

CORRELAÇÃO LINEAR

Em Estatística, a **correlação** é um parâmetro que indica o “grau de correspondência” entre duas variáveis. Ou seja, a correlação mostra a “intensidade” com a qual dois conjuntos de dados estão relacionados mutuamente.

A medida para o grau de correlação linear entre duas variáveis é o coeficiente de correlação de Pearson (conhecido como coeficiente de correlação linear), indicado por “r”, calculado por:

$$r = \frac{n \cdot \sum x_i \cdot y_i - (\sum x_i) \cdot (\sum y_i)}{\sqrt{(n \cdot \sum x_i^2 - (\sum x_i)^2) \cdot (n \cdot \sum y_i^2 - (\sum y_i)^2)}}$$

Os possíveis valores de correlação linear variam de -1 a 1, ou seja: $-1 \leq r \leq 1$.

Podemos proceder à seguinte classificação:

- $r = -1,00$: correlação negativa perfeita.
- $r = -0,75$: correlação negativa forte.
- $r = -0,50$: correlação negativa média.
- $r = -0,25$: correlação negativa fraca.
- $r = 0,00$: correlação linear inexistente.
- $r = +0,25$: correlação positiva fraca.
- $r = +0,50$: correlação positiva média.
- $r = +0,75$: correlação positiva forte.
- $r = +1,00$: correlação positiva perfeita.

Exemplo : Um pesquisador indagou a 7 pessoas (todas com 40 anos de idade) que aguardavam o trem em uma plataforma de metrô as seguintes questões:

- ✓ Qual a sua escolaridade, ou seja, quantos anos você estudou?
- ✓ Quantos livros você já leu?

Tabela : Número de anos que a pessoa estudou (x_i) e número de livros que a pessoa já leu (y_i).

x_i	3	5	7	9	10	14	16
y_i	1	2	3	5	7	10	13

x_i	y_i	x_i^2	y_i^2	$x_i \cdot y_i$
3	1	$(3)^2 = 9$	$(1)^2 = 1$	$(3) \cdot (1) = 3$
5	2	$(5)^2 = 25$	$(2)^2 = 4$	$(5) \cdot (2) = 10$
7	3	$(7)^2 = 49$	$(3)^2 = 9$	$(7) \cdot (3) = 21$
9	5	$(9)^2 = 81$	$(5)^2 = 25$	$(9) \cdot (5) = 45$
10	7	$(10)^2 = 100$	$(7)^2 = 49$	$(10) \cdot (7) = 70$
14	10	$(14)^2 = 196$	$(10)^2 = 100$	$(14) \cdot (10) = 140$
16	13	$(16)^2 = 256$	$(13)^2 = 169$	$(16) \cdot (13) = 208$
$\sum x_i = 64$	$\sum y_i = 41$	$\sum x_i^2 = 716$	$\sum y_i^2 = 357$	$\sum x_i \cdot y_i = 497$

O coeficiente de correlação linear é:

$$r = \frac{n \cdot \sum x_i \cdot y_i - (\sum x_i) \cdot (\sum y_i)}{\sqrt{(n \cdot \sum x_i^2 - (\sum x_i)^2) \cdot (n \cdot \sum y_i^2 - (\sum y_i)^2)}}$$

Para o exemplo 2, temos:

$$r = \frac{7 \cdot 497 - (64) \cdot (41)}{\sqrt{(7 \cdot 716 - (64)^2) \cdot (7 \cdot 357 - (41)^2)}} = \frac{3479 - 2624}{\sqrt{916 \cdot 818}} = 0,988$$