# 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:

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

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 |

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

O coeficiente de correlação linear é:

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

Para o exemplo 2, temos:

$$r = \frac{7.497 - (64).(41)}{\sqrt{(7.716 - (64)^2).(7.357 - (41)^2)}} = \frac{3479 - 2624}{\sqrt{916.818}} = 0,988$$

**Regressão Linear**

A regressão é o processo de traduzir o comportamento conjunto de duas variáveis na forma de uma "lei" matemática, denominada "equação da regressão".

Como, na prática, trabalhamos com diversos pontos experimentais, há inúmeras retas possíveis para o referido conjunto de dados. No entanto, há um critério, conhecido como "Método dos Mínimos Quadrados", que estabelece a "melhor" reta que se ajusta a todos os pontos experimentais

$a = r.(s_y/s_x)$

A equação geral de regressão ("reta interpoladora") é:

$$y = a.x + (\bar{y} - a\,\bar{x})$$

Onde:

$$\bar{x} = \frac{\sum x_i}{n}$$

$$s_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = \frac{\sum y_i}{n}$$

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

EXERCÍCIOS

**1.** A tabela a seguir mostra os resultados de uma pesquisa que apreciou o peso de um veículo (em ton) e o número médio de peças defeituosas que tiveram de ser repostas no primeiro ano de uso do automóvel. Pedem-se:

o coeficiente de correlação linear e a equação da regressão linear;

| $P_i$: Peso do veículo (ton) | $N_i$: Número de peças defeituosas |
|---|---|
| 1,00 | 2 |
| 1,25 | 5 |
| 1,50 | 5 |
| 1,75 | 7 |
| 2,00 | 10 |
| 2,25 | 11 |
| 2,50 | 15 |

Respostas: 0,98 e $N^* = 8{,}00.P_i - 6{,}14$

**2.** A tabela a seguir mostra os resultados de uma pesquisa realizada durante o mês de julho em um hospital pediátrico na qual foram apreciados: temperatura média do dia e número de atendimentos de casos de problemas respiratórios. Pedem-se:

o coeficiente de correlação linear e a equação da regressão linear;

| $T_i$: Temperatura média (°C) | $N_i$: Casos de problemas respiratórios |
|---|---|
| 9 | 28 |
| 11 | 26 |
| 14 | 22 |
| 15 | 22 |
| 17 | 22 |
| 18 | 16 |
| 20 | 12 |
| 21 | 6 |
| 22 | 6 |

Respostas: -0,94 e $N^* = -1{,}73.T_i + 46{,}09$