

Coeficiente de correlação de Pearson

Em estatística descritiva, o **coeficiente de correlação de Pearson**, também chamado de "coeficiente de correlação produto-momento" ou simplesmente de " ρ de Pearson" mede o grau da correlação (e a direcção dessa correlação - se positiva ou negativa) entre duas variáveis de escala métrica (intervalar ou de rácio/razão).

Este coeficiente, normalmente representado por ρ assume apenas valores entre -1 e 1.

- $\rho = 1$ Significa uma correlação perfeita positiva entre as duas variáveis.
- $\rho = -1$ Significa uma correlação negativa perfeita entre as duas variáveis - Isto é, se uma aumenta, a outra sempre diminui.
- $\rho = 0$ Significa que as duas variáveis não dependem linearmente uma da outra.

No entanto, pode existir uma dependência não linear. Assim, o resultado $\rho = 0$ deve ser investigado por outros meios.

Cálculo

Calcula-se o coeficiente de correlação de Pearson segundo a seguinte 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,

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

são as médias aritméticas de ambas as variáveis.

A análise correlacional indica a relação entre 2 variáveis lineares e os valores sempre serão entre +1 e -1. O sinal indica a direcção, se a correlação é positiva ou negativa, e o tamanho da variável indica a força da correlação.

Cabe observar que, como o coeficiente é concebido a partir do ajuste linear, então a fórmula não contém informações do ajuste, ou seja, é composta apenas dos dados.

Interpretando ρ

- 0.9 para mais ou para menos indica uma correlação muito forte.
- 0.7 a 0.9 positivo ou negativo indica uma correlação forte.
- 0.5 a 0.7 positivo ou negativo indica uma correlação moderada.
- 0.3 a 0.5 positivo ou negativo indica uma correlação fraca.
- 0 a 0.3 positivo ou negativo indica uma correlação desprezível.