

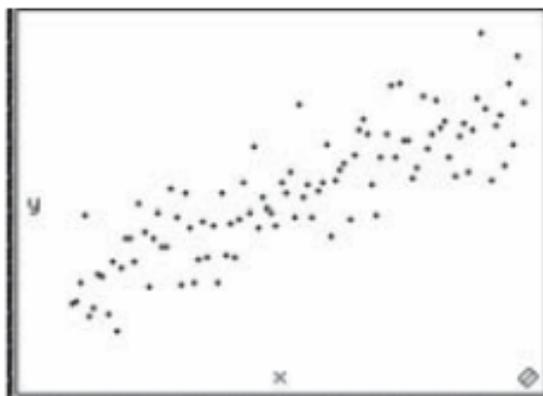
Correlação entre variáveis

Existe uma correlação entre duas variáveis quando os valores de uma variável estão relacionados, de alguma maneira, com os valores da outra variável.

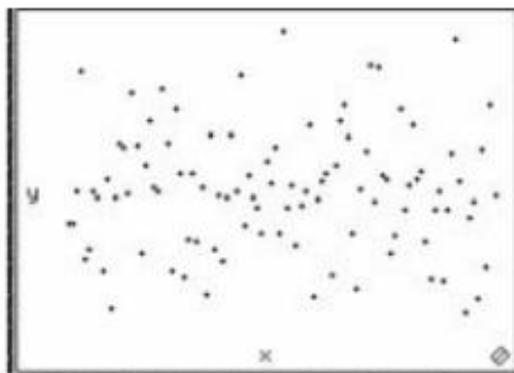
Uma vez definida a associação entre duas variáveis podemos identificar uma **equação** que melhor se ajusta aos dados e podemos usar essa equação para prever o valor de uma variável, conhecendo-se a outra.

Exploração dos dados

Diagramas de dispersão

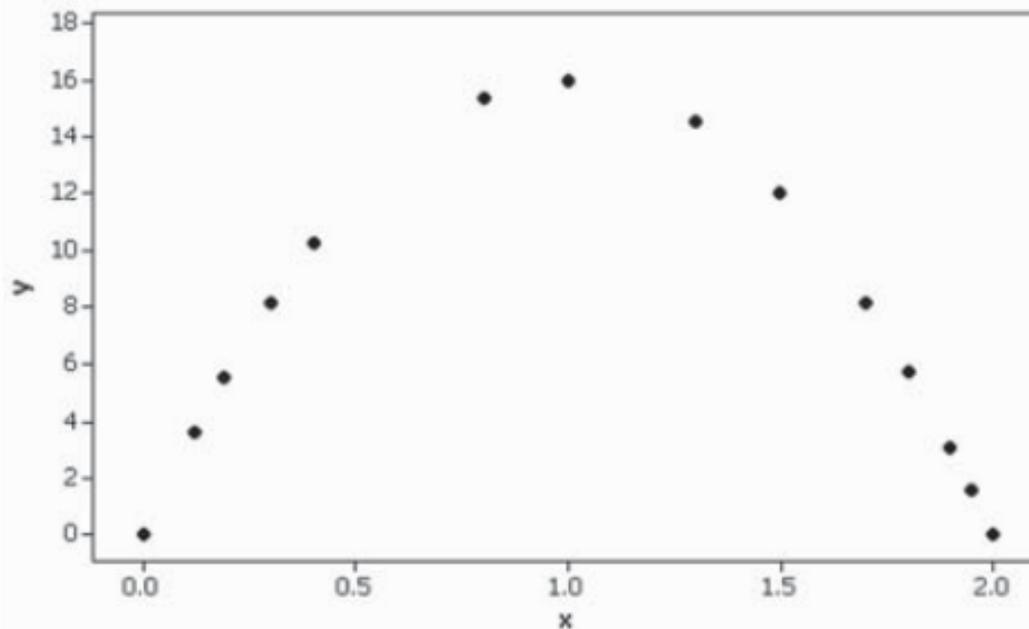


(a) **Correlação positiva**



(c) **Nenhuma correlação**

Minitab



(b) **Relação não linear**

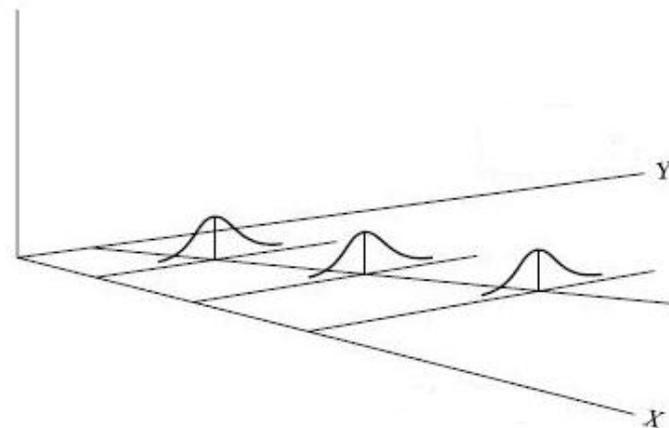
Coeficiente de correlação linear

O coeficiente de correlação linear r mede a força da correlação linear entre valores quantitativos emparelhados de x e y em uma amostra.

Dada qualquer coleção de dados quantitativos emparelhados, o coeficiente de correlação linear (r) pode sempre ser calculado, mas os seguintes requisitos devem ser satisfeitos ao se usarem dados amostrais para se tirar uma conclusão sobre correlação na população (ρ):

- 1) *Amostra aleatória simples*
- 2) *Distribuição normal bivariada*

- Exame visual do diagrama de dispersão sugere padrão de reta
- Ausência de outliers



Coeficiente de correlação linear (Pearson)

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

Notação:

N: número de pares de dados amostrais.

$\sum x$: soma dos valores x.

$\sum y$: soma dos valores y.

$\sum x^2$: soma dos valores x^2 .

$\sum xy$: soma dos valores dos produtos dos pares xy.

r : coeficiente correlação linear para amostras.

ρ : coeficiente correlação linear para população.

Coeficiente de correlação linear

Cálculo do coeficiente de correlação linear (r)

$$r = \frac{\sum \left[\frac{(x - \bar{x})}{s_x} \frac{(y - \bar{y})}{s_y} \right]}{n - 1}$$


$$r = \frac{\sum (z_x z_y)}{n - 1}$$

Se o valor absoluto de r , denotado por $|r|$, exceder o valor na Tabela de valores de r , conclua que há uma correlação linear. Caso contrário, não há evidência suficiente para apoiar a conclusão da existência de uma correlação.

Propriedades do coeficiente de correlação r

- 1) Os valores de r estão sempre compreendidos no intervalo $-1 \leq r \leq +1$
- 2) O valor de r não é alterado se todos os valores de qualquer uma das variáveis forem convertidos para outra escala
- 3) O valor de r não é afetado pela escolha de x como y ou vice-versa
- 4) “ r ” mede a intensidade de uma correlação **linear**
- 5) “ r ” é muito sensível a valores atípicos

Exemplo

Encontre o valor do coeficiente de correlação linear r para os custos emparelhados pizza/tarifa do metrô apresentados na Tabela abaixo e conclua se existe associação entre eles.

Elementary Statistics, Mario F Triola 11th

**Custos de uma Fatia de Pizza e Tarifa de Metrô
(em dólares)**

Custo da Pizza	0,15	0,35	1,00	1,25	1,75	2,00
Tarifa do Metrô	0,15	0,35	1,00	1,35	1,50	2,00

Requisitos (apenas para o caso de teste de hipótese)

- 1) Os dados são uma amostra aleatória simples de dados quantitativos
- 2) Os pontos do diagrama de dispersão se aproximam de uma reta
- 3) Não há outliers

X (pizza)	Y (metrô)	x^2	y^2	xy
0,15	0,15	0,0225	0,0225	0,0225
0,35	0,35	0,1225	1,1225	0,1225
1,00	1,00	1,00	1,00	1,00
1,25	1,35	1,5625	1,8225	1,6875
1,75	1,5	3,0625	2,25	2,6250
2,00	2,00	4,00	4,00	4,00
$\Sigma x=6,5$	$\Sigma y=6,35$	$\Sigma x^2 =9,77$	$\Sigma y^2 =9,2175$	$\Sigma xy=9,4575$

Usando a fórmula do Coeficiente de correlação linear

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

$$r = \frac{6(9,4575) - (6,5).(6,35)}{\sqrt{6(9,77) - (6,5)^2} \cdot \sqrt{6(9,2175) - (6,35)^2}}$$

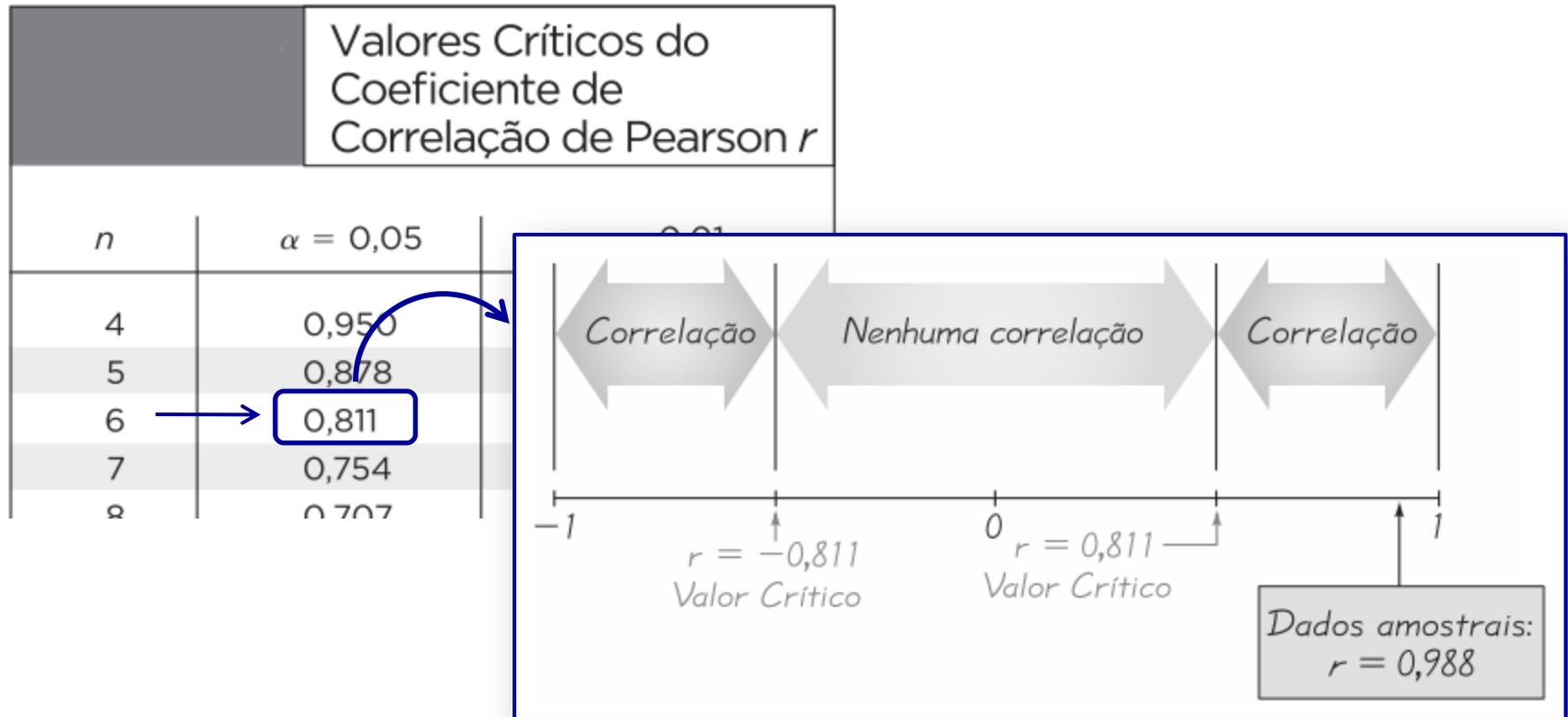
$$r = 0,988$$

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
15	0,514	0,641
16	0,497	0,623
17	0,482	0,606
18	0,468	0,590
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
30	0,361	0,463
35	0,334	0,430
40	0,312	0,403
45	0,294	0,380
50	0,279	0,361
55	0,266	0,345
60	0,254	0,330
65	0,244	0,317
70	0,235	0,306
75	0,227	0,296
80	0,220	0,286
85	0,213	0,278
90	0,207	0,270
95	0,202	0,263
100	0,197	0,256

Tabela de valores críticos do coeficiente de Correlação de Pearson

Interpretação do coeficiente de correlação linear (r)

Por meio da tabela de valores críticos



Aplicação 1

Encontre o valor do coeficiente de correlação linear r para os custos emparelhados pizza/tarifa do metrô apresentados na Tabela abaixo e conclua se existe associação entre eles.

Elementary Statistics, Mario F Triola 11th Edition

**Custos de uma Fatia de Pizza e Tarifa de Metrô
(em dólares)**

Custo da Pizza	0,15	0,35	1,00	1,25	1,75	2,00
Tarifa do Metrô	0,15	0,35	1,00	1,35	1,50	2,00

$$r = \frac{\sum(z_x z_y)}{n - 1} = \frac{4,93905}{5} = 0,988$$

Interpretação do coeficiente de correlação linear (r)

Por meio do conceito de variação explicada

O valor de r^2 é a proporção da variação em y que é explicada pela relação linear entre x e y .

Cuidado!

- 1) Correlação não implica causalidade
- 2) Correlação não pode ser baseada em dados obtidos de médias. As médias suprimem a variação individual e isso pode aumentar o coeficiente de correlação
- 3) Se não há correlação linear, pode existir outra correlação que não seja linear