Correlação Linear

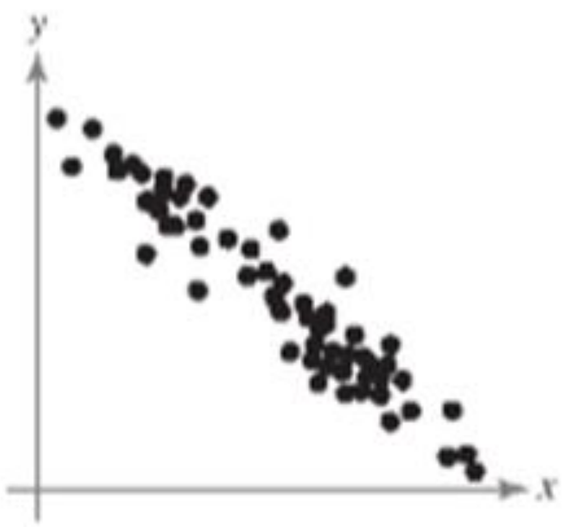
Prof. Luciano Galdino

CORRELAÇÃO LINEAR

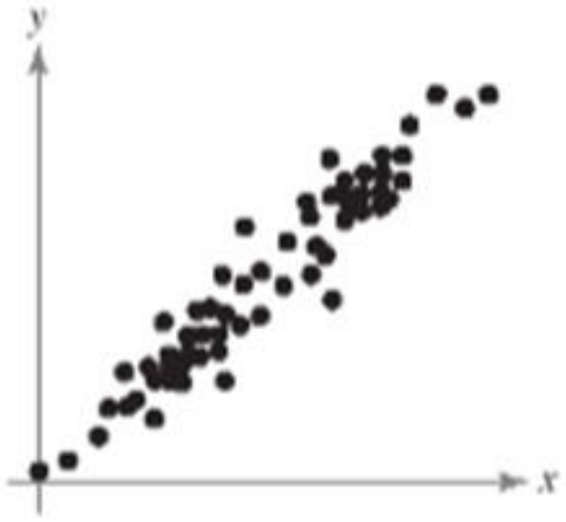
Relação linear entre duas variáveis.
Determinado através de gráficos de dispersão e do coeficiente de correlação.

Observação: Correlação não é causalidade.

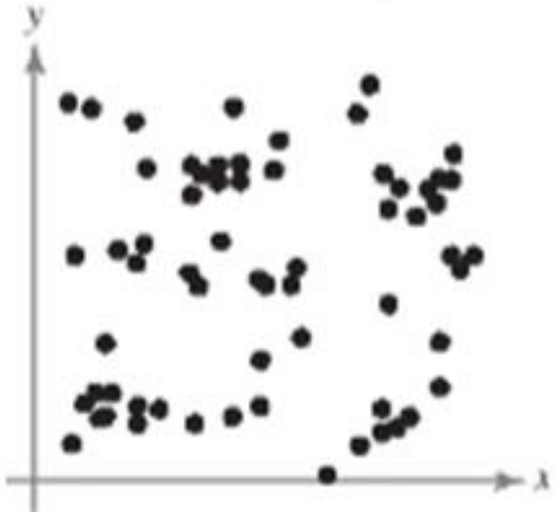
Correlação linear negativa



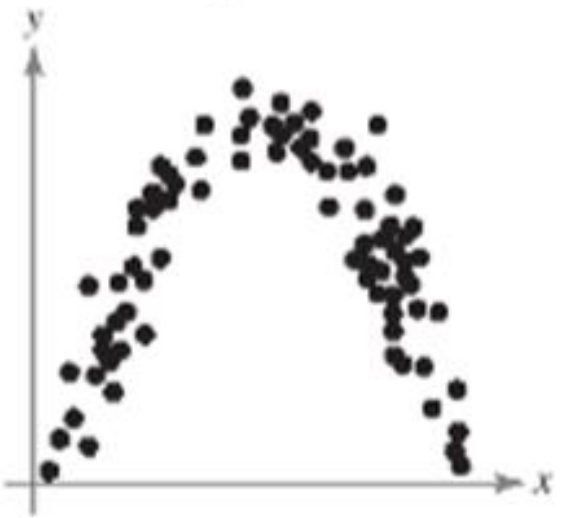
Correlação linear positiva



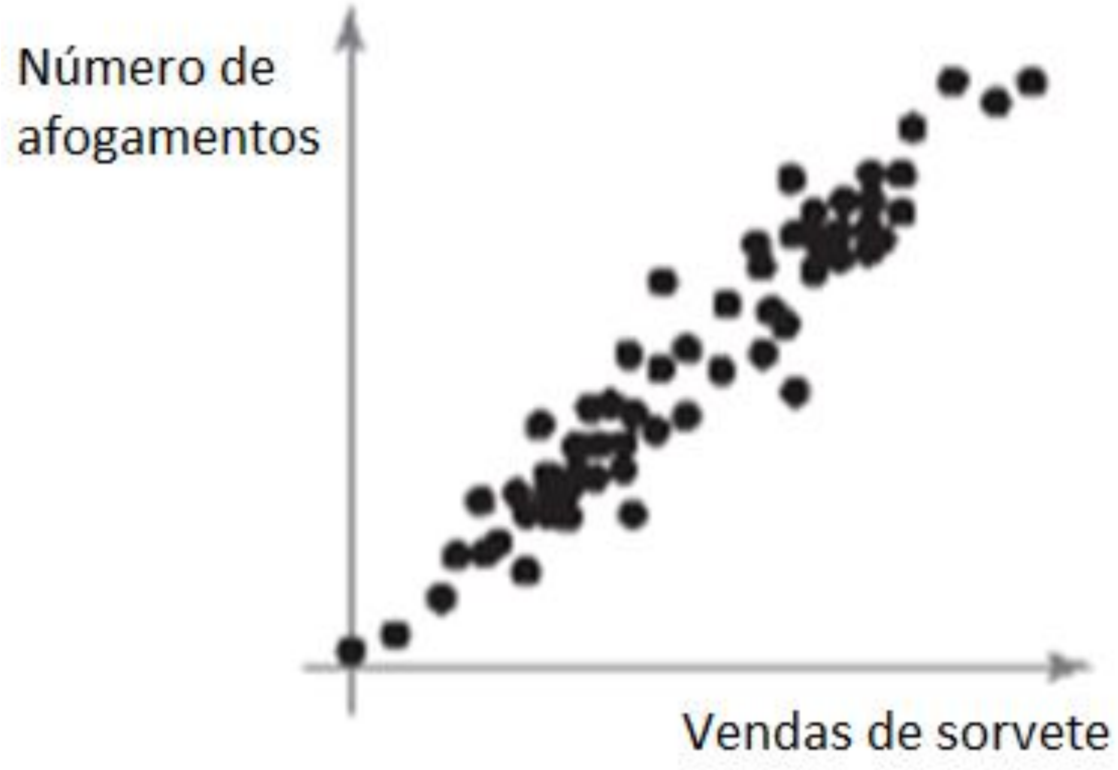
Não há correlação



Correlação não linear



CORRELAÇÃO x CAUSALIDADE



Tem correlação, mas a variação de uma variável não é a causa da variação da outra variável.

Coeficiente de Correlação Linear (Coeficiente de Pearson)

- Forma mais precisa de medir a correlação entre duas grandezas.
- Teste Paramétrico (Normalidade).

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

Coeficiente de correlação (r)	Correlação Positiva	Coeficiente de correlação (r)	Correlação Negativa
r = 1	Perfeita	r = - 1	Perfeita
0,95 ≤ r < 1	Muito forte	- 1 ≤ r < -0,95	Muito forte
0,8 ≤ r < 0,95	Forte	-0,95 ≤ r < -0,8	Forte
0,5 ≤ r < 0,8	Moderada	-0,8 ≤ r < -0,5	Moderada
0 ≤ r < 0,5	Fraca	-0,5 ≤ r < 0	Fraca

Coeficiente de determinação (r^2)

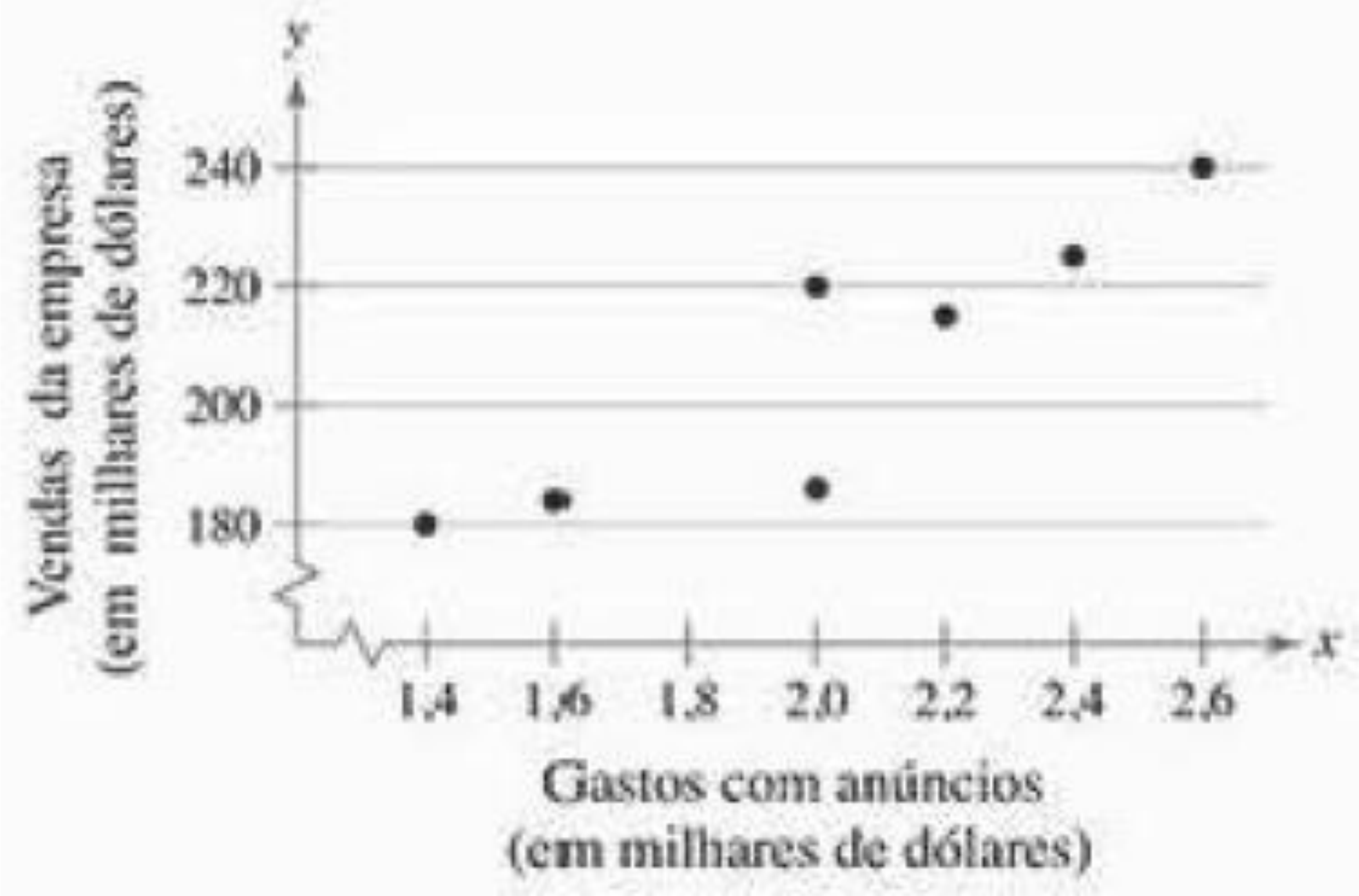
Porcentagem da variação de y que pode ser explicada pela relação de x e y.

Avalia a qualidade do ajuste de um modelo de regressão.

$$r^2 = coeficiente\ de\ correlação\ ao\ quadrado$$

Exemplo: Analisar a correlação linear entre os gastos de propaganda e as vendas de uma empresa com nível de significância de 0,05. Considere dados normalmente distribuídos.

Gastos com propaganda (1.000s de $), x	Vendas da empresa (1.000s de $), y
2,4	225
1,6	184
2,0	220
2,6	240
1,4	180
1,6	184
2,0	186
2,2	215



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

Gastos com propaganda (1.000s de $), x	Vendas da empresa (1.000s de $), y	xy	x^2	y^2
2,4	225	540	5,76	50.625
1,6	184	294,4	2,56	33.856
2,0	220	440	4	48.400
2,6	240	624	6,76	57.600
1,4	180	252	1,96	32.400
1,6	184	294,4	2,56	33.856
2,0	186	372	4	34.596
2,2	215	473	4,84	46.225
$\Sigma x = 15{,}8$	$\Sigma y = 1.634$	$\Sigma xy = 3.289{,}8$	$\Sigma x^2 = 32{,}44$	$\Sigma y^2 = 337.558$

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

$$r = \frac{8(3289{,}8) - (15{,}8)(1634)}{\sqrt{8(32{,}44) - (15{,}8)^2}\,\sqrt{8(337558) - (1634)^2}}$$

$$r = 0{,}9129\ (coeficiente\ de\ correlação)$$

$$r^2 = 0{,}8334\ coeficiente\ de\ determinação$$

n	$\alpha = 0{,}05$	$\alpha = 0{,}01$
4	0,950	0,990
5	0,878	0,959
6	0,811	0,917
7	0,754	0,875
8	0,707	0,834
9	0,666	0,798
10	0,632	0,765
11	0,602	0,735
12	0,576	0,708
13	0,553	0,684
14	0,532	0,661
19	0,456	0,575
20	0,444	0,561
21	0,433	0,549
22	0,423	0,537
23	0,413	0,526
24	0,404	0,515
25	0,396	0,505
26	0,388	0,496
27	0,381	0,487
28	0,374	0,479
29	0,367	0,471

Validação do coeficiente de correlação

- **Se |r| > valor da tabela:** coeficiente significante.

- **Se |r| ≤ valor da tabela:** coeficiente não significante.

Conclusão: Como 0,9129 > 0,707, então o coeficiente é significante.

Coeficiente de correlação de postos de Spearman

Teste Não paramétrico.

Medida da força da relação entre duas variáveis.

Pode ser utilizado na relação de dados lineares e também não lineares, assim como também para dados no nível ordinal.

Cálculo do Coeficiente de Spearman

$$r_R = 1 - \frac{6\Sigma_i {d_i}^2}{n(n^2 - 1)}$$

n = número amostras.
di = diferença de alcance de cada elemento.

Coeficiente de correlação	Correlação Positiva	Coeficiente de correlação	Correlação Negativa
$r_R = 1$	Perfeita	$r_R = -1$	Perfeita
$0,95 \leq r_R < 1$	Muito forte	$-1 \leq r_R < -0,95$	Muito forte
$0,8 \leq r_R < 0,95$	Forte	$-0,95 \leq r_R < -0,8$	Forte
$0,5 \leq r_R < 0,8$	Moderada	$-0,8 \leq r_R < -0,5$	Moderada
$0 \leq r_R < 0,5$	Fraca	$-0,5 \leq r_R < 0$	Fraca

Coeficiente de correlação de Kendall

Teste não paramétrico indicado para número pequeno de amostras ou para populações com grandes quantidades de empates (valores repetidos).

Pode ser utilizado juntamente com o Spearman para comparação.

É mais conservador que o teste de Spearman.

Cálculo do Coeficiente de Kendall

$x_i > x_j$ e $y_i > y_j$ ou se $x_i < x_j$ e $y_i < y_j$.　　$x_i > x_j$ e $y_i < y_j$ ou se $x_i < x_j$ e $y_i > y_j$.

$$\tau = \frac{(\text{quantidade de pares concordantes}) - (\text{quantidade de pares discordantes})}{n(n-1)/2}$$

Coeficiente de correlação	Correlação Positiva	Coeficiente de correlação	Correlação Negativa
τ = 1	Perfeita	τ = - 1	Perfeita
0,95 ≤ τ < 1	Muito forte	- 1 ≤ τ < -0,95	Muito forte
0,8 ≤ τ < 0,95	Forte	-0,95 ≤ τ < -0,8	Forte
0,5 ≤ τ < 0,8	Moderada	-0,8 ≤ τ < -0,5	Moderada
0 ≤ τ < 0,5	Fraca	-0,5 ≤ τ < 0	Fraca