

Correlação de Pearson: Uma Análise Estatística de Relações Lineares

Júlio Henrique da Costa Santana

Faculdade de Educação Tecnológica do Estado do Rio de Janeiro – Paracambi

1 – Resumo.

Este artigo explora o conceito e a aplicação da correlação de Pearson, uma medida estatística amplamente utilizada para avaliar a relação linear entre duas variáveis. Começamos por definir o coeficiente de correlação de Pearson, suas propriedades e como ele é calculado. Em seguida, apresentamos exemplos de situações em que a correlação de Pearson é aplicável e discutimos suas limitações. Por fim, discutimos a interpretação dos resultados da correlação de Pearson e sua importância nas ciências sociais, econômicas e naturais.

1.1 – Abstract

This article explores the concept and application of Pearson correlation, a widely used statistical measure to assess the linear relationship between two variables. We start by defining the Pearson correlation coefficient, its properties and how it is calculated. Next, we present examples of situations in which the Pearson correlation is applicable and discuss its limitations. Finally, we discuss the interpretation of Pearson correlation results and their importance in social, economic and natural sciences.

2 – Introdução.

A correlação é uma técnica estatística fundamental usada para avaliar a força e a direção da relação entre duas variáveis. Entre as várias medidas de correlação, o coeficiente de correlação de Pearson, ou simplesmente "correlação de Pearson", é amplamente empregado em pesquisa científica. Esta medida é particularmente útil quando se deseja avaliar a relação linear entre duas variáveis quantitativas. Ela foi desenvolvida por Karl Pearson no início do século XX e tem sido um dos métodos mais comuns na análise estatística. A literatura existente abrange vários aspectos da correlação de Pearson, incluindo sua definição, cálculo, interpretação e aplicações em diversas áreas do conhecimento.

3 – Método.

O coeficiente de correlação de Pearson (r) é calculado pela fórmula:

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$

onde x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n são os valores medidos de ambas as variáveis. Para além disso:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad \text{e} \quad \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i$$

são as médias aritméticas de ambas as variáveis.

Pearson, K. (1900). Na *Philosophical Magazine*, Series 5, 50(302), 157-175 fala que a correlação de Pearson é definida como um coeficiente que varia de -1 a 1, indicando a força e a direção de uma relação linear entre duas variáveis. Ela é calculada usando a fórmula mencionada anteriormente. A literatura inicial de Pearson, como seu trabalho de 1900, estabeleceu os fundamentos dessa medida.

De acordo com Dancey, C. P., & Reidy, J. (2007). A interpretação do coeficiente de correlação de Pearson é essencial. Um valor de r próximo a 1 indica uma forte correlação positiva, enquanto um valor próximo a -1 indica uma forte correlação negativa. Um valor próximo a 0 sugere uma correlação fraca ou ausente.

Embora amplamente utilizada, Barbiero, A., & Vettorato, D. (2008) fala que a correlação de Pearson tem algumas limitações. Ela assume uma relação linear, é sensível a valores atípicos e não pode detectar relações não lineares. Outras medidas, como a correlação de Spearman e a correlação de Kendall, podem ser mais adequadas em certos casos.

Em comparação com outras a correlação de Pearson, Zar, J. H. (1999) disse que outras medidas de correlação, como a correlação de Spearman, a correlação de Kendall e a correlação de point-biserial. Cada uma delas tem suas vantagens e desvantagens, e a escolha da medida depende da natureza dos dados e do objetivo da análise.

4 - Resultados e Discussão.

A correlação de Pearson varia de -1 a 1, onde -1 indica uma correlação negativa perfeita, 0 indica ausência de correlação e 1 indica uma correlação positiva perfeita. Um valor de r próximo a 0 sugere uma correlação fraca, enquanto valores próximos a -1 ou 1 indicam uma correlação forte. A correlação de Pearson é amplamente utilizada em diversas áreas, incluindo psicologia, economia, medicina e ciências naturais. Ela fornece uma medida quantitativa da relação linear entre duas variáveis, permitindo a identificação de padrões e tendências. Cohen, J. (1988) alega que correlação de Pearson é aplicada em diversas áreas, incluindo psicologia, economia, medicina, ecologia e ciências sociais. Ela é usada para analisar relações entre variáveis como idade e desempenho acadêmico, consumo de produtos e vendas, entre outras. No entanto, é importante lembrar que a

correlação não implica causalidade. Além disso, a correlação de Pearson é sensível a valores atípicos, o que pode distorcer os resultados.

5 - Conclusão:

A correlação de Pearson é uma ferramenta valiosa na análise estatística, permitindo a quantificação da relação linear entre variáveis. Ao entender seus fundamentos e limitações, os pesquisadores podem aplicar essa medida de forma eficaz em suas investigações e interpretar os resultados com confiança.

6 – Referências.

Pearson, K. (1900). *Philosophical Magazine*, Series 5, 50(302), 157-175.

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. Prentice Hall.

Barbiero, A., & Vettorato, D. (2008). When is the Pearson coefficient not a suitable tool for examining relationships between variables? *Proceedings of the National Academy of Sciences of the United States of America*, 105(4), E6.

Zar, J. H. (1999). *Biostatistical analysis*. Prentice Hall.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.