REGRESSÃO LINEAR Por: Ricardo Reis

INTRODUÇÃO

CORRELAÇÃO

REGRESSÃO LINEAR SIMPLES

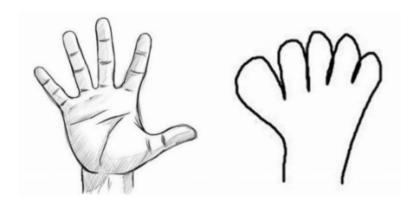
CASO - BOSTON HOUSING

REFERÊNCIAS

INTRODUÇÃO

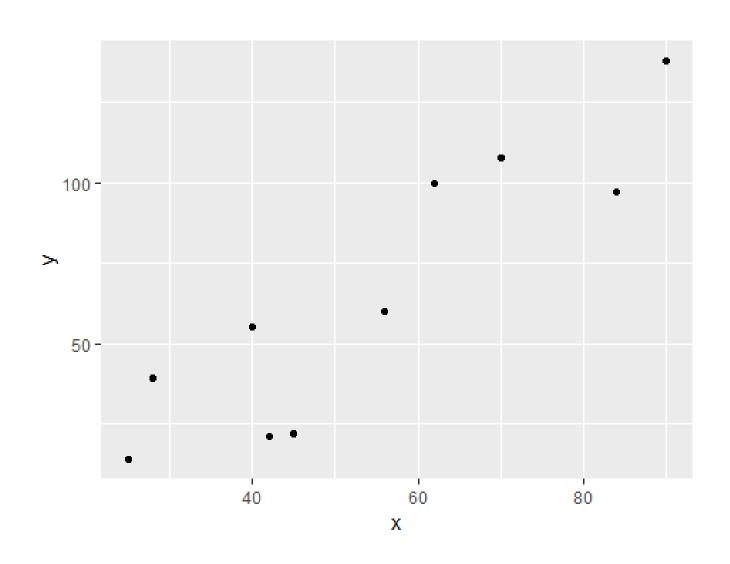
- Como o número de quartos e a área útil influenciam na formação de preços de aluguel de imóveis?
- Meu investimento em propaganda está me dando retorno?
- Eu posso prever as vendas desse mês a partir do investimento em propaganda na rádio, TV e jornal ?
- Se os vendedores fizerem determinado treinamento, terão melhor performance de vendas ?
- A SELIC, o dóllar e a inflação influenciam no valor das ações de uma determinada empresa na bolsa?

Modelos são uma simplificação ou aproximação da realidade.

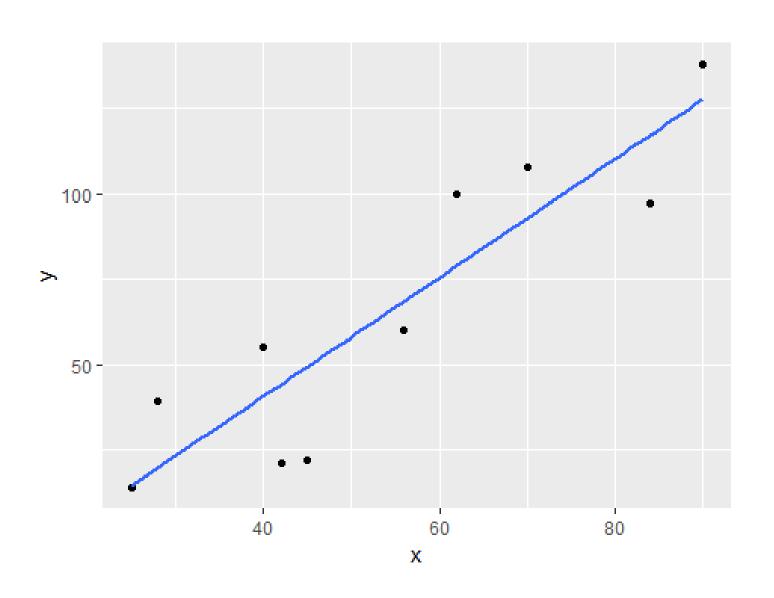


CORRELAÇÃO

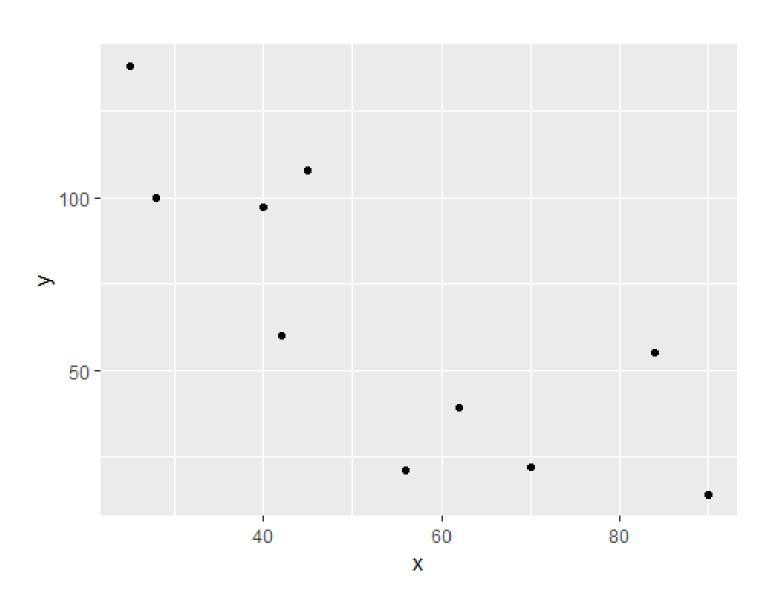
O gráfico sugere relação linear entre as variáveis?



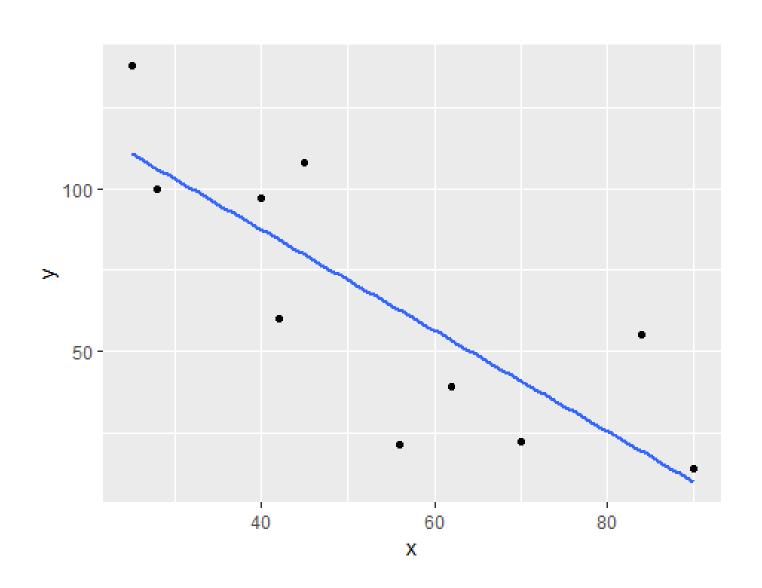
correlação = 0,9



Existe tendência?



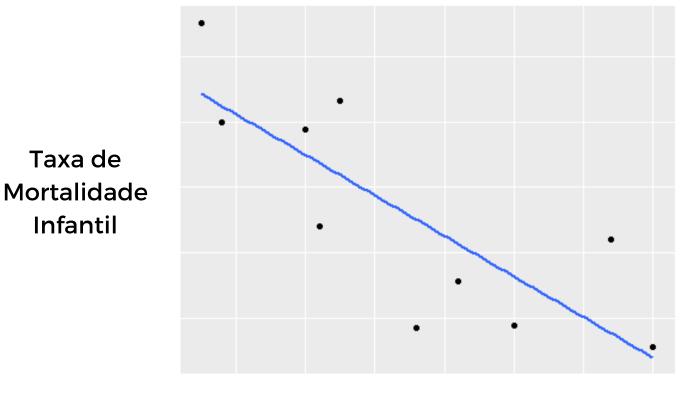
correlação = -0,8



Uma maneira de quantificar a força da relação entre 2 variáveis.

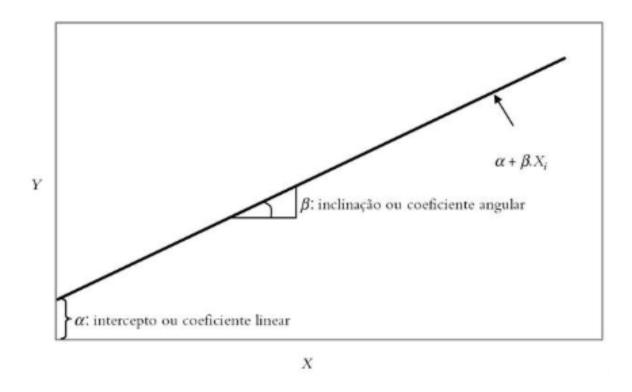


Correlação **NÃO** é causalidade!



TV's/Lar

REGRESSÃO LINEAR SIMPLES



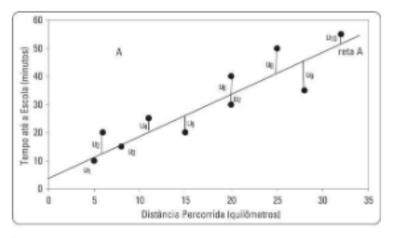
$$Y = a + bx + e$$

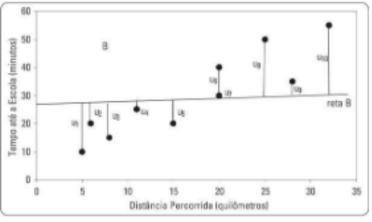
- Y: Variável explicada ou dependente
- a: Intercepto (o que significa?)
- b: Coeficiente angular (o que significa?)
- x: Variável explicativa ou independente
- e: Resíduo

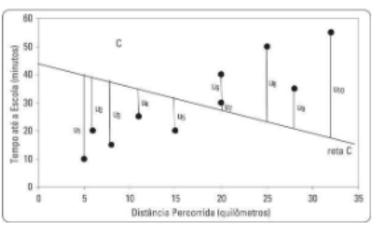
Mínimos Quadrados Ordinários

$$\sum_{i=1}^{n} u_i = 0$$

$$\sum_{i=1}^{n} u_i^2 = \min$$







Análise de Indicadores

```
call:
lm(formula = Vendas.de.Sclomps \sim Anuncios, data = anuncios)
Residuals:
    Min 1Q Median 3Q
                                  Max
-253.616 -127.006 -1.364 133.479 244.140
Coefficients:
          Estimate Std. Error t value Pr(>|t|)
Anuncios 9.9806 0.1025 97.40 <2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 146.8 on 327 degrees of freedom
Multiple R-squared: 0.9667, Adjusted R-squared: 0.9666
F-statistic: 9486 on 1 and 327 DF, p-value: < 2.2e-16
```

Análise de Indicadores

```
call:
lm(formula = Vendas.de.Sclomps ~ Anuncios, data = anuncios)
Residuals:
          1Q Median
    Min
                               3Q
                                      Max
-253.616 -127.006 -1.364 133.479 244.140
Coefficients:
           Fstimate Std. Error t value Pr(>|t|)
(Intercept) 258.2462
                    18.4332 14.01 <2e-16 ***
Anuncios
                                       <2e-16 ***
             9.9806
                       0.1025 97.40
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 146.8 on 327 degrees of freedom
Multiple R-squared: 0.9667,
                             Adjusted R-squared: 0.9666
F-statistic: 9486 on 1 and 327 DF, p-value: < 2.2e-16
```

Problema

Satisfação global

Deseja-se prever a satisfação global pela satisfação com o atendimento e pela satisfação com o produto.

satisfação global	satisfação com o atendimento	satisfação com o produto	
7	6	7	
8	7	7	
9	7	8	
4	5	4	
5	5	5	
7	6	7	
8	7	8	
6	6	6	
7	6	6	

Satisfação global em função da satisfação com o atendimento

Estatística de regi	ressão					
R múltiplo	0,943					
R-Quadrado	0,890					
R-quadrado ajustado	0,874					
Erro padrão	0,555					
Observações	9					
ANOVA						/
7410771	gl	SQ	MQ	F	F de significação	•
Regressão	1	17,396	17,396	56,401	0,000	
Resíduo	7	2,159	0,308			
Total	8	19,556				
	Coeficientes	Erro padrão	Stat t	valor-P	95% inferiores	95% superiores
Interseção	-4,75	1,546	-3,072	0,018	-8,406	-1,094
satisfação com atendimo	1,88636364	0,251	7,510	0,000	1,292	2,480

Satisfação global em função da satisfação com o produto

Estatística de reg	ressão					
R múltiplo	0,953					
R-Quadrado	0,908					
R-quadrado ajustado	0,895					
Erro padrão	0,508					
Observações	9					
ANOVA	gl	SQ	MQ	F	F de significação	
Regressão	1	17,751	17,751	68,852	0,000	
Resíduo	7	1,805	0,258			
Total	8	19,556				
	Coeficientes E	rro padrão	Stat t	valor-P	95% inferiores	95% superiores
Interseção	-0,421875	0,884	-0,477	0,648	-2,512	1,669
satisfação com o produt	1,1171875	0,135	8,298	0,000	0,799	1,436

Satisfação global em função das satisfações com o produto e com o atendimento

global = $\beta_0 + \beta_1$. atendimento + β_2 . produto + resíduo

Estatística de reg	gressão						
R múltiplo	0,971						
R-Quadrado	0,943						
R-quadrado ajustad	0,925 🖍	0,925			Pelo	menos um	
Erro padrão	0,430				coeficiente é		
Observações	9					ente de zero!	
ANOVA							
	gl	SQ	MQ	F	de significação	0	
Regressão	2	18,448	9,224	49,989	0,000		
Resíduo	6	1,107	0,185				
Total	8	19,556				•	
	Coeficientes E	peficientes Erro padrão		valor-P	95% inferiores	95% superiores	
Interseção	-2,821	1,443	Stat t -1,955	0,098	-6,353		
satisfação com aten	0,893	0,459	1,944	0,100	-0,231	2,017	
satisfação com o pro	0,643	0,269	2,388	0,054	-0,016	1,302	

ATENÇÃO

- Os coeficientes (betas) de todas as variáveis mudam quando uma outra variável é adicionada.
- Uma variável X sozinha pode ser significante para explicar Y, mas quando uma outra variável muito correlacionada com X entra no modelo, X pode se tornar não significante (Colinearidade).
- Deve-se colocar ou retirar do modelo uma variável por vez, para verificar se as outras variáveis se mantêm significantes.
- A alta correlação entre as duas variáveis independentes gera problemas de interpretação dos coeficientes e de significância.

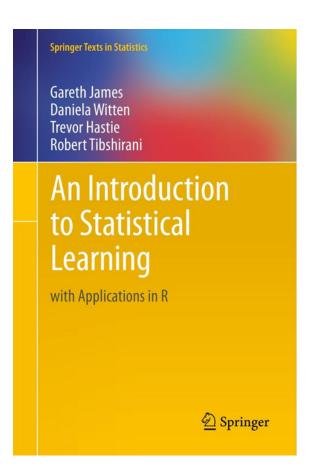
CASO BOSTON HOUSING

REFERÊNCIAS BIBLIOGRÁFICAS

Gareth James et al.

An Introduction to Statistical Learning

with Applications in R



Luiz Paulo Fávero

Análise de Dados

Modelos de Regressão com Excel, Stata e SPSS

