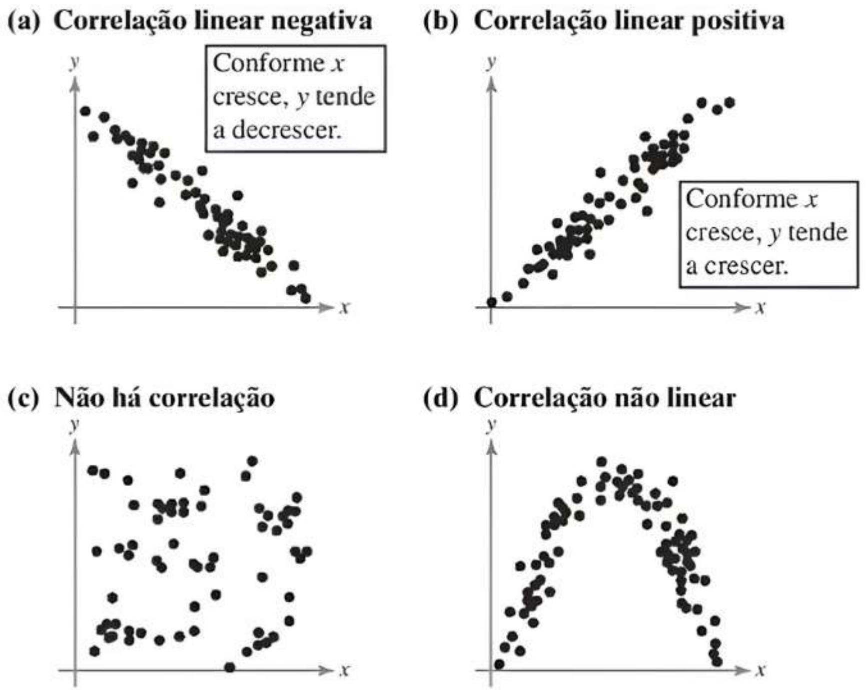
Estatística Inferencial

Regressão Linear

Correlação Linear

- Uma correlação refere-se à relação entre duas variáveis, onde os dados podem ser expressos como pares ordenados (x, y).
- x representa a variável independente, e y é a variável dependente.
- Se essa relação puder ser expressa por meio de função linear, temos uma correlação linear.

Correlação Linear



Fonte: Larson e Farber (2015, p. 439).

$$r = \frac{n \cdot \sum xy - (\sum x)(\sum y)}{\sqrt{\left(n \cdot \sum x^2 - (\sum x)^2\right)\left(n \cdot \sum y^2 - (\sum y)^2\right)}}$$

Exemplo

- Em uma competição de corrida, foram registrados dados sobre a idade e o peso dos participantes por meio de uma amostragem aleatória simples, resultando em uma amostra composta por 4 mulheres e 10 homens. Os dados das mulheres foram organizados na tabela. Como podemos classificar a correlação entre a idade (x) e o peso (y) das mulheres apresentados?

Mulheres	
Idade	Peso
18	60
20	62
23	65
27	58

Fonte: elaborado pela autora.

$$r = \frac{n \cdot \sum xy - (\sum x)(\sum y)}{\sqrt{\left(n \cdot \sum x^2 - (\sum x)^2\right)\left(n \cdot \sum y^2 - (\sum y)^2\right)}}$$

x	y	x^2	y^2	$x \cdot y$
18	60	324	3600	1080
20	62	400	3844	1240
23	65	529	4225	1495
27	58	729	3364	1566
$\sum x = 88$	$\sum y = 245$	$\sum x^2 = 1982$	$\sum y^2 = 15033$	$\sum xy = 5381$

Fonte: elaborado pela autora.

$\sum x = 88 \quad \sum y = 245 \quad \sum x^2 = 1982 \quad \sum y^2 = 15033 \quad \sum xy = 5381$

$$r = \frac{n \cdot \sum xy - (\sum x)(\sum y)}{\sqrt{\left(n \cdot \sum x^2 - (\sum x)^2\right)\left(n \cdot \sum y^2 - (\sum y)^2\right)}}$$

$$r = \frac{4 \cdot 5381 - 88 \cdot 245}{\sqrt{(4 \cdot 1982 - (88)^2)(4 \cdot 15033 - (245)^2)}} = -\frac{36}{\sqrt{19688}}$$
$$\approx -0{,}26$$

Coeficiente de Determinação

- Refere-se ao quadrado do coeficiente de correlação.
- Expressa a proporção da variação da variável independente que é explicada pela variável dependente, servindo como uma métrica para avaliar a qualidade do ajuste.

$$r^2 = \frac{\text{variação explicada}}{\text{variação total}}$$

Regressão Linear

- Expressamos a variação da variável dependente com base nas variáveis independentes.
- Reta da regressão

$$y = ax + b$$

- Coeficiente angular

$$a = \frac{n \cdot \sum xy - (\sum x)(\sum y)}{n \cdot \sum x^2 - (\sum x)^2}$$

- Coeficiente linear

$$b = \frac{(\sum x^2)(\sum y) - (\sum xy)(\sum x)}{n \cdot \sum x^2 - (\sum x)^2}$$

Exemplo

- Imagine que você é um gerente de uma loja de eletrônicos e está interessado em prever as vendas mensais com base em fatores específicos. Para simplificar, suponha que você tenha dados históricos de vendas mensais (em unidades) e o gasto mensal em publicidade (em reais) para os últimos 6 meses.

Gasto com publicidade (multiplicar por R$1000,00)	10	11	12,2	13,8	14,4	15,5
Unidades vendidas (multiplicar por 10000)	9,8	9,7	12,6	14,4	13,6	16,2

Fonte: elaborado pela autora.

Exemplo

- Com base nesses dados, é possível que você preveja as vendas mensais futuras com base no gasto com publicidade?
- Se você gastar R$18000,00 em publicidade qual será a previsão de vendas?

x	y	x^2	y^2	$x \cdot y$
10	9,8	100	96,04	98
11	9,7	121	94,09	106,7
12,2	12,6	148,84	158,76	153,72
13,8	14,4	190,44	207,36	198,72
14,4	13,6	207,36	184,96	195,84
15,5	16,2	240,25	262,44	251,1
$\boldsymbol{\sum x = 76,9}$	$\boldsymbol{\sum y = 76,3}$	$\boldsymbol{\sum x^2 = 1007,89}$	$\boldsymbol{\sum y^2 = 1003,59}$	$\boldsymbol{\sum xy = 1004,08}$

Fonte: elaborado pela autora.

$\sum x = 76,9$	$\sum y = 76,3$	$\sum x^2 = 1007,89$	$\sum y^2 = 1003,59$	$\sum xy = 1004,08$

$$r = \frac{n \cdot \sum xy - (\sum x)(\sum y)}{\sqrt{\left(n \cdot \sum x^2 - (\sum x)^2\right)\left(n \cdot \sum y^2 - (\sum y)^2\right)}}$$

$$r = \frac{6 \cdot 1004{,}08 - 76{,}9 \cdot 76{,}3}{\sqrt{(6 \cdot 1007{,}89 - (76{,}9)^2)(6 \cdot 1003{,}59 - (76{,}3)^2)}}$$

$$= \frac{157{,}01}{\sqrt{26725{,}9405}} = 0{,}96$$

$\sum x = 76,9$	$\sum y = 76,3$	$\sum x^2 = 1007,89$	$\sum y^2 = 1003,59$	$\sum xy = 1004,08$

O coeficiente angular da reta é dado por:

$$a = \frac{n \cdot \sum xy - (\sum x)(\sum y)}{n \cdot \sum x^2 - (\sum x)^2}$$

$$a = \frac{6 \cdot 1004{,}08 - 76{,}9 \cdot 76{,}3}{6 \cdot 1007{,}89 - (76{,}9)^2} = \frac{157{,}01}{133{,}73} \approx 1{,}17$$

$\sum x = 76,9$	$\sum y = 76,3$	$\sum x^2 = 1007,89$	$\sum y^2 = 1003,59$	$\sum xy = 1004,08$

O coeficiente linear da reta é dado por:

$$b = \frac{(\sum x^2)(\sum y) - (\sum xy)(\sum x)}{n \cdot \sum x^2 - (\sum x)^2}$$

$$b = \frac{1007,89 \cdot 76,3 - 1004,08 \cdot 76,9}{6 \cdot 1007,89 - (76,9)^2} = \frac{-311,745}{133,73} \approx -2,33$$

- A reta de regressão correspondente é $y = 1{,}17x - 2{,}33$.
- Realizando a previsão se for gasto R$18000,00

$$y = 1{,}17x - 2{,}33$$

$$y = 1{,}17 \cdot 18 - 2{,}33$$

$$y = 21{,}06 - 2{,}33 = 18{,}73$$

- Se forem gastos R$18000,00 em publicidade há uma previsão de que sejam vendidas 187300 unidades