



Medidas de Associação e Correlação

Estatística Aplicada

Ana Maria Nogales Vasconcelos

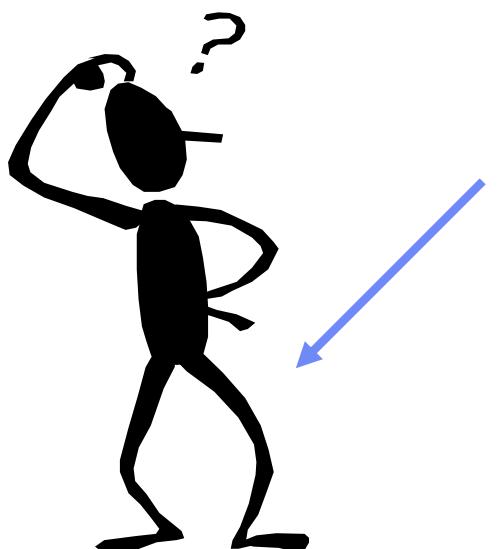
Maria Teresa Leão Costa



Medidas de Associação para Variáveis Quantitativas

O Problema ...

- ◆ Deseja-se verificar se existe associação entre duas variáveis quantitativas X e Y.



$$(X_i, Y_i) \quad i = 1, 2, \dots, n$$

***Como medir a
intensidade da relação
entre 2 variáveis ?***



Exemplo

1

Em um estudo sobre a existência de associação entre experiência e produtividade foram coletadas as seguintes informações em uma amostra de 10 agentes de seguro. Para cada agente de seguro, a experiência foi medida em anos de serviço e a produtividade em número de clientes.

Agente	Anos de Serviço	Número de Clientes
A	2	48
B	3	50
C	4	56
D	5	52
E	4	43
F	6	60
G	7	62
H	8	58
I	8	64
J	10	72
Total	57	565



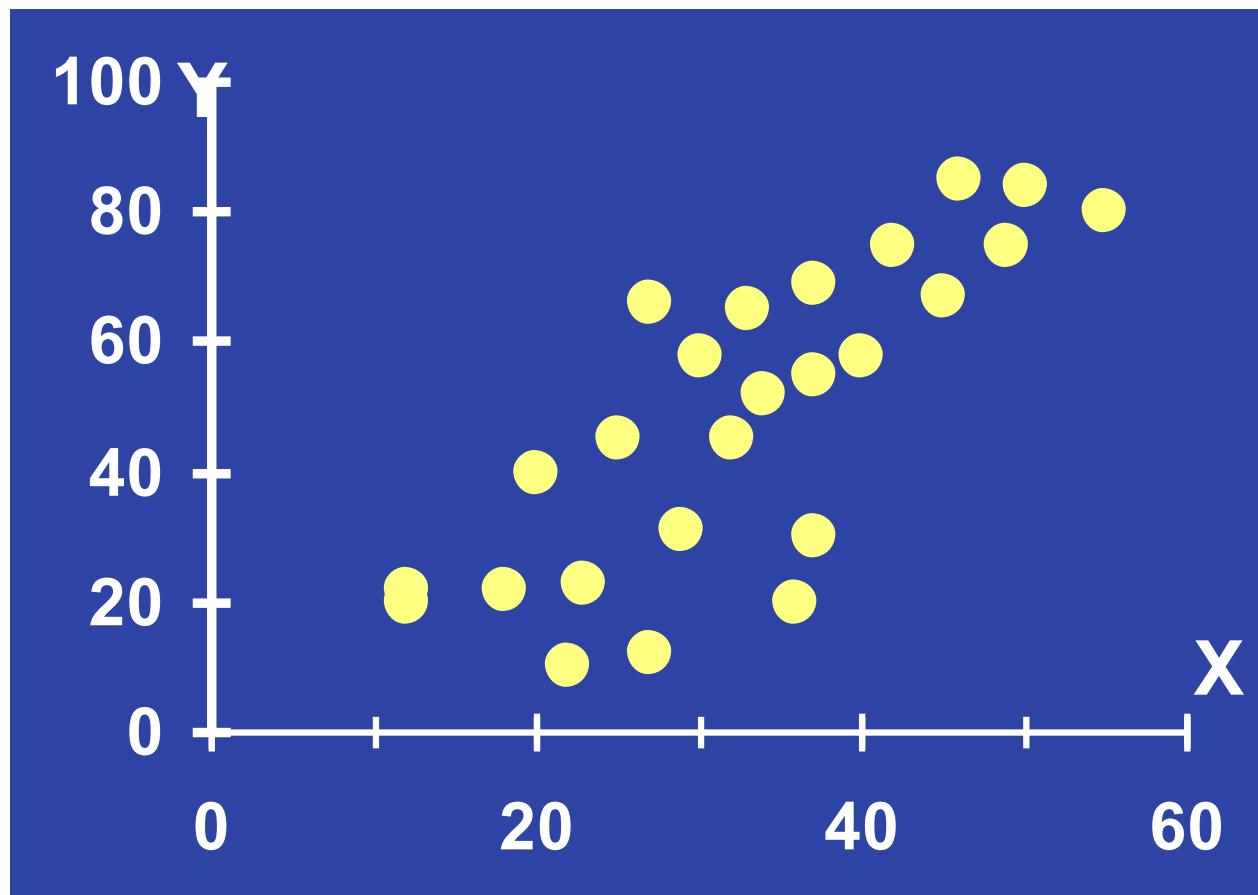
Os Dados

i	X_i	Y_i
1	X_1	Y_1
2	X_2	Y_2
:	:	:
n	X_n	Y_n



Diagrama de Dispersão

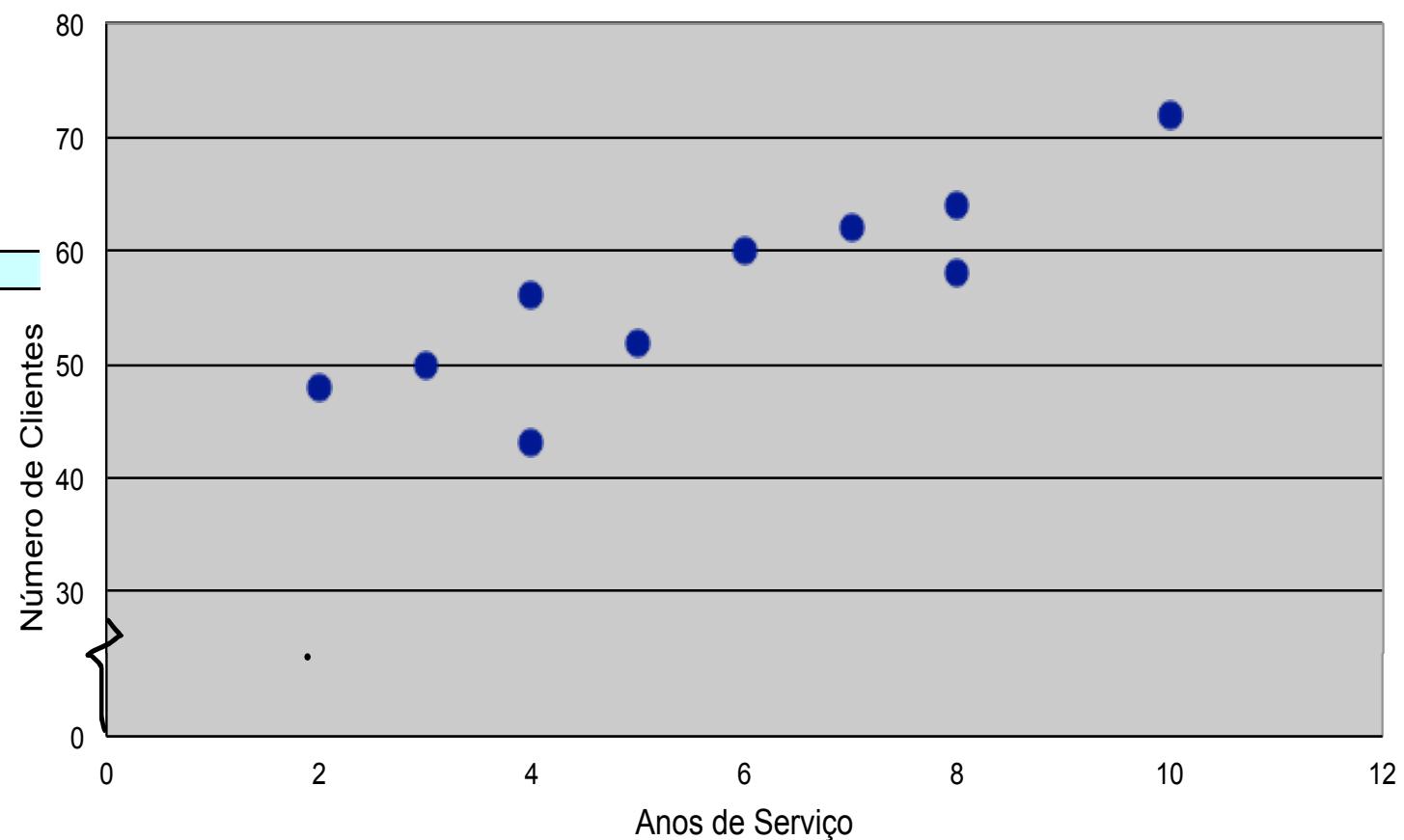
- ◆ Representação dos pares de valores $(x_i, y_i), i = 1, 2, \dots, n$ num sistema cartesiano;
- ◆ Recurso gráfico bastante útil para se verificar associação entre variáveis quantitativas.

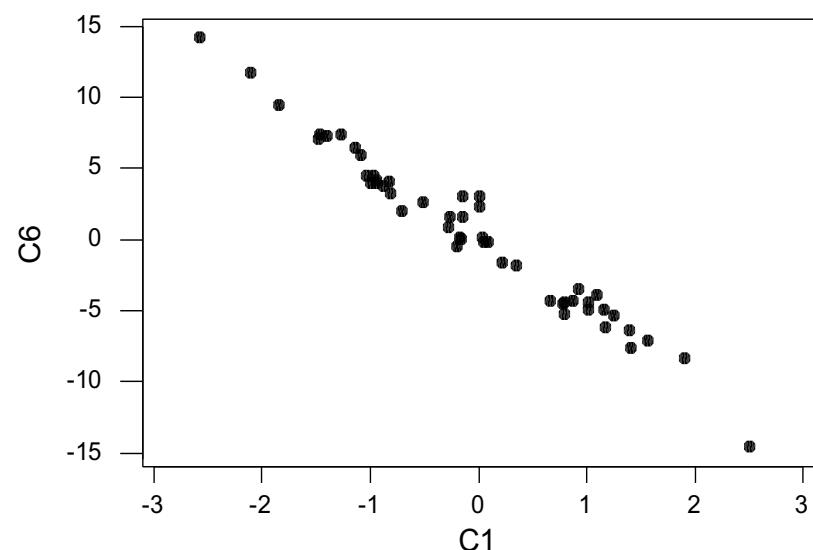
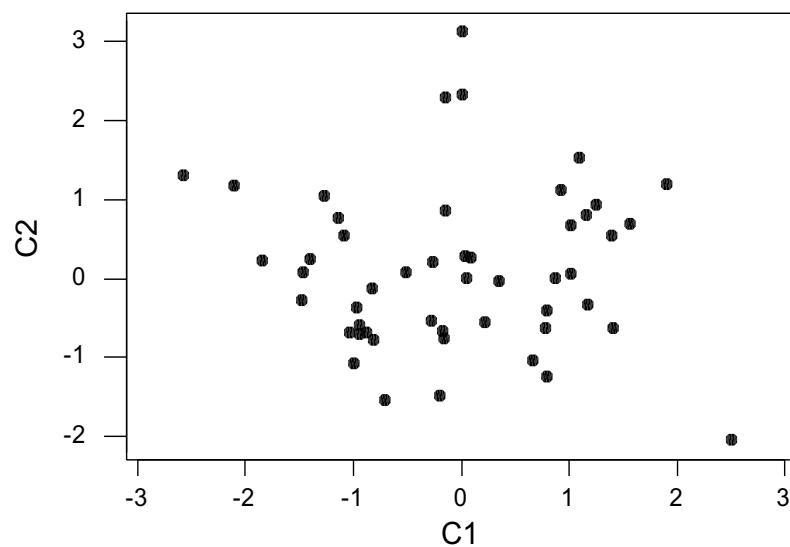
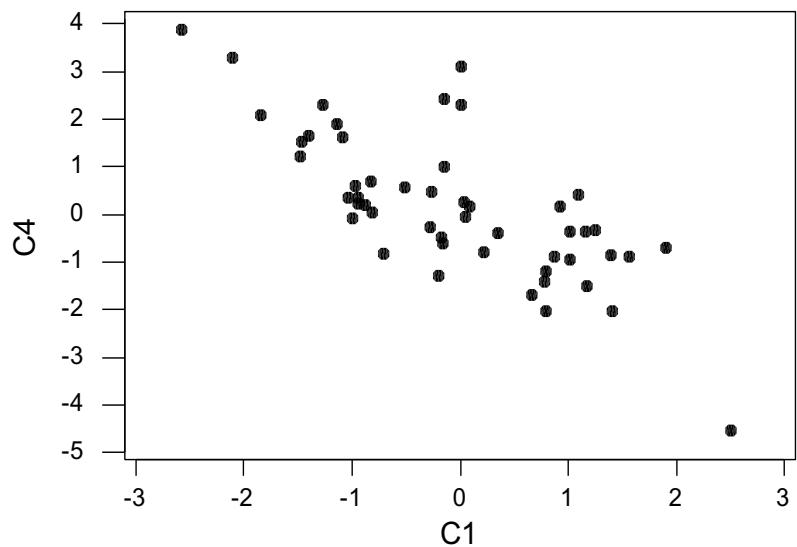


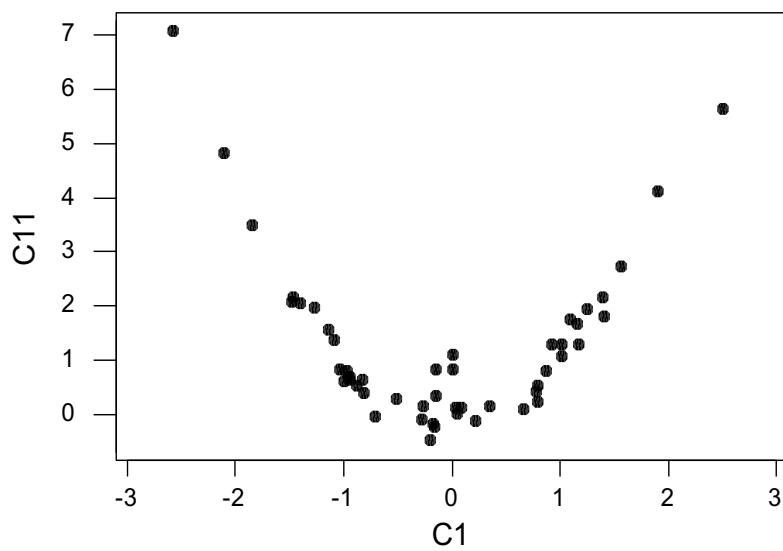
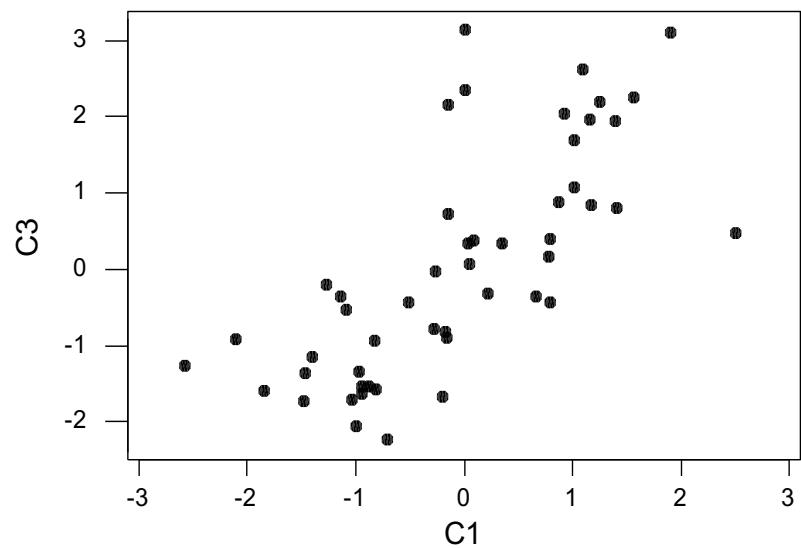
Exemplo



Agente	Anos de Serviço	Número de Clientes
A	2	48
B	3	50
C	4	56
D	5	52
E	4	43
F	6	60
G	7	62
H	8	58
I	8	64
J	10	72
Total	57	565





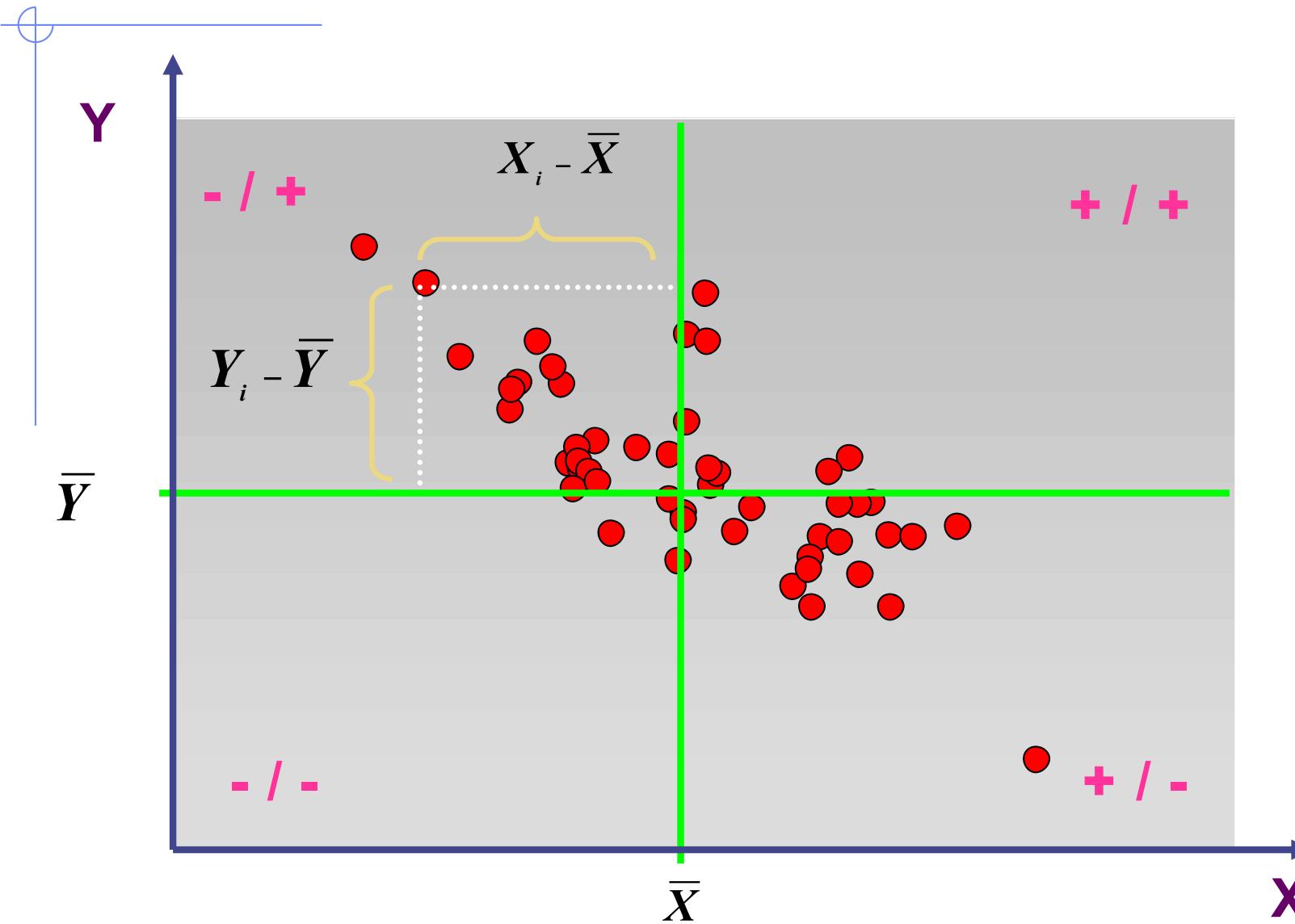




Coeficiente de Correlação Linear

■ Medida do grau de relacionamento **LINEAR** entre duas variáveis quantitativas.

- ρ (rho) - Coeficiente de correlação populacional
- r – Coeficiente de correlação amostral

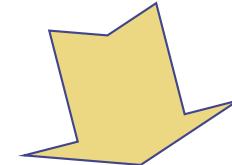




- ◆ Posição relativa de um valor qualquer num conjunto de valores:

$$\left. \begin{array}{l} z_x = \frac{x_i - \bar{x}}{S_x} \\ z_y = \frac{y_i - \bar{y}}{S_y} \end{array} \right\}$$

Padronização das variáveis



Torna comparáveis grupos de valores, mesmo que os desvios padrões dos grupos sejam diferentes

Coeficiente de Correlação Linear Amostral

■ Coeficiente de Correlação Momento- produto de Pearson:

$$r = \frac{\sum_{i=1}^n z_x \times z_y}{n-1} = \frac{1}{n-1} \sum_{i=1}^n \frac{(X_i - \bar{X})}{s_x} \times \frac{(Y_i - \bar{Y})}{s_y}$$



Coeficiente de Correlação Linear Amostral

■ Fórmula Operacional:

$$r = \frac{\sum_{i=1}^n X_i Y_i - n \bar{X} \bar{Y}}{\sqrt{\sum_{i=1}^n X_i^2 - n \bar{X}^2} \cdot \sqrt{\sum_{i=1}^n Y_i^2 - n \bar{Y}^2}}$$



Cálculo de r



i	X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$
1	X_1	Y_1	X_1^2	Y_1^2	$X_1 Y_1$
2	X_2	Y_2	X_2^2	Y_2^2	$X_2 Y_2$
:	:	:	:	:	:
n	X_n	Y_n	X_n^2	Y_n^2	$X_n Y_n$
Σ	ΣX_i	ΣY_i	ΣX_i^2	ΣY_i^2	$\Sigma X_i Y_i$



Exemplo

Agente	x_i	y_i
	Anos de Serviço	Número de Clientes
A	2	48
B	3	50
C	4	56
D	5	52
E	4	43
F	6	60
G	7	62
H	8	58
I	8	64
J	10	72
Total	57	565

$\sum x_i$

$\sum y_i$

Exemplo

	x_i	y_i	
Agente	Anos de Serviço	Número de Clientes	x_i^2
A	2	48	4
B	3	50	9
C	4	56	16
D	5	52	25
E	4	43	16
F	6	60	36
G	7	62	49
H	8	58	64
I	8	64	64
J	10	72	100
Total	57	565	383

$$\sum x_i$$

$$\sum y_i$$

$$\sum x_1^2$$

Exemplo

	x_i	y_i		
Agente	Anos de Serviço	Número de Clientes	x_i^2	y_i^2
A	2	48	4	2304
B	3	50	9	2500
C	4	56	16	3136
D	5	52	25	2704
E	4	43	16	1849
F	6	60	36	3600
G	7	62	49	3844
H	8	58	64	3364
I	8	64	64	4096
J	10	72	100	5184
Total	57	565	383	32581

$$\sum x_i$$

$$\sum y_i$$

$$\sum x_i^2$$

$$\sum y_i^2$$



Exemplo

	x_i	y_i				
Agente	Anos de Serviço	Número de Clientes	x_i^2	y_i^2	$x_i \times y_i$	
A	2	48	4	2304	96	
B	3	50	9	2500	150	
C	4	56	16	3136	224	
D	5	52	25	2704	260	
E	4	43	16	1849	172	
F	6	60	36	3600	360	
G	7	62	49	3844	434	
H	8	58	64	3364	464	
I	8	64	64	4096	512	
J	10	72	100	5184	720	
Total	57	565	383	32581	3392	

$$\sum x_i$$

$$\sum y_i$$

$$\sum x_i^2$$

$$\sum y_i^2$$

$$\sum x_i y_i$$

Exemplo

$$\bar{x} = \frac{\sum x_i}{n} = \frac{57}{10} = 5,7$$

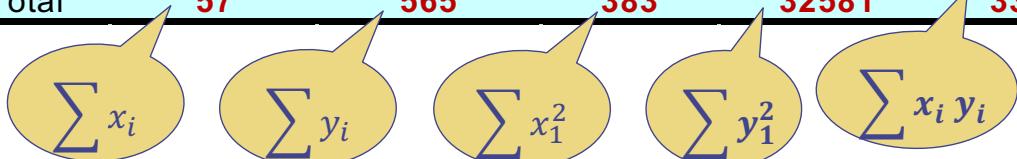
$$\bar{y} = \frac{\sum y_i}{n} = \frac{565}{10} = 56,5$$

Como:

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum x_i^2 - n \bar{x}^2} \sqrt{\sum y_i^2 - n \bar{y}^2}}$$

$$r = \frac{3392 - 10 \times 5,7 \times 56,5}{\sqrt{383 - 10 \times 5,7^2} \sqrt{32581 - 10 \times 56,5^2}} =$$

Agente	<i>x_i</i>	<i>y_i</i>	<i>x_i²</i>	<i>y_i²</i>	<i>x_i × y_i</i>
	Anos de Serviço	Número de Clientes			
A	2	48	4	2304	96
B	3	50	9	2500	150
C	4	56	16	3136	224
D	5	52	25	2704	260
E	4	43	16	1849	172
F	6	60	36	3600	360
G	7	62	49	3844	434
H	8	58	64	3364	464
I	8	64	64	4096	512
J	10	72	100	5184	720
Total	57	565	383	32581	3392





Exemplo

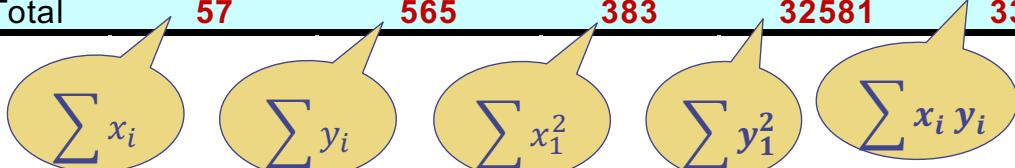
$$\bar{x} = \frac{\sum x_i}{n} = \frac{57}{10} = 5,7$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{565}{10} = 56,5$$

Como:

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum x_i^2 - n \bar{x}^2} \sqrt{\sum y_i^2 - n \bar{y}^2}}$$

Agente	x_i	y_i	x_i^2	y_i^2	$x_i \times y_i$
	Anos de Serviço	Número de Clientes			
A	2	48	4	2304	96
B	3	50	9	2500	150
C	4	56	16	3136	224
D	5	52	25	2704	260
E	4	43	16	1849	172
F	6	60	36	3600	360
G	7	62	49	3844	434
H	8	58	64	3364	464
I	8	64	64	4096	512
J	10	72	100	5184	720
Total	57	565	383	32581	3392

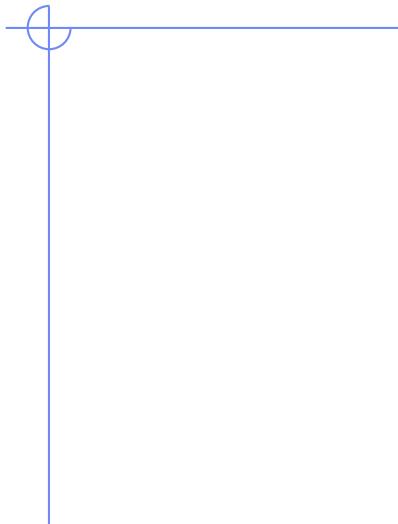


$$r = \frac{3392 - 10 \times 5,7 \times 56,5}{\sqrt{383 - 10 \times 5,7^2} \sqrt{32581 - 10 \times 56,5^2}} =$$

$$= \frac{3392 - 3220,5}{\sqrt{383 - 324,9} \sqrt{32581 - 31922,5}} = \frac{171,5}{\sqrt{58,1} \times \sqrt{658,5}} = 0,877$$



Interpretação de r





Interpretação de r

**Coeficiente de
Correlação**

r



*Sentido
da
relação
(Sinal de r)*

*Intensidade
da
relação
($|r|$)*



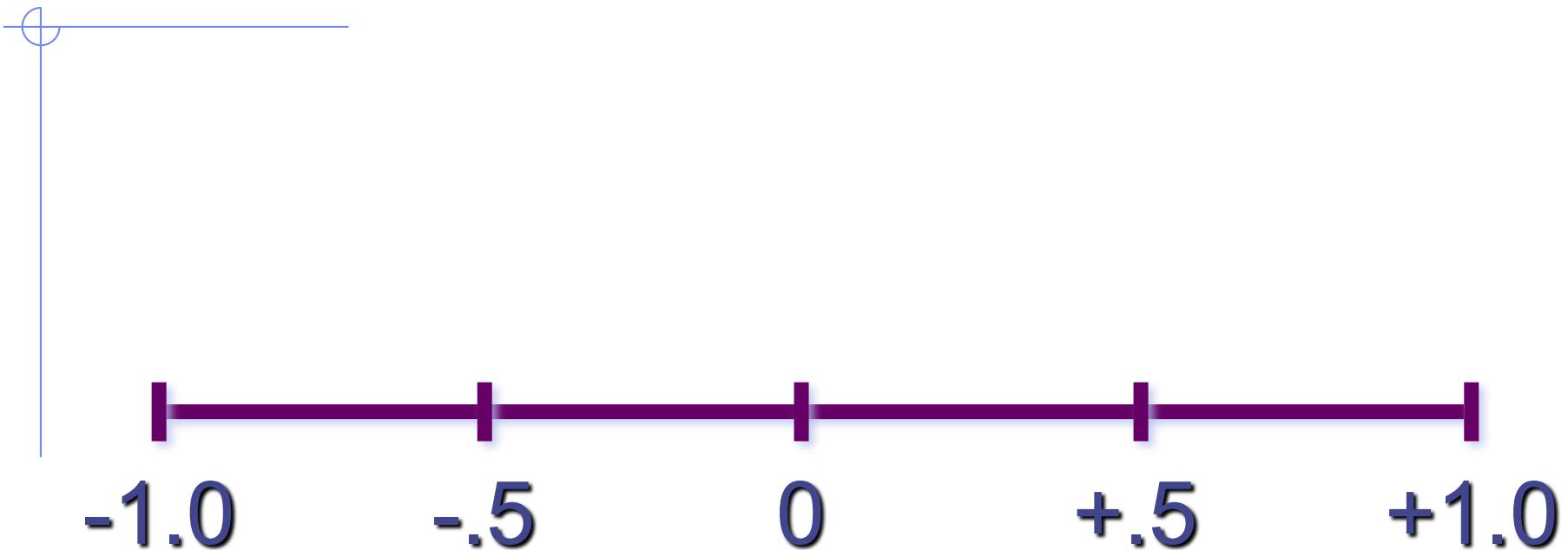
Sentido da Relação

- ◆ **Correlação Positiva** - quanto maior o valor de X, maior o valor de Y

- ◆ **Correlação Negativa** - quanto maior o valor de X, menor o valor de Y

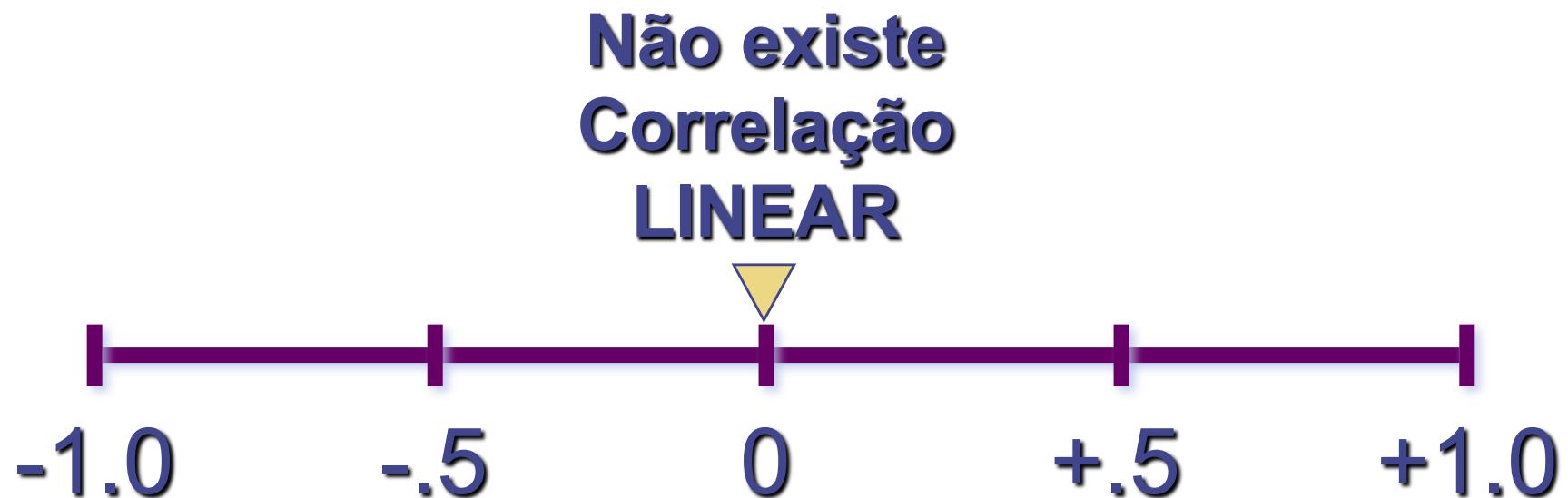


Interpretação de r



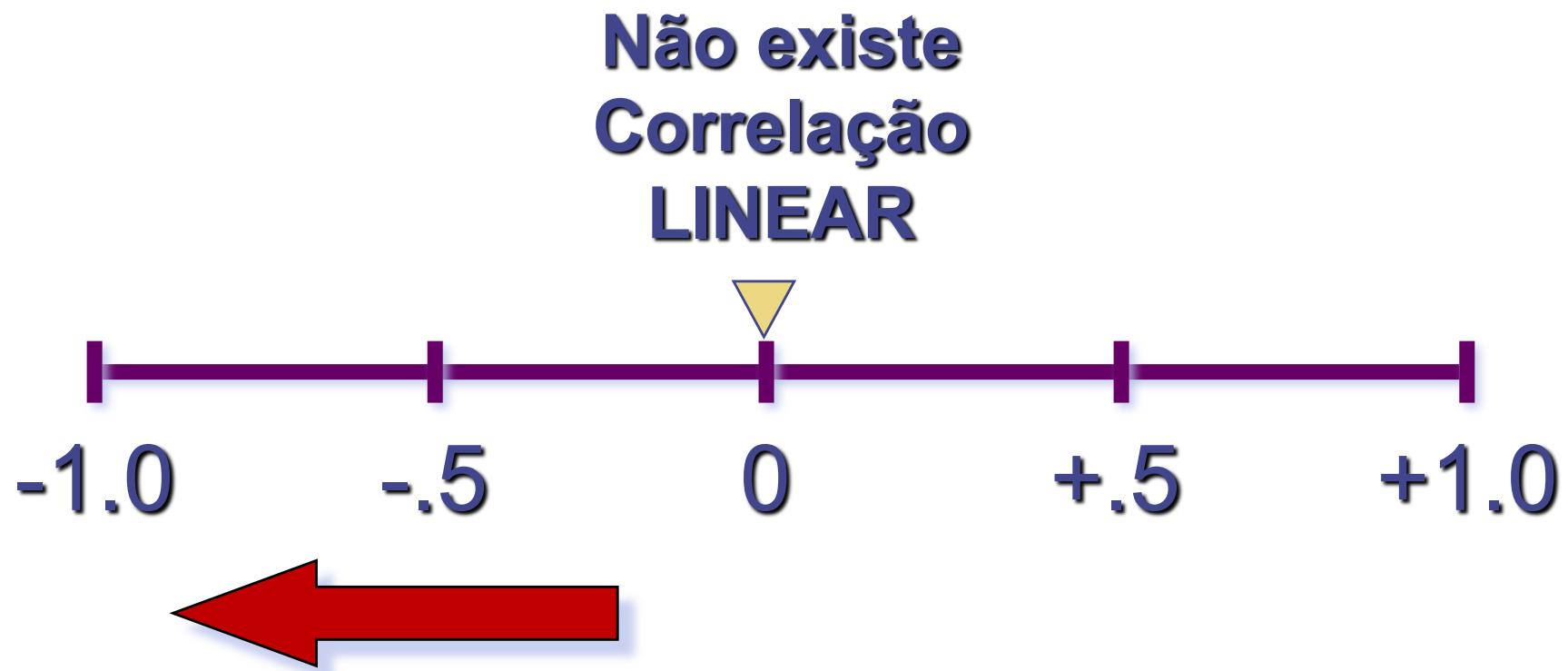


Interpretação de r





Interpretação de r



**Grau crescente de
correlação negativa**



Interpretação de r

**Correlação
Negativa
Perfeita**



-1.0

**Não existe
Correlação
LINEAR**



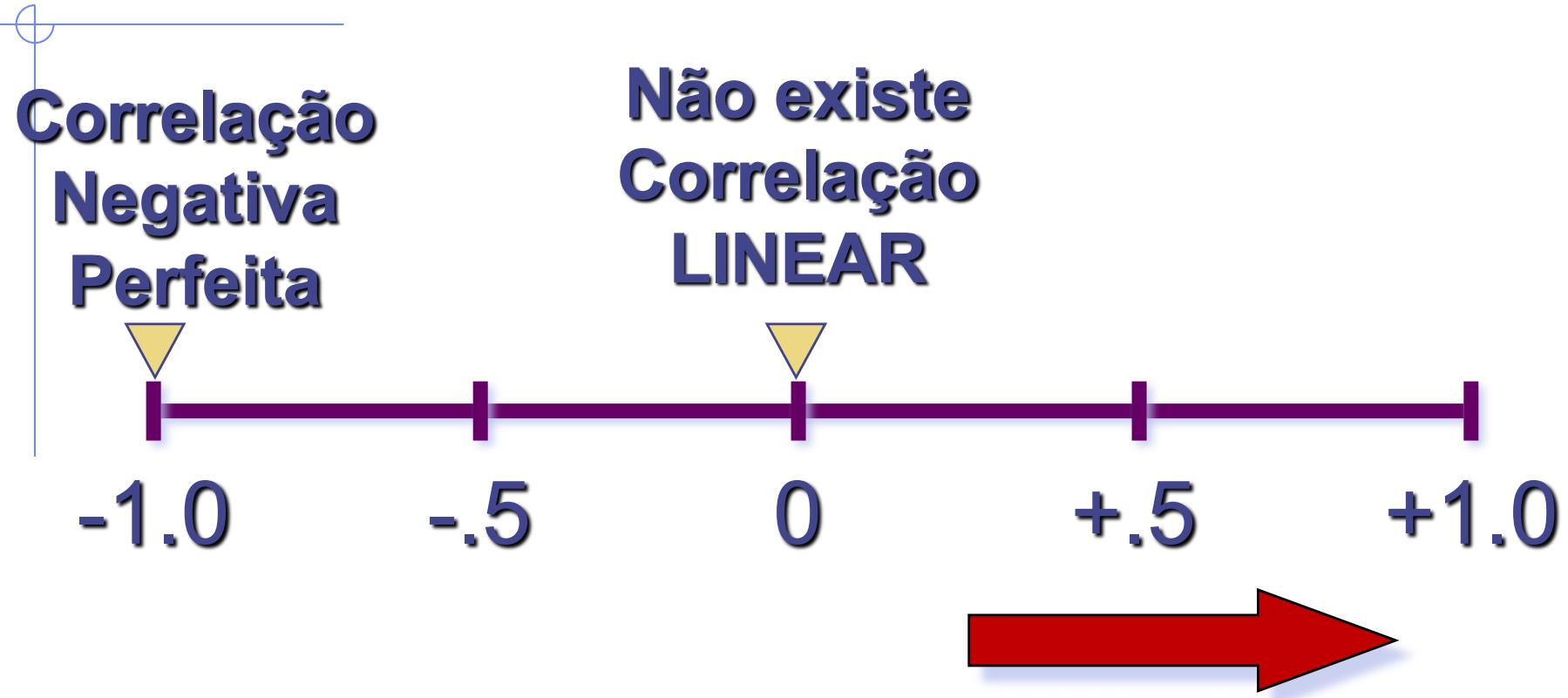
0

.5

+1.0

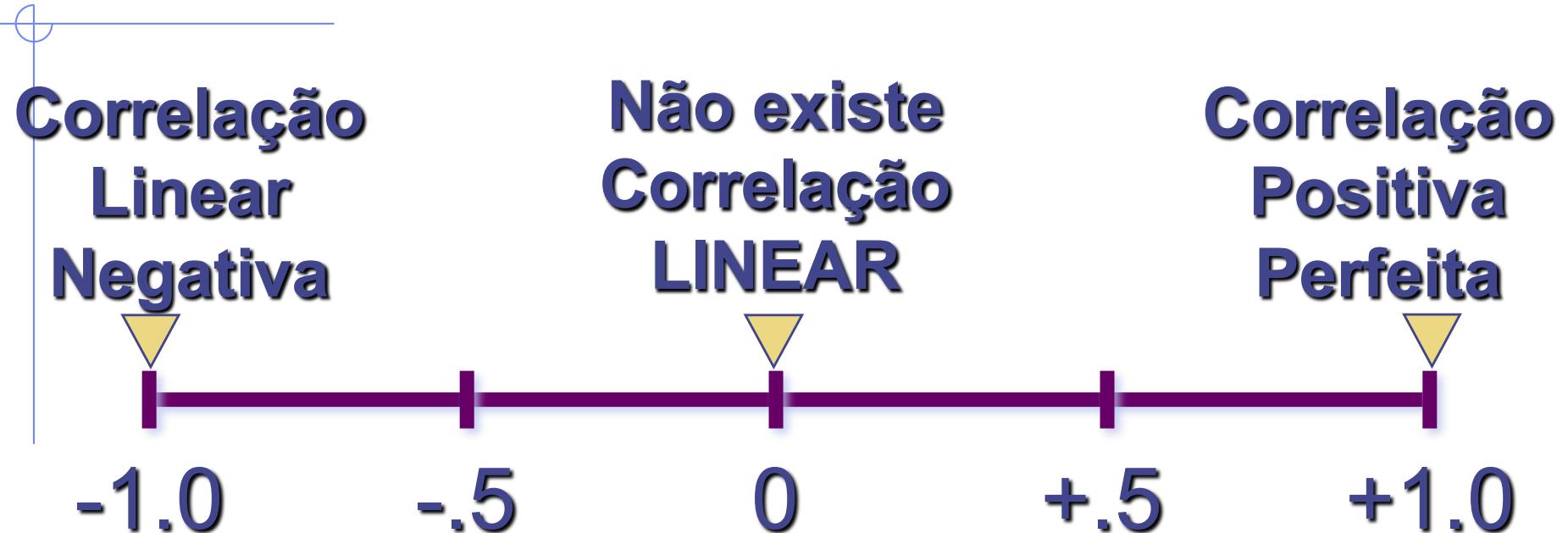


Interpretação de r





Interpretação de r





Exemplo

$$\bar{x} = \frac{\sum x_i}{n} = \frac{57}{10} = 5,7$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{565}{10} = 56,5$$

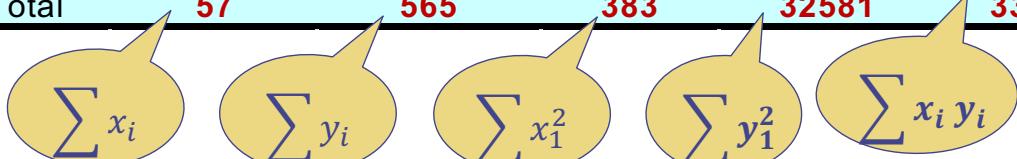
Como:

$$r = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sqrt{\sum x_i^2 - n \bar{x}^2} \sqrt{\sum y_i^2 - n \bar{y}^2}}$$

$$r = \frac{3392 - 10 \times 5,7 \times 56,5}{\sqrt{383 - 10 \times 5,7^2} \sqrt{32581 - 10 \times 56,5^2}} =$$

$$= \frac{3392 - 3220,5}{\sqrt{383 - 324,9} \sqrt{32581 - 31922,5}} = \frac{171,5}{\sqrt{58,1} \times \sqrt{658,5}} = 0,877$$

Agente	x_i	y_i	x_i^2	y_i^2	$x_i \times y_i$
	Anos de Serviço	Número de Clientes			
A	2	48	4	2304	96
B	3	50	9	2500	150
C	4	56	16	3136	224
D	5	52	25	2704	260
E	4	43	16	1849	172
F	6	60	36	3600	360
G	7	62	49	3844	434
H	8	58	64	3364	464
I	8	64	64	4096	512
J	10	72	100	5184	720
Total	57	565	383	32581	3392



Correlação forte e positiva