua relação.^[54] Em particular, se a média condicional de Y dado X , denotado $E(Y|X)$, não for linear em X , o coeficiente de correlação não determinará completamente $E(Y|X)$. A imagem ao lado mostra gráficos de dispersão do quarteto de Anscombe, um conjunto de quatro pares diferentes de variáveis criadas por Francis Anscombe. As quatro variáveis y têm a mesma



Os quatro conjunto de dados possuem coeficiente de correlação de 0,816. No primeiro gráfico, a distribuição parece ser normalmente distribuída e corresponde ao que se esperaria ao considerar duas variáveis correlacionadas e seguindo o pressuposto de normalidade. No segundo gráfico, a distribuição não é normalmente distribuída. Embora uma relação óbvia entre duas variáveis possa ser observada, ela não é linear. Neste caso, o coeficiente de correlação de Pearson não indica que há uma relação funcional exata – apenas a medida em que esta relação pode ser aproximada por uma relação linear. No terceiro gráfico, a relação linear é perfeita com exceção de um *outlier* que exerce influencia suficiente para reduzir o coeficiente de correlação de 1 para 0,816. No quarto gráfico, há outro exemplo de quando um *outlier* é suficiente para produzir um alto coeficiente de correlação, mesmo se a relação entre as duas variáveis não for linear.

média (7,5), variância (4,12), correlação (0,816) e regressão linear($y = 3 + 0,5x$).^[55] Entretanto, a distribuição das variáveis é muito diferente. De acordo com os gráficos ao lado, o coeficiente de correlação como uma estatística de resumo não pode substituir a análise visual dos dados. Nota-se que os exemplos às vezes indicam que o coeficiente de correlação de Pearson presume que os dados seguem uma distribuição normal. No entanto, isto não é verdade.^[56]

Ver também

- Autocorrelação
- Biometria
- Cópula (estatística)
- Correlação espúria
- Covariância
- Econometria
- Estatística
- Lambda de Goodman e Kruskal
- R^2
- Relação cruzada
- Regressão
- Regressão espúria

Referências

1. Bussab, Wilton de O.; Morettin, Pedro A. (2010). *Estatística Básica* 6ª ed. [S.l.]: Saraiva. p. 73. 540 páginas
2. Bussab, Wilton de O.; Morettin, Pedro A. (2010). *Estatística Básica* 6ª ed. [S.l.]: Saraiva. p. 76. 540 páginas
3. Mann, Prem S. (2010). *Introductory Statistics* 7ª ed. [S.l.]: John Wiley & Sons. p. 592 — 593. 625 páginas
4. Cooper, Donald R.; Schindler, Pamela S. (2016). *Métodos de Pesquisa em Administração* 12ª ed. [S.l.]: AMGH. p. 480. 695 páginas
5. Bussab, Wilton de O.; Morettin, Pedro A. (2010). *Estatística Básica* 6ª ed. [S.l.]: Saraiva. p. 83 — 84. 540 páginas
6. Martins, Maria Eugénia Graça (2014). «Coeficiente de Correlação Amostral» (https://www.fc.up.pt/pessoas/jfgomes/pdf/vol_2_num_2_69_art_coeficienteCorrelacaoAmostral.pdf) (PDF). *Revista de Ciência Elementar*. p. 1. Consultado em 19 de maio de 2017
7. Rodgers, J. L.; Nicewander, W. A. (1988). «Thirteen ways to look at the correlation coefficient». *The American Statistician*. **42** (1): 59–66. JSTOR 2685263 (<https://www.jstor.org/stable/2685263>). doi:10.1080/00031305.1988.10475524 (<https://dx.doi.org/10.1080%2F00031305.1988.10475524>)
8. Croxton, Frederick Emory; Cowden, Dudley Johnstone; Klein, Sidney (1968) *Applied General Statistics*, Pitman. ISBN 9780273403159 (page 625)
9. Dietrich, Cornelius Frank (1991) *Uncertainty, Calibration and Probability: The Statistics of Scientific and Industrial Measurement* 2nd Edition, A. Higler. ISBN 9780750300605 (Page 331)
10. Aitken, Alexander Craig (1957) *Statistical Mathematics* 8th Edition. Oliver & Boyd. ISBN 9780050013007 (Page 95)
11. Dodge, Yadolah (2008). *The Concise Encyclopedia of Statistics*. [S.l.]: Springer. p. 115 — 119
12. Memória, José Maria Pompeu (2004). «Breve História da Estatística» (http://www2.ee.ufpe.br/codec/historia_e_statistica.pdf) (PDF): 22 — 25
13. Piovani, PIJuan Ignacio (2013). «De Objeto a Método: Notas Históricas sobre Estatística e Pesquisa Social» (<http://www.scielo.br/pdf/sant/v3n5/2238-3875-sant-03-05-0245.pdf>) (PDF). *Sociologia & Antropologia*. **3** (5): 245 — 270
14. Caire, Elaine (2013). «A História da Curva Normal» (https://repositorio.unesp.br/bitstream/handle/11449/91024/caire_e_me_rcla.pdf?sequence=1) (PDF): 27 — 30
15. Barbe, Lluís (2010). *Francis Ysidro Edgeworth: A Portrait with Family and Friends*. [S.l.]: Edward Elgar. p. 169 — 171. 291 páginas
16. «POPULAÇÃO E AMOSTRA» (http://web.unifoa.edu.br/portal/plano_aula/arquivos/04680/1.%20Nota%20de%20Aula%20-%20POPULA%C3%87%C3%83O%20E%20AMOSTRA.pdf) (PDF). Centro Universitário de Volta Redonda. p. 1. Consultado em 19 de maio de 2017
17. Razdolsky, Leo (2014). *Probability—Based Structural Fire Load*. [S.l.]: Cambridge University Press. p. 63. 335 páginas
18. «Covariance and Correlation» (<http://www.stat.rice.edu/~dcox/Stat421/Supp2/node3.html>). Rice University. Consultado em 1 de junho de 2017
19. Dowdy, S. and Wearden, S. (1983). "Statistics for Research", Wiley. ISBN 0-471-08602-9 pp 230
20. Hirsch, Robert P. (2016). *Introduction to Biostatistical Applications in Health Research with Microsoft Office Excel*. [S.l.]: Wiley. p. 150. 392 páginas
21. 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>)
22. Dogde, Yadolah (2008). *The Concise Encyclopedia of Statistics*. [S.l.]: Springer. p. 88 — 91
23. Rodrigues, Lígia Henriques (18 de maio de 2015). «Regressão Linear Simples» (<https://www.ime.usp.br/~fmachado/MAE229/AULA10.pdf>) (PDF). IMEUSP. p. 7. Consultado em 19 de maio de 2017
24. Gniazdowski, Zenon (2013). «Geometric interpretation of a correlation» (http://zeszyty-naukowe.wusi.edu.pl/zeszyty/zeszyt9/Geometric_interpretation_of_a_correlatio

- n.pdf) (PDF). *Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki*. 7 (9): 27-35. Consultado em 8 de junho de 2017
25. Shalhehi, Ahmad Hedayatpanah (2013). «Using Eta correlation ratio in analyzing strongly nonlinear relationship between two Variables in Practical researches» ([http://webcache.googleusercontent.com/search?q=cache:4ZqwwszJ5_EJ:www.tjmcs.com/include_s/files/articles/Vol7_Iss3_213-220_Using_Eta_\(*\)_correlation_ratio_in_a.pdf+&cd=1&hl=pt-BR&ct=clnk&gl=br](http://webcache.googleusercontent.com/search?q=cache:4ZqwwszJ5_EJ:www.tjmcs.com/include_s/files/articles/Vol7_Iss3_213-220_Using_Eta_(*)_correlation_ratio_in_a.pdf+&cd=1&hl=pt-BR&ct=clnk&gl=br)) (PDF). *Journal of mathematics and computer science*. p. 2. Consultado em 19 de maio de 2017
 26. LEWANDOWSKI, DANIEL; COOKE, ROGER M.; TEBBENS, RADBOUD J. DUINTJER (2007). «Sample-based Estimation of Correlation Ratio with Polynomial Approximation» (<https://pdfs.semanticscholar.org/a520/d60be17ac548b363360074df1bb37d69307e.pdf>) (PDF). Semantic Scholar. Consultado em 14 de junho de 2017
 27. Pontes, Antonio Carlos Fonseca. «ENSINO DA CORRELAÇÃO DE POSTOS NO ENSINO MÉDIO» (<http://www.ime.unicamp.br/sinape/sites/default/files/EnsinoCorrelacaoDePostos.pdf>) (PDF). Universidade Federal do Acre. Consultado em 19 de maio de 2017
 28. Falcão, António Jorge Teixeira (2012). «Detecção de Correlação e Causalidade em Séries Temporais não Categóricas» (https://run.unl.pt/bitstream/10362/7814/1/Falcao_2012.pdf) (PDF). Universidade nova de Lisboa. Consultado em 19 de maio de 2017
 29. 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
 30. Kendall, M. G. (1955) "Rank Correlation Methods", Charles Griffin & Co.
 31. Devore, Jay L. (2006). *Probabilidade e Estatística para Engenharia e Ciências*. [S.l.]: Cengage Learning. p. 432 — 433. 692 páginas
 32. Filho, Luiz Medeiros de Araujo Lima. «Correlação e Regressão» (<http://www.de.ufpb.br/~luiz/AED/Aula9.pdf>) (PDF). UNIVERSIDADE FEDERAL DA PARAÍBA. p. 6. Consultado em 19 de maio de 2017
 33. Viali, Lorí. «Estatística Básica» (http://www.pucrs.br/famat/viali/graduacao/engenharias/material/apostilas/Apostila_5.pdf) (PDF). Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS). Consultado em 7 de junho de 2017
 34. Filho, Kepler de Souza Oliveira (10 de abril de 2013). «Mínimos Quadrados» (<http://astro.if.ufrgs.br/minq/>). Universidade Federal do Rio Grande do Sul. p. 1. Consultado em 22 de maio de 2017
 35. Farber, Larson (2010). *Estatística Aplicada*. São Paulo: Pearson. 396 páginas. ISBN 978-85-7605-372-9
 36. Farber, Larson (2010). *Estatística Aplicada*. São Paulo: Pearson. 396 páginas. ISBN 9788576053729
 37. Mahdavi Damghani B. (2013). «The Non-Misleading Value of Inferred Correlation: An Introduction to the Cointelation Model». *Wilmott Magazine*. doi:10.1002/wilm.10252 (<https://dx.doi.org/10.1002%2Fwilm.10252>)
 38. Abbara, Omar M. F. (2009). «Modelagem de Dependência em Séries Financeiras Multivariadas» (http://www.ime.unicamp.br/sinape/sites/default/files/Abbara,OmarMuheddineFranco_M.pdf) (PDF). Campinas: Instituto de Matemática, Estatística e Computação Científica - Unicamp. p. 18. Consultado em 14 de junho de 2017
 39. Székely, G. J. Rizzo; Bakirov, N. K. (2007). «Measuring and testing independence by correlation of distances». *Annals of Statistics*. 35 (6): 2769–2794. doi:10.1214/009053607000000505 (<https://dx.doi.org/10.1214%2F009053607000000505>)
 40. Székely, G. J.; Rizzo, M. L. (2009). «Brownian distance covariance». *Annals of Applied Statistics*. 3 (4): 1233–1303. doi:10.1214/09-AOAS312 (<https://dx.doi.org/10.1214%2F09-AOAS312>)
 41. 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>)
 42. Plischke, Elmar (2010). «An Adaptive Correlation Ratio Method» (http://ac.els-cdn.com/S1877042810013388/1-s2.0-S1877042810013388-main.pdf?_tid=81854404-4a2d-11e7-a1ec-00000aaccb361&acdnat=1496694676_e4c32c794391cd720dbc925fac00e301) (PDF). *Procedia Social and Behavioral Sciences*: 7722 – 7723
 43. Cao, Longbing (2015). «Coupling learning of complex interactions» (<https://web.archive.org/web/20170705130854/http://www-staff.it.uts.edu.au/%7Elbcao/publication/JIPM-online.pdf>) (PDF). *Information Processing and Management*: 167 – 186. Arquivado do original (<https://www-staff.it.uts.edu.au/~lbcao/publication/JIPM-online.pdf>) (PDF) em 5 de julho de 2017
 44. «19. Simple Linear Regression IV — The Coefficient of Determination R^2 » (<http://pages.stern.nyu.edu/~churvic/MBA/Handouts/19-Reg4.pdf>) (PDF). Leonard N. Stern School of Business — New York University. Consultado em 7 de junho de 2017
 45. Komenda, Stanislav (1993). «"Dressing up" with correlation» (<http://www.statapol.cz/oldstat/bulletiny/lib-93-a.pdf>) (PDF). *Information Bulletin of Czech Statistical Society*. 4. Consultado em 8 de junho de 2017
 46. Thorndike, Robert Ladd (1947). *Research problems and techniques (Report No. 3)*. Washington DC: US Govt. print. off.
 47. Nikolić, D; Muresan, RC; Feng, W; Singer, W (2012). «Scaled correlation analysis: a better way to compute a cross-correlogram». *European Journal of Neuroscience*: 1–21. doi:10.1111/j.1460-9568.2011.07987.x (<https://dx.doi.org/10.1111%2Fj.1460-9568.2011.07987.x>)
 48. «Chapter 14: Describing Relationships: Scatterplots and Correlation» (<http://math.wayne.edu/~bert/courses/1020/Handouts/chap14.pdf>) (PDF). Wayne State University. Consultado em 7 de junho de 2017
 49. Jacob Cohen. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.)
 50. Agarwal, B.L. (2006). *Basics Statistics* (<https://books.google.com.br/books?id=qT2srNLLxB0C&pg=PA396&lpg=PA396&dq=coefficient+of+correlation+is+always+possible+to+be+calculate?&source=bl&ots=W3SPGCOY0j&sig=6WDOtwLYvq-UH0iTRBd1n7MQbx0&hl=pt-BR&sa=X&ved=0ahUKEwi-1pKhtb7UAhWRqZAKHcCZCEM4ChDoAQhVMAY#v=onepage&q=coefficient%20of%20correlation%20is%20always%20possible%20to%20be%20calculate%3F&f=false>) 4 ed. Delhi: New Age International Publishers. p. 396. ISBN 81-224-1814-7
 51. Garlick, Andy. *Estimating Risk: A Management Approach*. [S.l.]: Gower. p. 187. ISBN 9780566087769
 52. Paiva, Delhi; Lauretto, Marcelo de Souza (2011). «Análise Exploratória de Dados» (http://www.each.usp.br/lauretto/SIN5008_2011/aula01/aula1). SIN5008 -

- Estatística computacional. Consultado em 14 de junho de 2017
53. Aldrich, John (1995). «Correlations Genuine and Spurious in Pearson and Yule». *Statistical Science*. **10** (4): 364–376. JSTOR 2246135 (<https://www.jstor.org/stable/2246135>). doi:10.1214/ss/1177009870 (<https://dx.doi.org/10.1214/ss/1177009870>)
54. Mahdavi Damghani, Babak (2012). «The Misleading Value of Measured Correlation». *Wilmott*. **2012** (1): 64–73. doi:10.1002/wilm.10167 (<https://dx.doi.org/10.1002/wilm.10167>)
55. Anscombe, Francis J. (1973). «Graphs in statistical analysis». *The American Statistician*. **27**: 17–21. JSTOR 2682899 (<https://www.jstor.org/stable/2682899>). doi:10.2307/2682899 (<https://dx.doi.org/10.2307/2682899>)
56. **Erro de citação: Código <ref> inválido; não foi fornecido texto para as refs de nome thirteenways2**

Leitura adicional

- COHEN, Jacob et al. **Applied multiple regression/correlation analysis for the behavioral sciences**. Routledge, 2013.
- ZANIBONI, GUSTAVO TADEU et al. Fusao bayesiana de imagens utilizando coeficientes de correlação localmente adaptáveis. **Anais do IX SBSR**. Santos, SP, 1998.
-

Ligações externas

- [Correlation Comparison](http://peaks.informatik.uni-erlangen.de/cgi-bin/usignificance.cgi) (<http://peaks.informatik.uni-erlangen.de/cgi-bin/usignificance.cgi>)
- [Finding Order in Randomness: Correlation](http://nagysandor.eu/AsimovTeka/correlation_en/index.html) (http://nagysandor.eu/AsimovTeka/correlation_en/index.html)
- [Interpretação do Coeficiente de Correlação](http://leg.ufpr.br/~silvia/CE003/node74.html) (<http://leg.ufpr.br/~silvia/CE003/node74.html>)
- [Interpreting Correlations](http://rpsychologist.com/d3/correlation/) (<http://rpsychologist.com/d3/correlation/>)
- [Pearson's Correlation Coefficient](http://www.vias.org/tmdatanaleng/cc_corr_coeff.html) (http://www.vias.org/tmdatanaleng/cc_corr_coeff.html)
- [Understanding Correlation](http://www.mega.nu:8080/ampp/rummel/uc.htm) (<http://www.mega.nu:8080/ampp/rummel/uc.htm>)
- [Weighted Correlation Matrix](http://www.mathworks.com/matlabcentral/fileexchange/20846-weighted-correlation-matrix?requestedDomain=www.mathworks.com) (<http://www.mathworks.com/matlabcentral/fileexchange/20846-weighted-correlation-matrix?requestedDomain=www.mathworks.com>)
- [Correlation \(in statistics\)](http://www.encyclopediaofmath.org/index.php?title=Correlation_(in_statistics)&oldid=11629). ([http://www.encyclopediaofmath.org/index.php?title=Correlation_\(in_statistics\)&oldid=11629](http://www.encyclopediaofmath.org/index.php?title=Correlation_(in_statistics)&oldid=11629)) A.V. Prokhorov (originator), *Encyclopedia of Mathematics*.
- Weisstein, Eric W. [Correlation Coefficient](http://mathworld.wolfram.com/CorrelationCoefficient.html). (<http://mathworld.wolfram.com/CorrelationCoefficient.html>) From *MathWorld*--A Wolfram Web Resource.

Obtida de "<https://pt.wikipedia.org/w/index.php?title=Correlação&oldid=56939697>"

Esta página foi editada pela última vez às 15h21min de 11 de dezembro de 2019.

Este texto é disponibilizado nos termos da licença Atribuição-CompartilhaIgual 3.0 Não Adaptada (CC BY-SA 3.0) da Creative Commons; pode estar sujeito a condições adicionais. Para mais detalhes, consulte as [condições de utilização](#).