

Aprendizado de Máquina

Aprendizado Supervisionado
Regressão

Prof. Dr^a. Andreza Sartori
asartori@furb.br

Documentos Consultados/Recomendados

- RUSSEL, Stuart; NORVIG, Peter. **Inteligência artificial**. Rio de Janeiro : GEN LTC, 2013. 1 recurso online. Disponível em: <https://integrada.minhabiblioteca.com.br/books/9788595156104> . Acesso em: 26 jul. 2021.
- NG, Andrew; Guestrin, Carlos; Charikar, Moses. **Machine Learning**. Stanford University. Disponível em: <http://cs229.stanford.edu/materials.html>
- MALIK, Jitendra. **Computer Vision**. UC Berkeley. Disponível em: <https://www-inst.eecs.berkeley.edu/~cs280/sp15/index.html>
- IA Expert Academy. Plataforma de Cursos sobre Inteligência Artificial. Disponível em: <https://www.youtube.com/channel/UCaGrIWpwjWXT6OIQh9W4Riw>
- Louppe, Gilles. Deep Learning, ULiège, 2022. Disponível em: <https://github.com/glouppe/info8010-deep-learning>

Conteúdo Programático:

Unidade 1: Fundamentos de Aprendizado de Máquina

Unidade 2: Aprendizado Supervisionado

Unidade 3: Aprendizado Não Supervisionado

Unidade 4: Redes Neurais Artificiais

Unidade 5: Aplicações de Aprendizado de Máquina



Conteúdo Programático:

Unidade 1: Fundamentos de Aprendizado de Máquina

Unidade 2: Aprendizado Supervisionado

Unidade 3: Aprendizado Não Supervisionado

Unidade 4: Redes Neurais Artificiais

Unidade 5: Aplicações de Aprendizado de Máquina



Conteúdo Programático:

Unidade 1: Fundamentos de Aprendizado de Máquina

Unidade 2: Aprendizado Supervisionado

2.1 Regressão

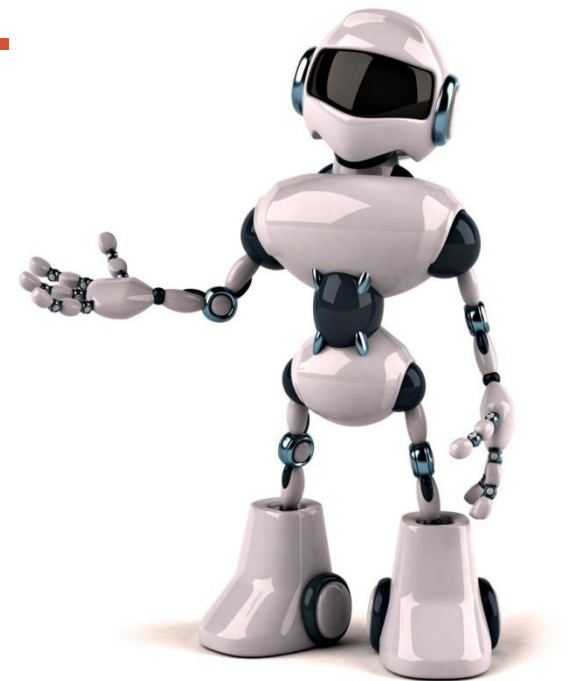
- 2.1.1 Regressão linear (simples e múltipla)
- 2.1.2 Regressão polinomial
- 2.1.4 Regressão com Vetores de Suporte (SVR)

2.2 Classificação

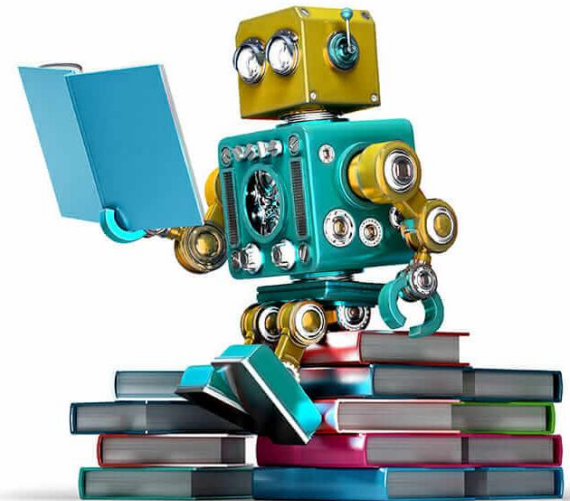
- 2.2.1 Aprendizagem Bayesiana (Naive Bayes)
- 2.2.2 Árvores de decisão
- 2.2.3 k-Nearest Neighbour (kNN)
- 2.2.4 Regressão logística
- 2.2.5 Support Vector Machine (SVM)



Recapitulando...



Como funciona o processo de aprendizagem?



Teoria da Aprendizagem

Dado um conjunto de treinamento de N pares de exemplos de entrada e saída

$$(x_1, y_1), (x_2, y_2), \dots (x_n, y_n),$$

onde cada valor de y pode ser encontrado por uma função desconhecida:

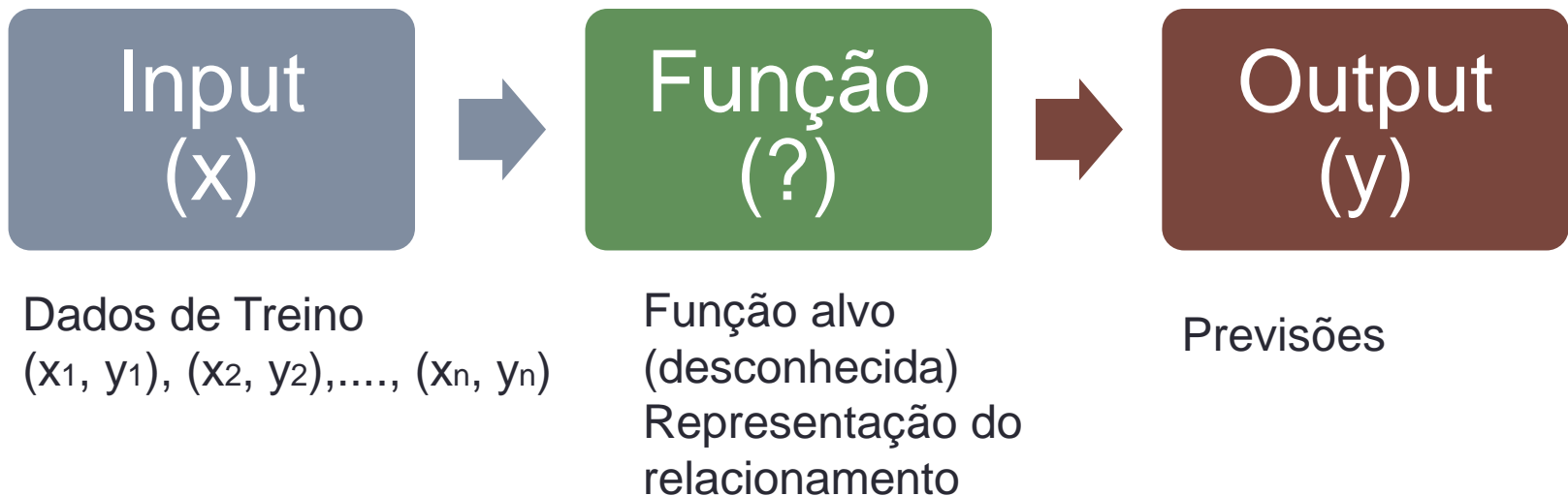
$$y = f(x),$$

o objetivo da aprendizagem é descobrir uma função **h (hipótese)** que se aproxime da função verdadeira f

Teoria da Aprendizagem

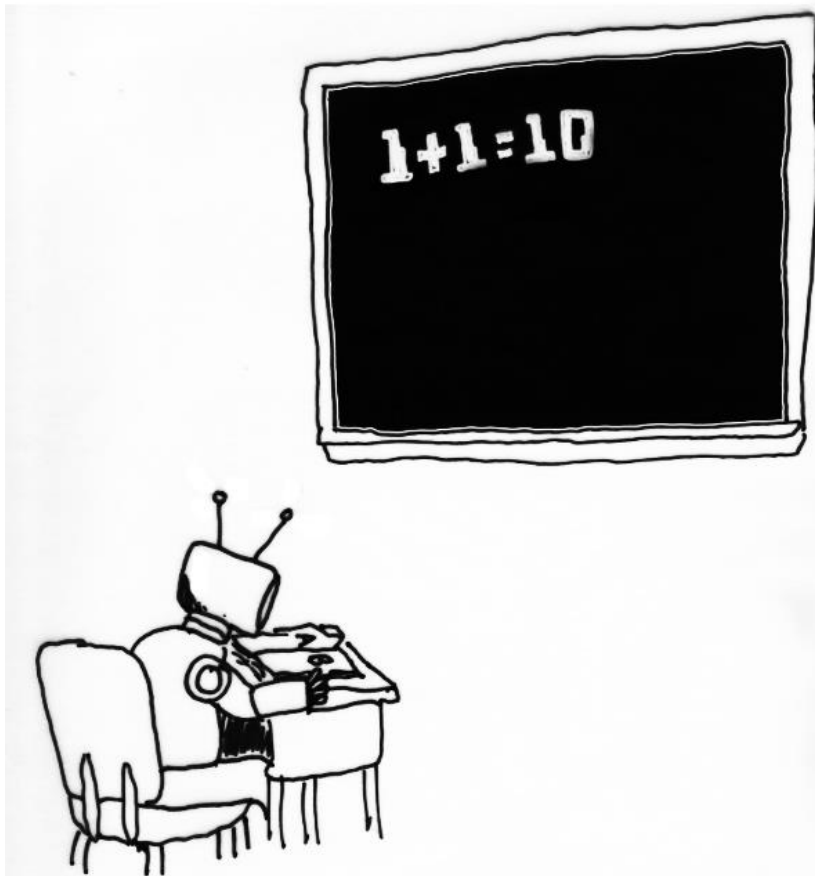
O objetivo da aprendizagem é descobrir uma função **h** (**hipótese**) que se aproxime da função verdadeira **f**

$$y = f(x)$$

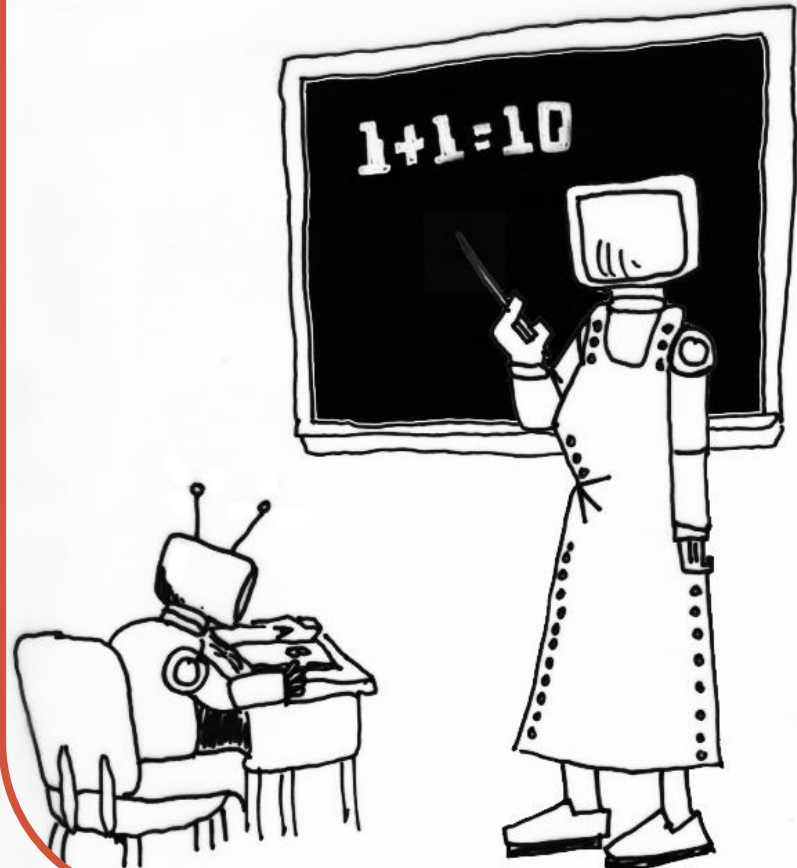


Nesta aula

UNSUPERVISED MACHINE LEARNING



SUPERVISED MACHINE LEARNING

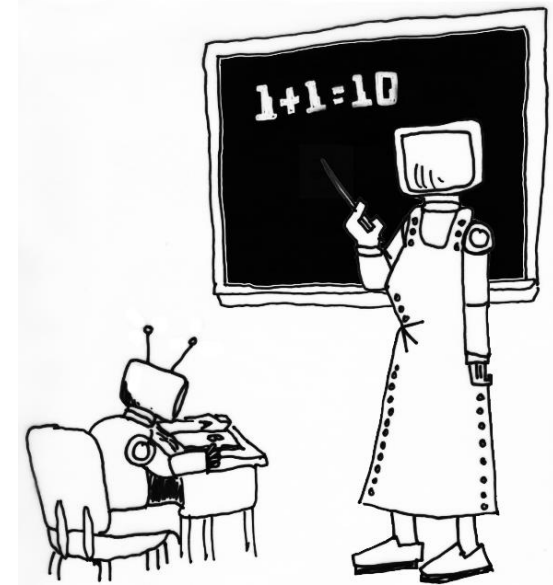


Aprendizado Supervisionado

- Damos ao sistema a “**resposta correta**” durante o processo de treinamento.
- Dado um conjunto de entradas de treinamento e saídas correspondentes, produz os resultados "corretos" para novas entradas.

SUPERVISED MACHINE LEARNING

- É eficiente pois o sistema pode trabalhar diretamente com informações corretas.



Abordagens do Aprendizado Supervisionado

- **Classificação:**

- Responde se uma determinada “entrada” pertence a uma certa classe.
- Dada a imagem de uma fruta: informa que fruta é (dentro um número finito de classes).

- **Regressão:**

- Faz uma predição a partir de exemplos.
- Prever o valor dos imóveis, dados os valores por metro quadrado.

Abordagens do Aprendizado Supervisionado

- **Classificação:**

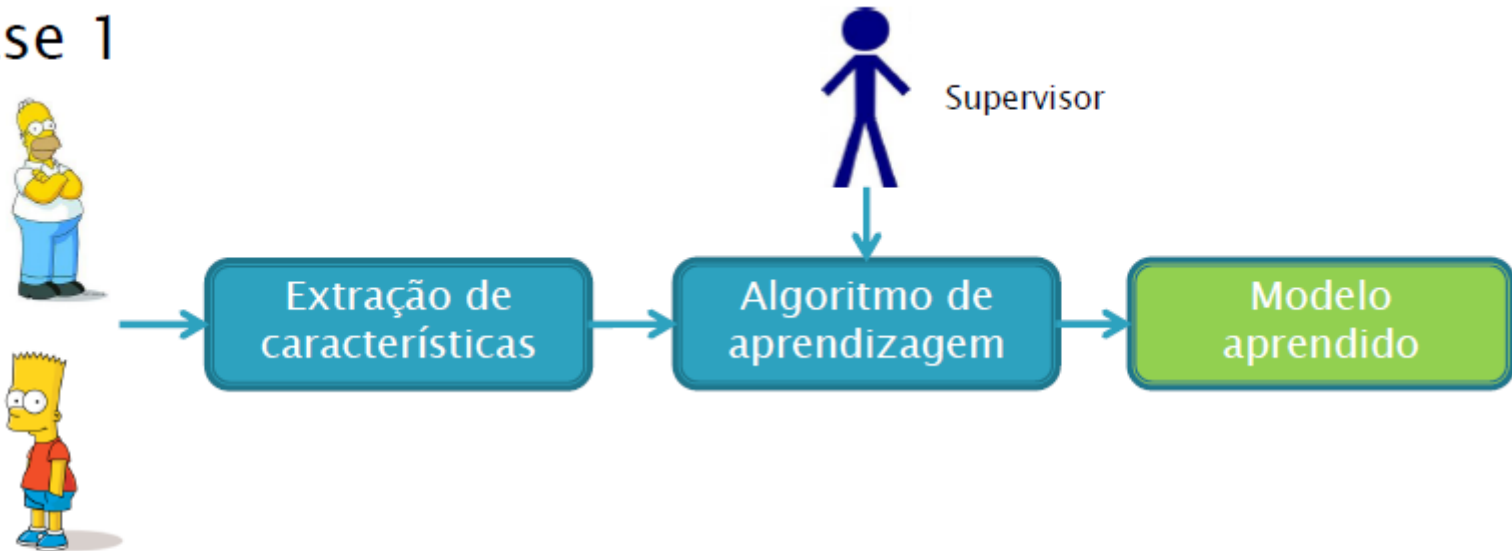
- Responde se uma determinada “entrada” pertence a uma certa classe.
- Dada a imagem de uma fruta: informa que fruta é (dentro um número finito de classes).

- **Regressão:**

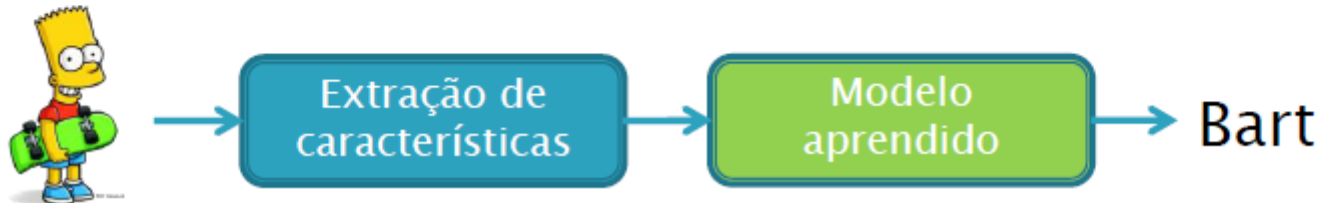
- Faz uma predição a partir de exemplos.
- Prever o valor dos imóveis, dados os valores por metro quadrado.

Aprendizado Supervisionado

Fase 1

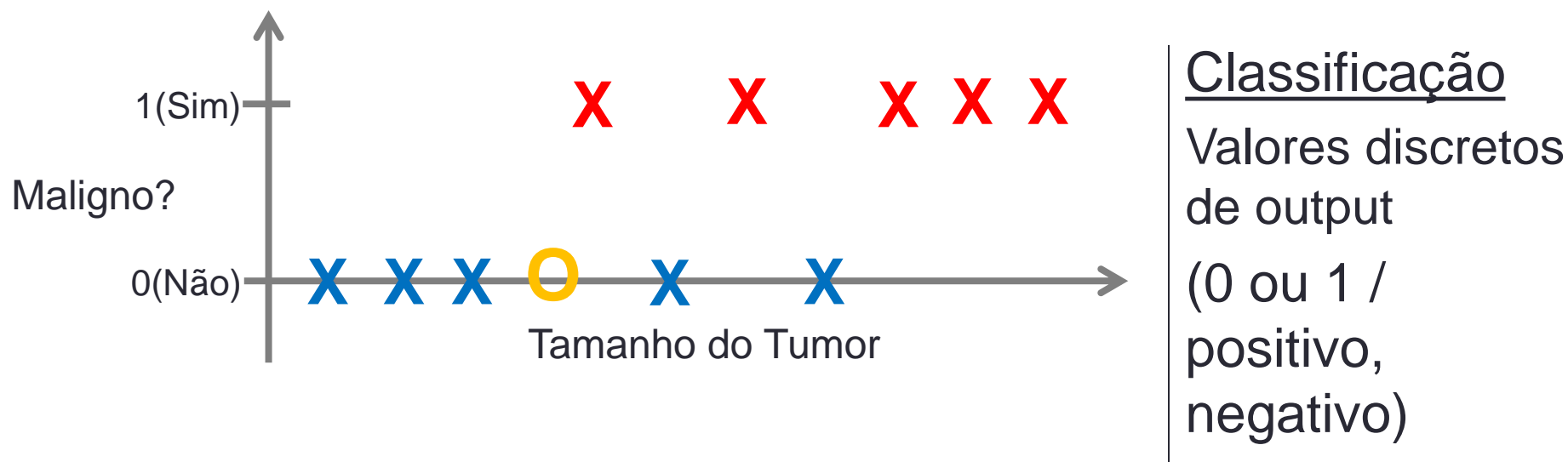


Fase 2



Aprendizado Supervisionado: Classificação

Prever se tumor na mama é Maligno ou Benigno.



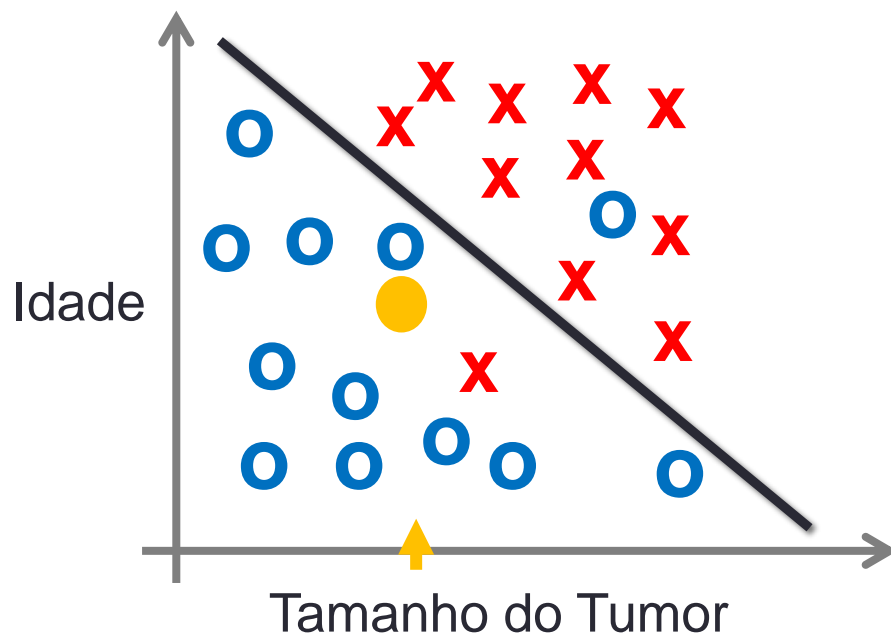
Qual é a probabilidade / chance de um tumor ser maligno ou benigno?

Pode ter mais de dois valores para valores possíveis de saída (multiclasse).

Exemplo: 0 (benigno), 1 (câncer tipo 1), 2 (câncer tipo 2), 3,n

Aprendizado Supervisionado: Classificação

Prever se tumor na mama é Maligno ou Benigno.



Abordagens do Aprendizado Supervisionado

- Classificação:

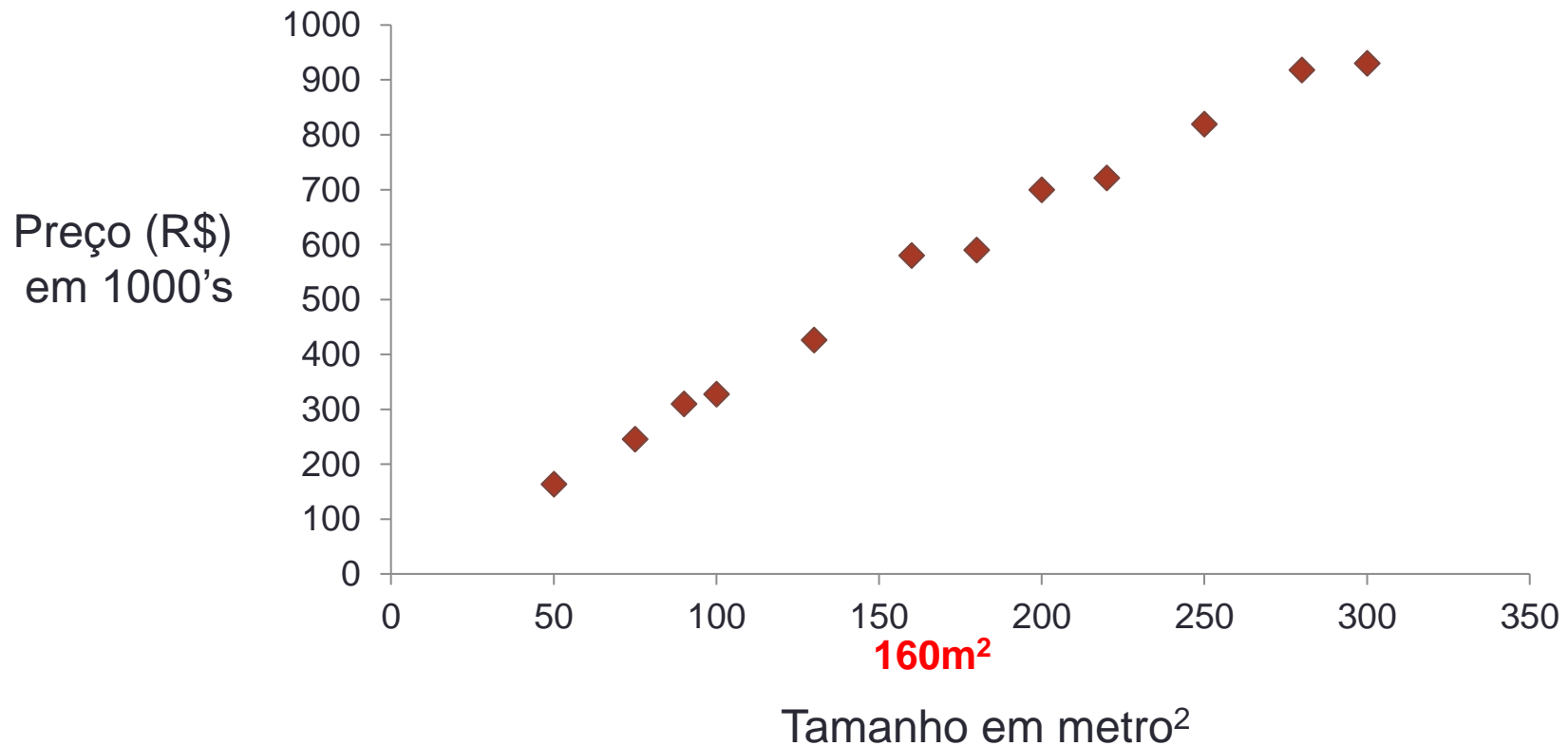
- Responde se uma determinada “entrada” pertence a uma certa classe.
- Dada a imagem de uma fruta: que fruta é (dentre um número finito).

- Regressão:

- Faz uma predição a partir de exemplos.
- Prever o valor dos imóveis, dados os valores por metro quadrado.

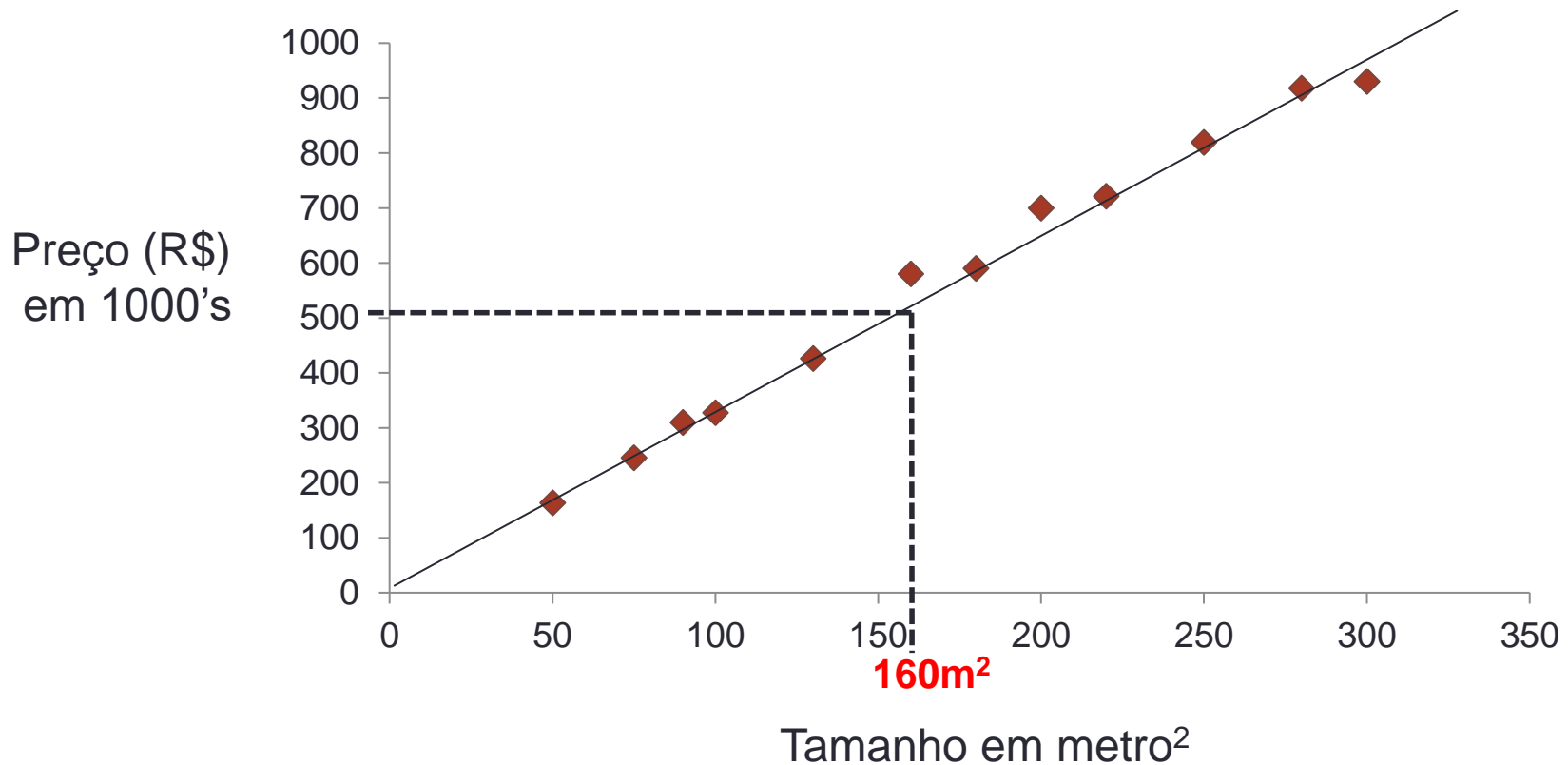
Aprendizado Supervisionado: Regressão

Prever o Preço de Imóveis



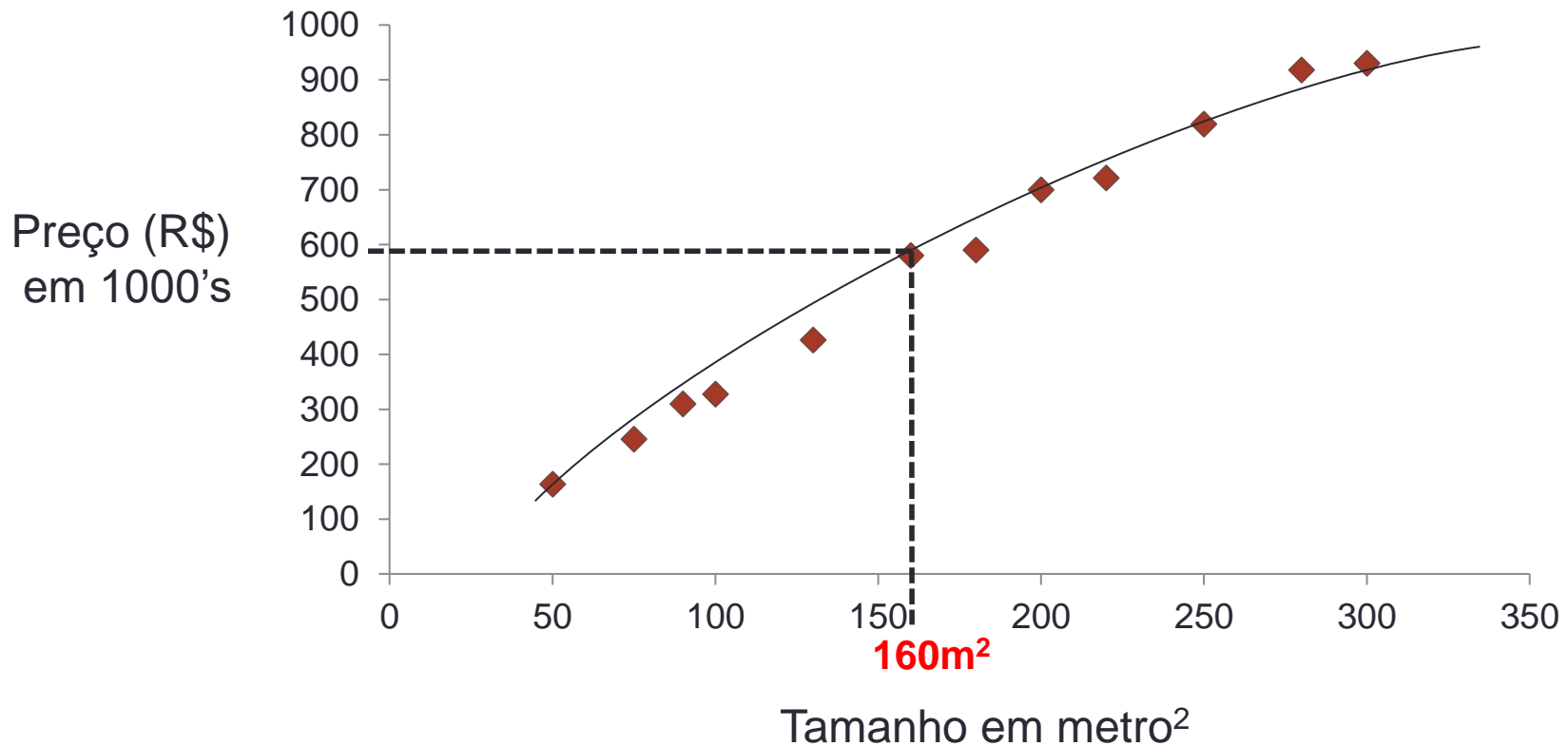
Aprendizado Supervisionado: Regressão

Prever o Preço de Imóveis



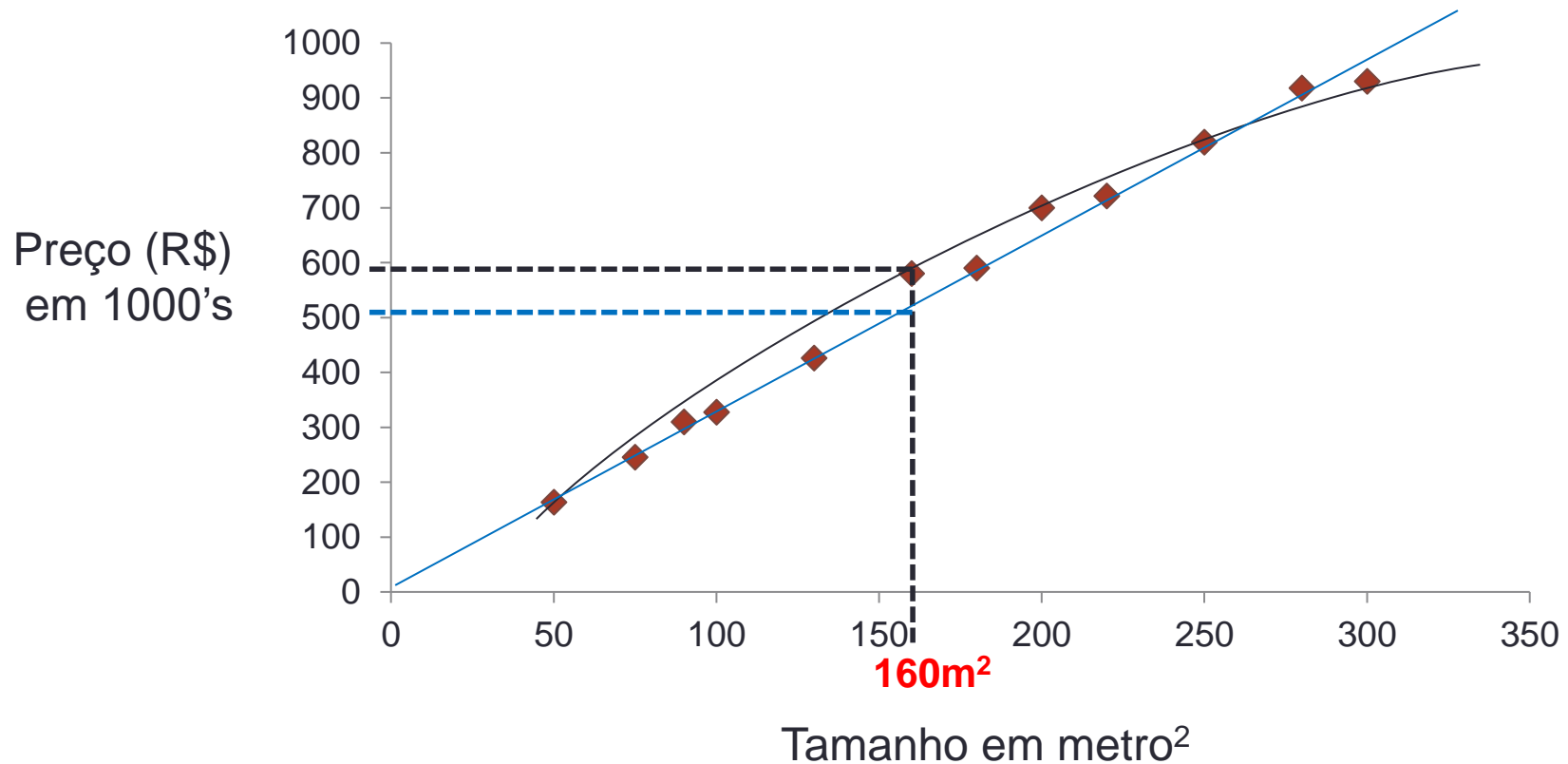
Aprendizado Supervisionado: Regressão

Prever o Preço de Imóveis



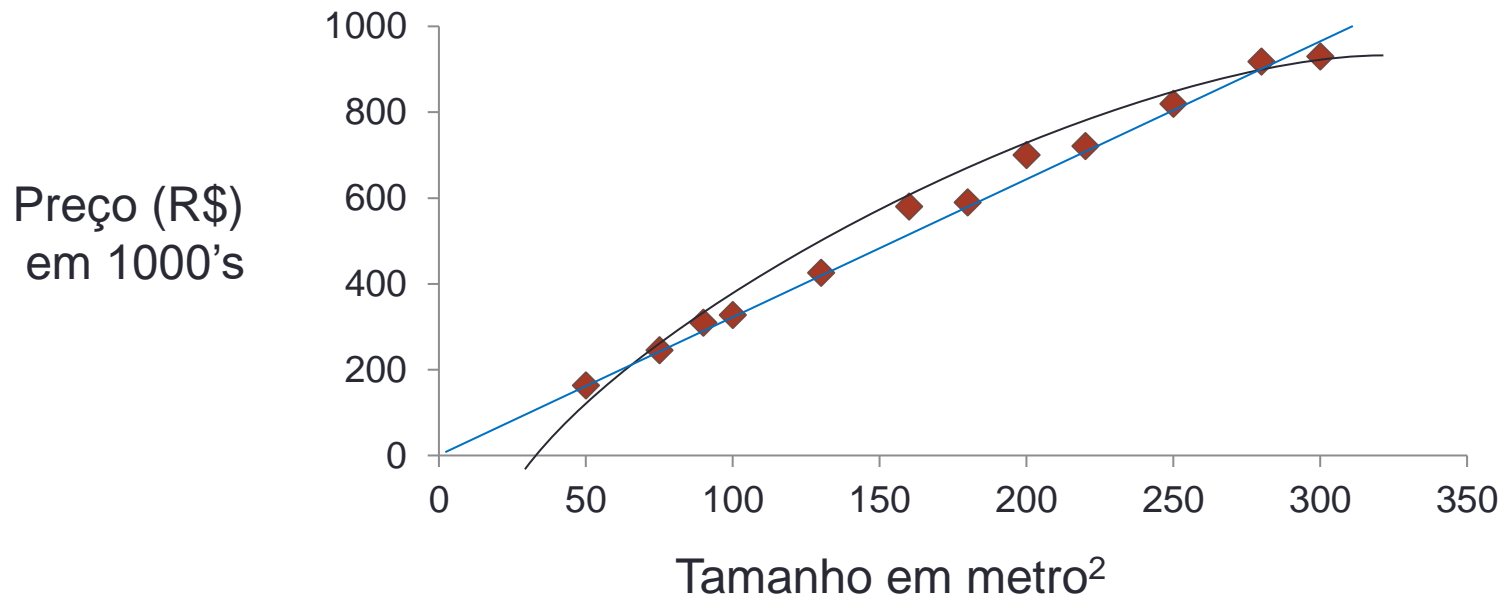
Aprendizado Supervisionado: Regressão

Prever o Preço de Imóveis



Aprendizado Supervisionado: Regressão

Prever o Preço de Imóveis



Aprendizado Supervisionado
“respostas certas” são dadas

Regressão: Prevê valores de
saída(output) contínuo - preço

Abordagens do Aprendizado Supervisionado

- Classificação:

- Responde se uma determinada “entrada” pertence a uma certa classe.
- Dada a imagem de uma fruta: que fruta é (dentro um número finito).

- Regressão:

- Faz uma predição a partir de exemplos.
- Prever o valor dos imóveis, dados os valores por metro quadrado.
 - Regressão Linear Simples
 - Regressão Linear Múltipla
 - Regressão Não Linear/Polinomial (Simples e Múltipla)
 - Regressão com Vetores de Suporte (SVR)
 - ...

Regressão Linear

Regressão
Linear

```
graph TD; A[Regressão Linear] --> B[Simples]; A --> C[Múltipla]
```

Simples

Múltipla

1. Análise de Correlação

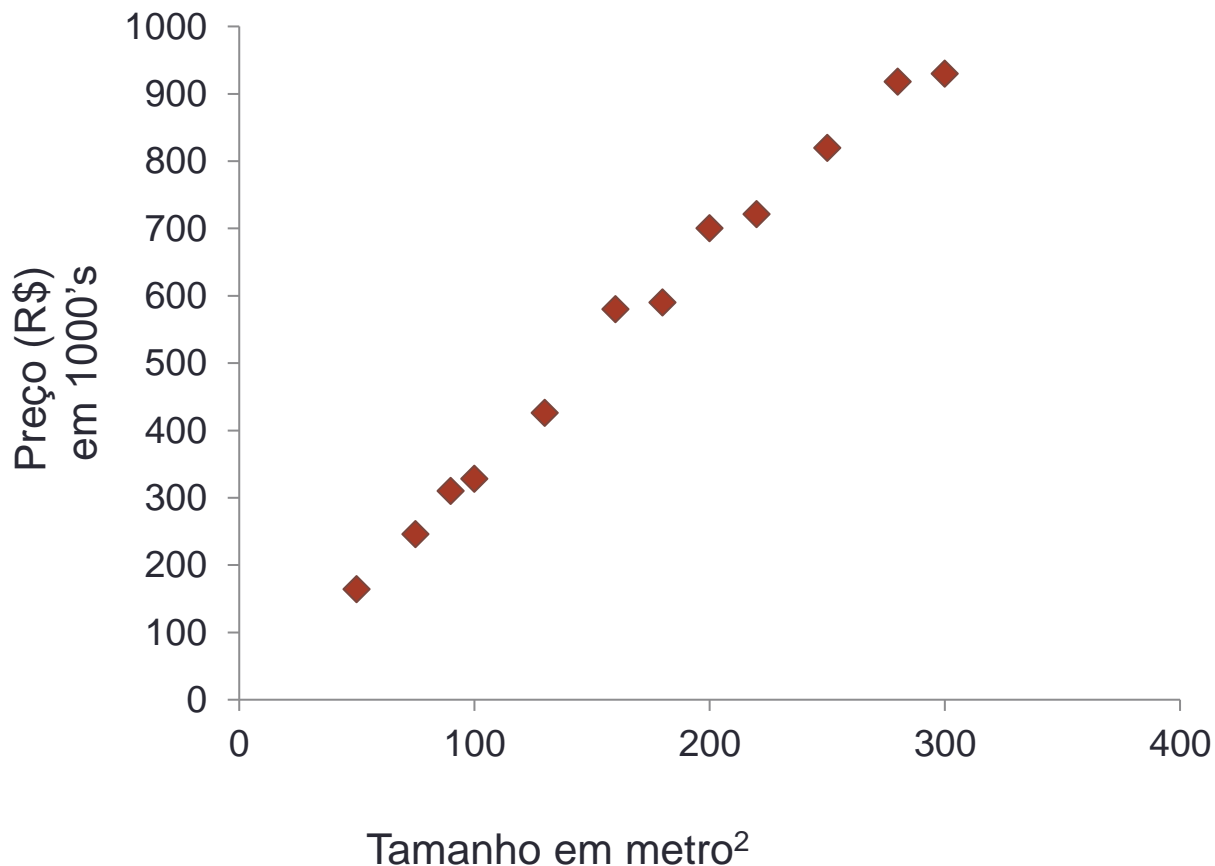
1. Análise de Correlação

Duas variáveis estão relacionadas se a mudança de uma provoca a mudança na outra.

- Exemplo: Tamanho em m^2 x Preço da Casa

1. Análise de Correlação

Tamanho em m ²	Preço (R\$) em 1000's
50	164
75	246
90	310
100	328
130	426
160	580
180	590
200	700
220	721
250	820
280	918
300	930



1. Análise de Correlação

Duas variáveis estão relacionadas se a mudança de uma provoca a mudança na outra.

- Exemplo: Tamanho em m² x Preço da Casa

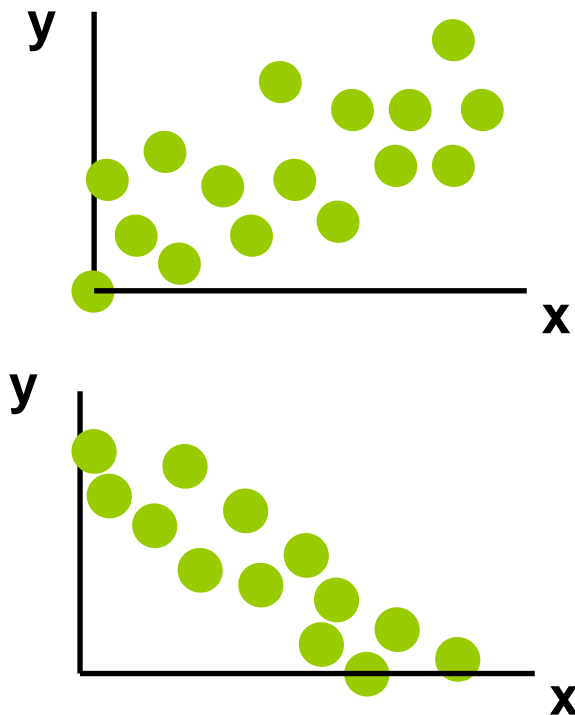
Correlação:

- É utilizada para medir o quanto uma variável está associada a outra.

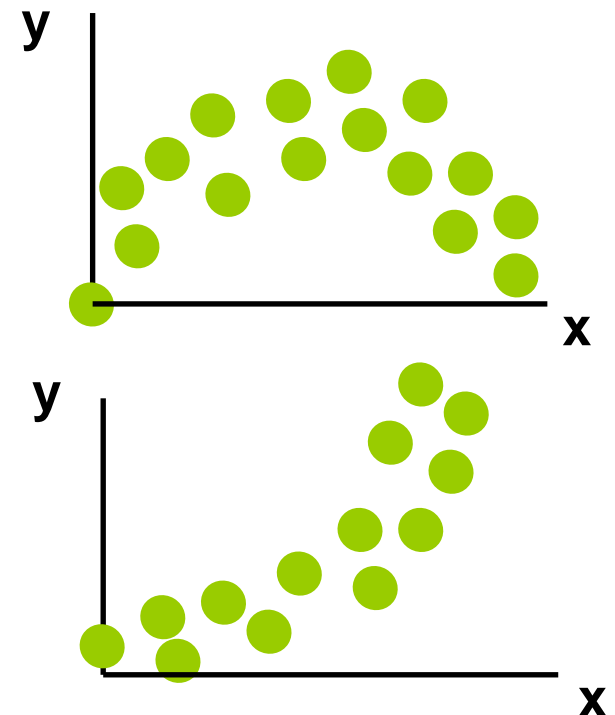
1. Análise de Correlação

Gráfico (Diagrama) de Dispersão: usado para mostrar a relação entre duas variáveis quantitativas, medidas sobre os mesmos indivíduos.

Relação Linear



Relação Curvilinear



1. Análise de Correlação

Duas variáveis estão relacionadas se a mudança de uma provoca a mudança na outra.

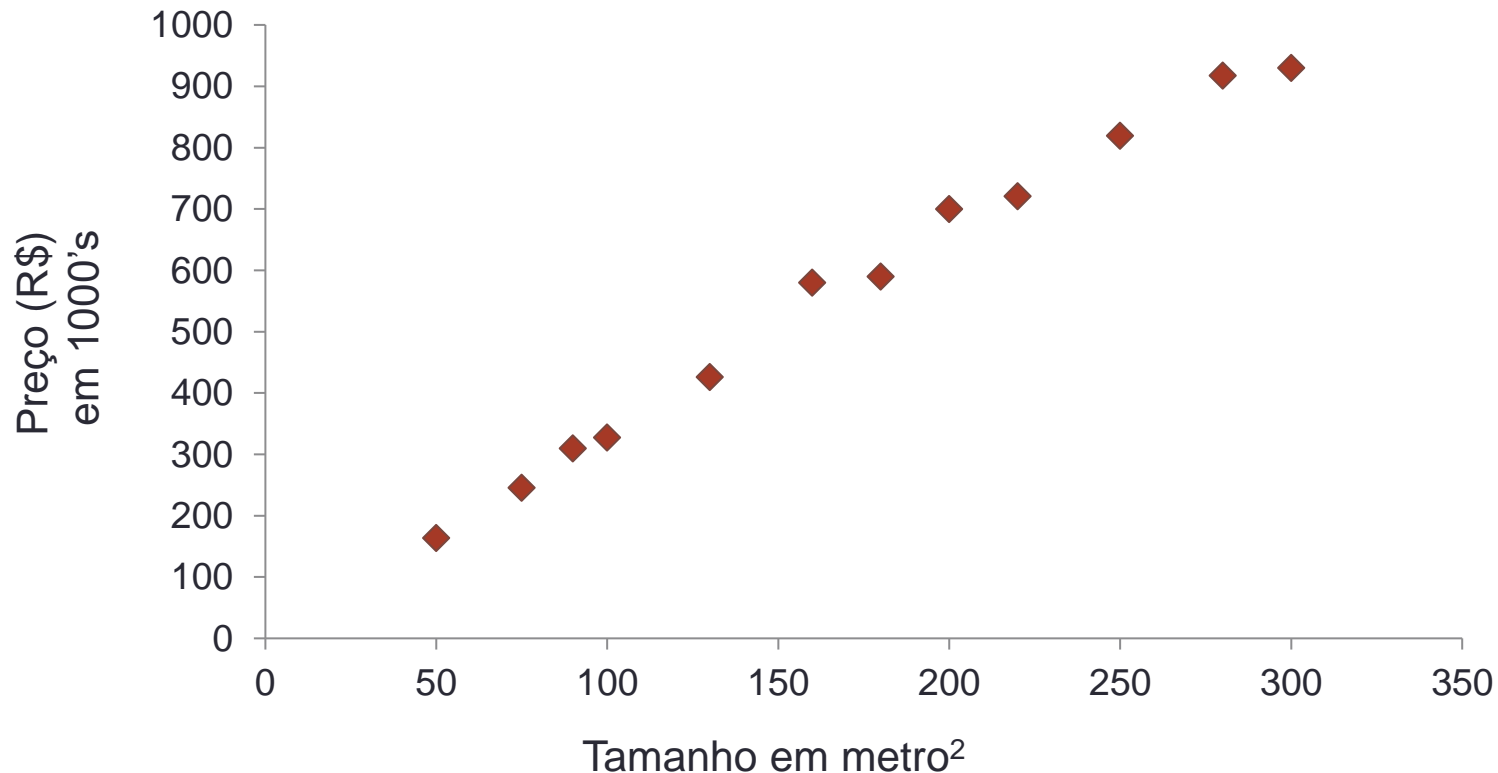
- Exemplo: Tamanho em m^2 x Preço da Casa

Correlação:

- É utilizada para medir o quanto uma variável está associada a outra.
- Quando a alteração no valor de uma variável **(independente (x) – Tamanho em m^2)** provoca alterações no valor da outra variável **(dependente (y) - Preço da Casa)**

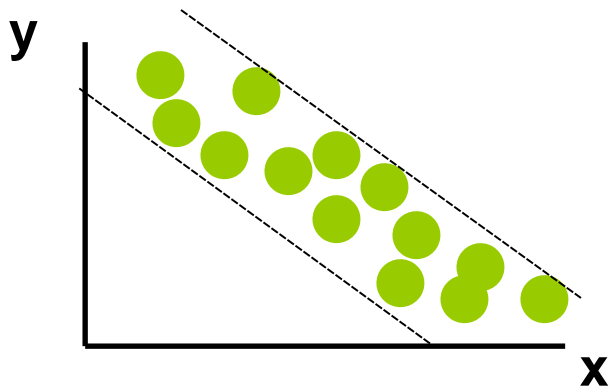
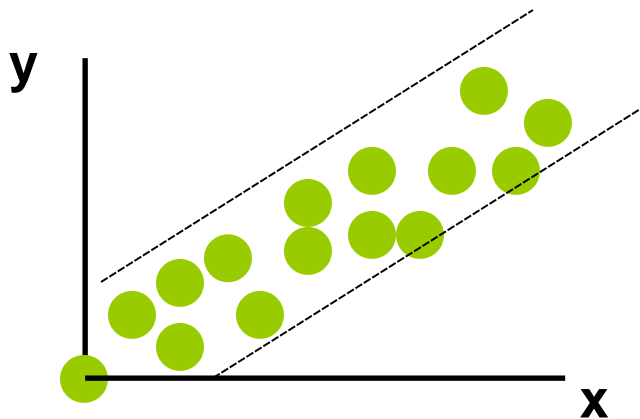
1. Análise de Correlação

- **Eixo x – Tamanho:** variável independente
- **Eixo y – Preço:** variável dependente (muda de acordo com as mudanças na variável x)

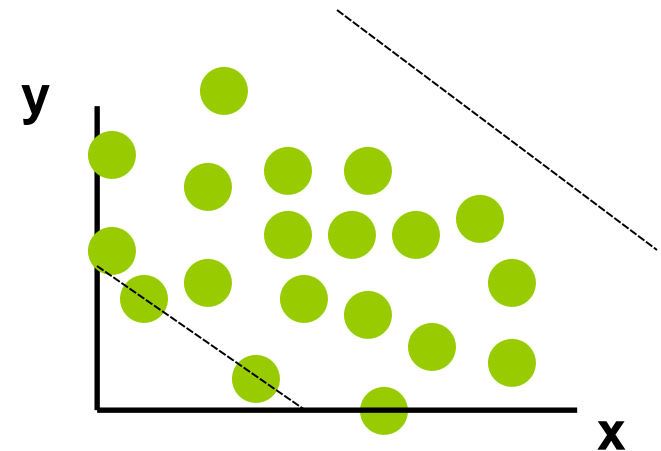
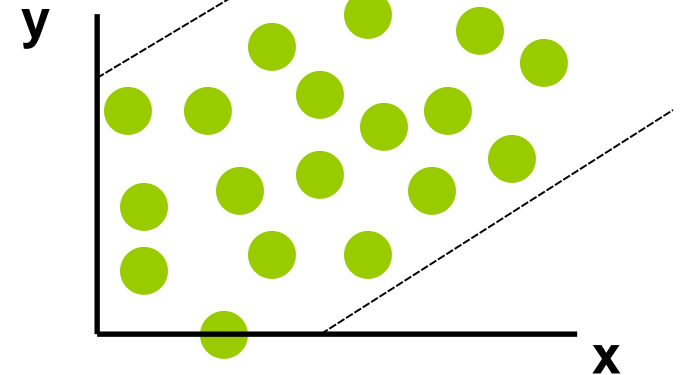


Exemplos de Gráfico de Dispersão

Relações Fortes

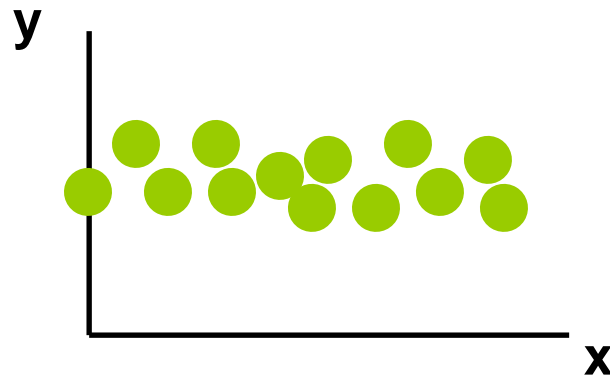
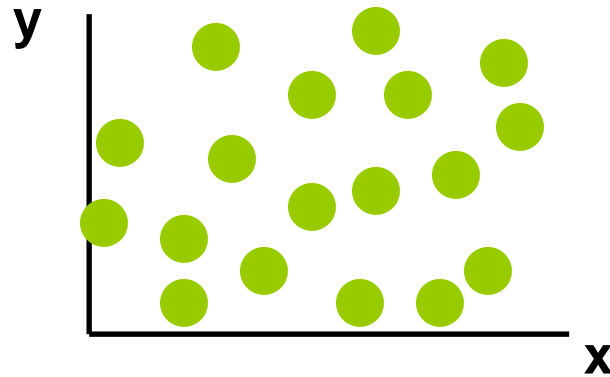


Relações Fracas



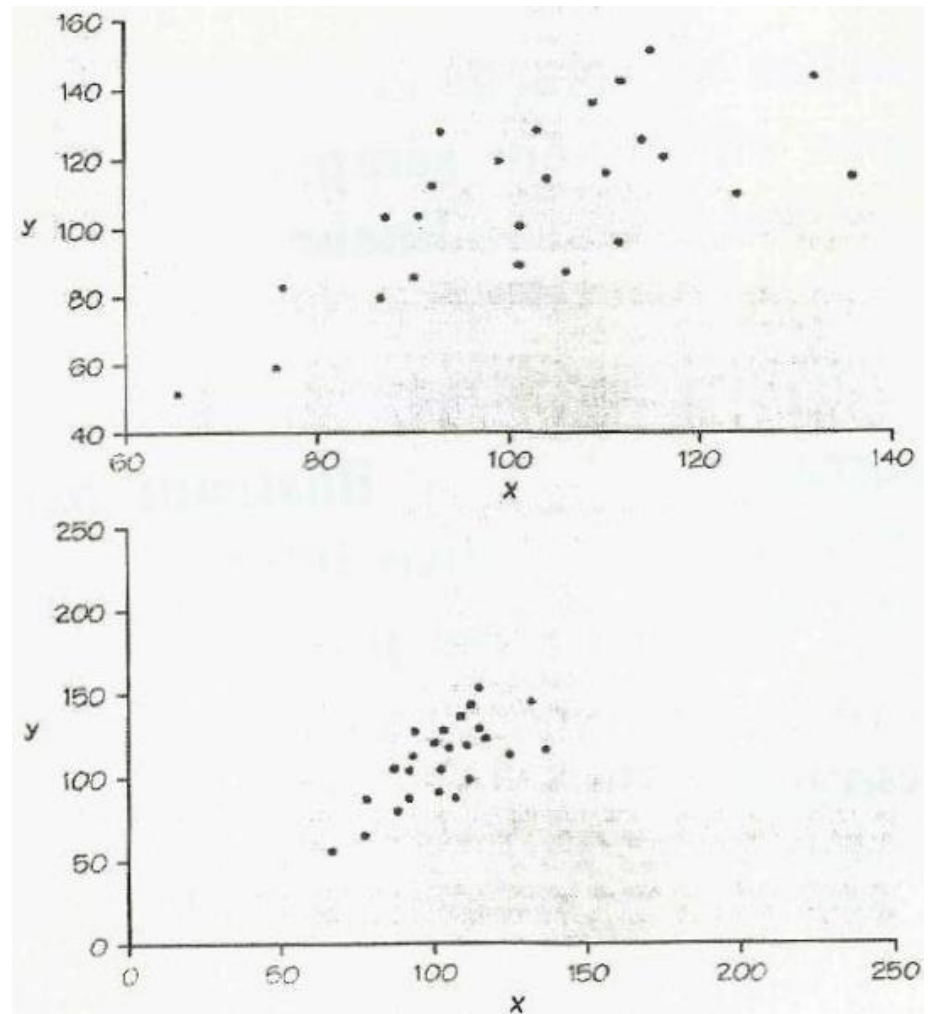
Exemplos de Gráfico de Dispersão

Nenhuma Relação



Alguns problemas da análise gráfica

- Nem sempre conseguimos ver exatamente a intensidade de uma relação linear.
- Gráfico ao lado: mesmos dados, porém em uma escala diversa.
- Para este problema utilizamos uma medida numérica: **Coeficiente de Correlação**



Coeficiente de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

r = mede o grau de relacionamento linear entre valores x e y , isto é, o **Coeficiente de Correlação**.

Mede a intensidade e a direção da relação linear entre duas variáveis quantitativas.

x = variável independente

y = variável dependente

Coeficiente de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Soma ((x – média de x) * (y – média de y))

Coeficiente de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Tamanho (m2)	Preço	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
30	57.000			
39	69.000			
49	77.000			
60	90.000			
44,5 (média) 12,92 (desvio padrão)	73.350 (média) 13.865,42 (dp)			

Coeficiente de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Tamanho (m2)	Preço	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
30	57.000	-14,5		
39	69.000	-5,5		
49	77.000	4,5		
60	90.000	15,5		
44,5 (média) 12,92 (desvio padrão)	73.350 (média) 13.865,42 (dp)			

Coeficiente de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Tamanho (m2)	Preço	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
30	57.000	-14,5	-16.250	
39	69.000	-5,5	-4.250	
49	77.000	4,5	3.750	
60	90.000	15,5	16.750	
44,5 (média) 12,92 (desvio padrão)	73.350 (média) 13.865,42 (dp)			

Coeficiente de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Tamanho (m2)	Preço	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$
30	57.000	-14,5	-16.250	235.625
39	69.000	-5,5	-4.250	23.375
49	77.000	4,5	3.750	16.875
60	90.000	15,5	16.750	259.625
44,5 (média) 12,92 (desvio padrão)	73.350 (média) 13.865,42 (dp)			535.500 (soma)

Regressão Linear e Análise de Correlação

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{[\sum (x - \bar{x})^2][\sum (y - \bar{y})^2]}}$$

Soma ((x – média de x) * (y – média de y))

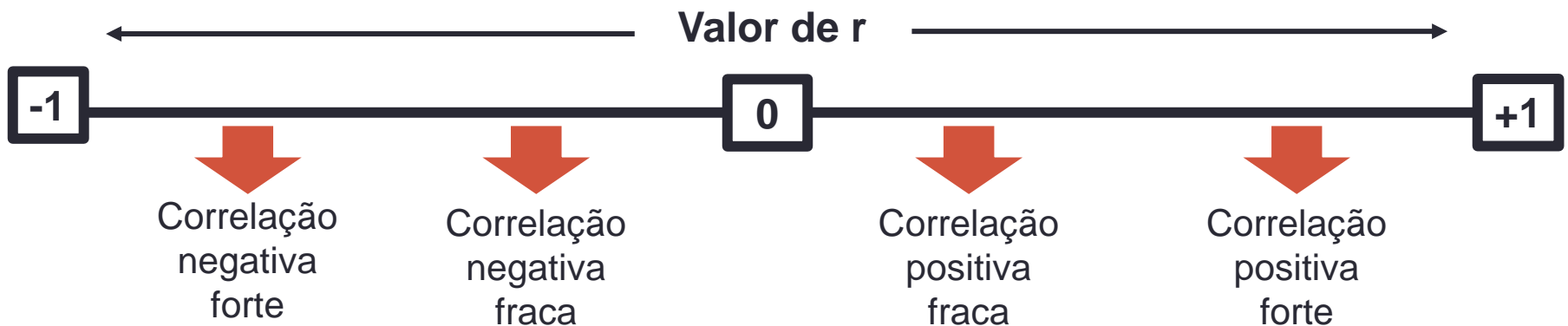
- Dividido -

Raiz quadrada ((soma de (x – média de x)²) *
(soma de (y – média de y)²))

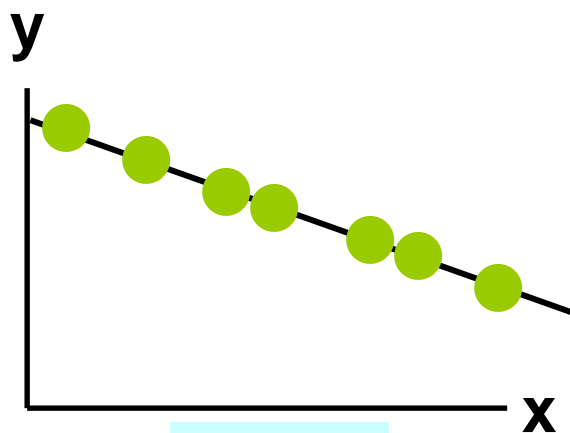
Análise de Correlação

'r' será um valor entre
-1 e 1

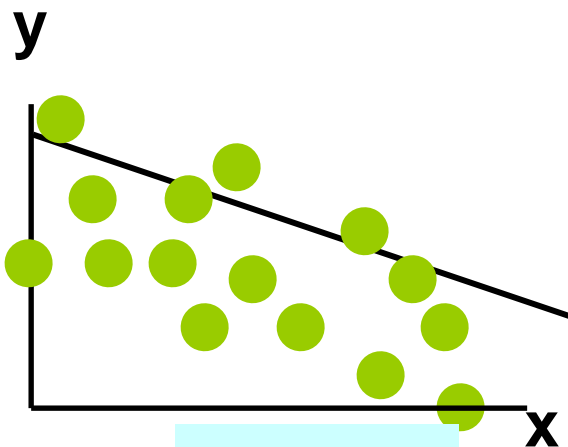
- Quanto mais próximo de -1 : maior correlação negativa
- Quanto mais próximo de 1 : maior correlação positiva
- Quanto mais próximo de 0 : menor a correlação linear



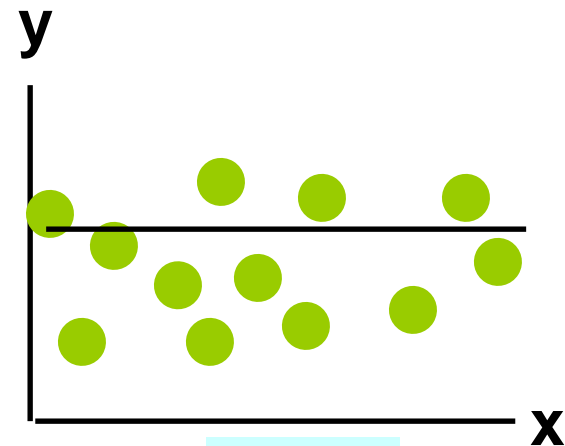
Análise de Correlação: Gráfico de Dispersão



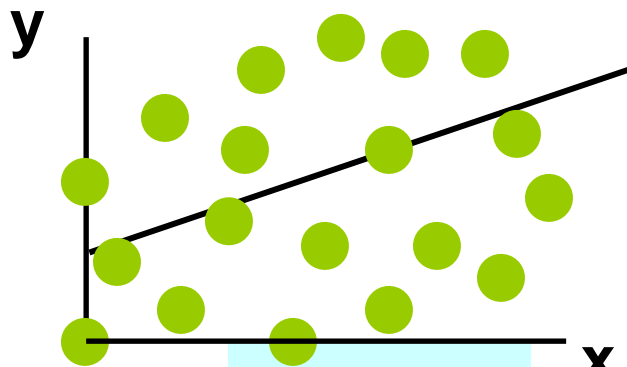
$r = -1$



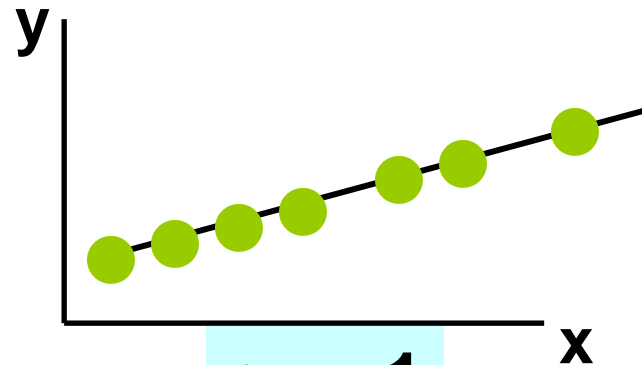
$r = -0.6$



$r = 0$



$r = +0.3$

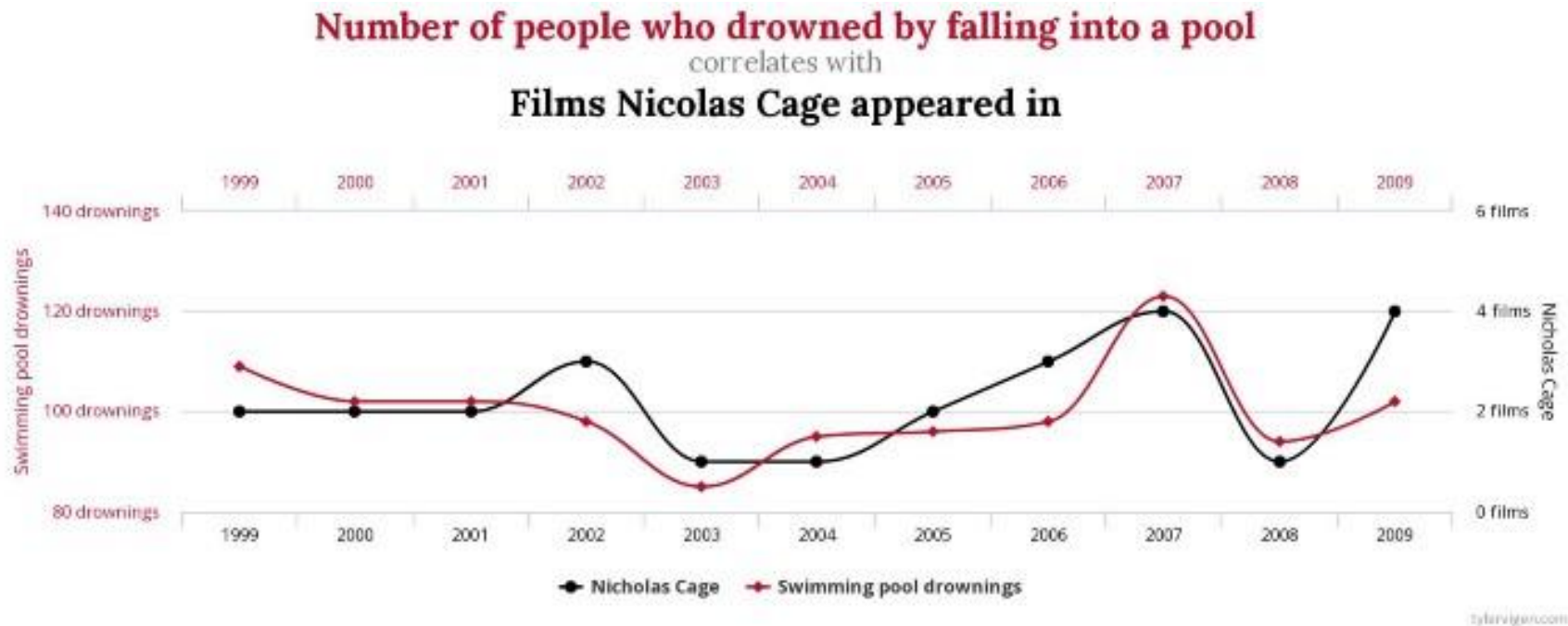


$r = +1$

Análise de Correlação

Correlação	Interpretação
0,00 a 0,19 ou 0,00 a -0,19	Correlação bem fraca
0,20 a 0,39 ou -0,20 a -0,39	Correlação fraca
0,40 a 0,69 ou -0,40 a -0,69	Correlação moderada
0,70 a 0,89 ou -0,70 a -0,89	Correlação forte
0,90 a 1,00 ou -0,90 a -1,00	Correlação muito forte

Correlação Não é Causa



2. Regressão Linear

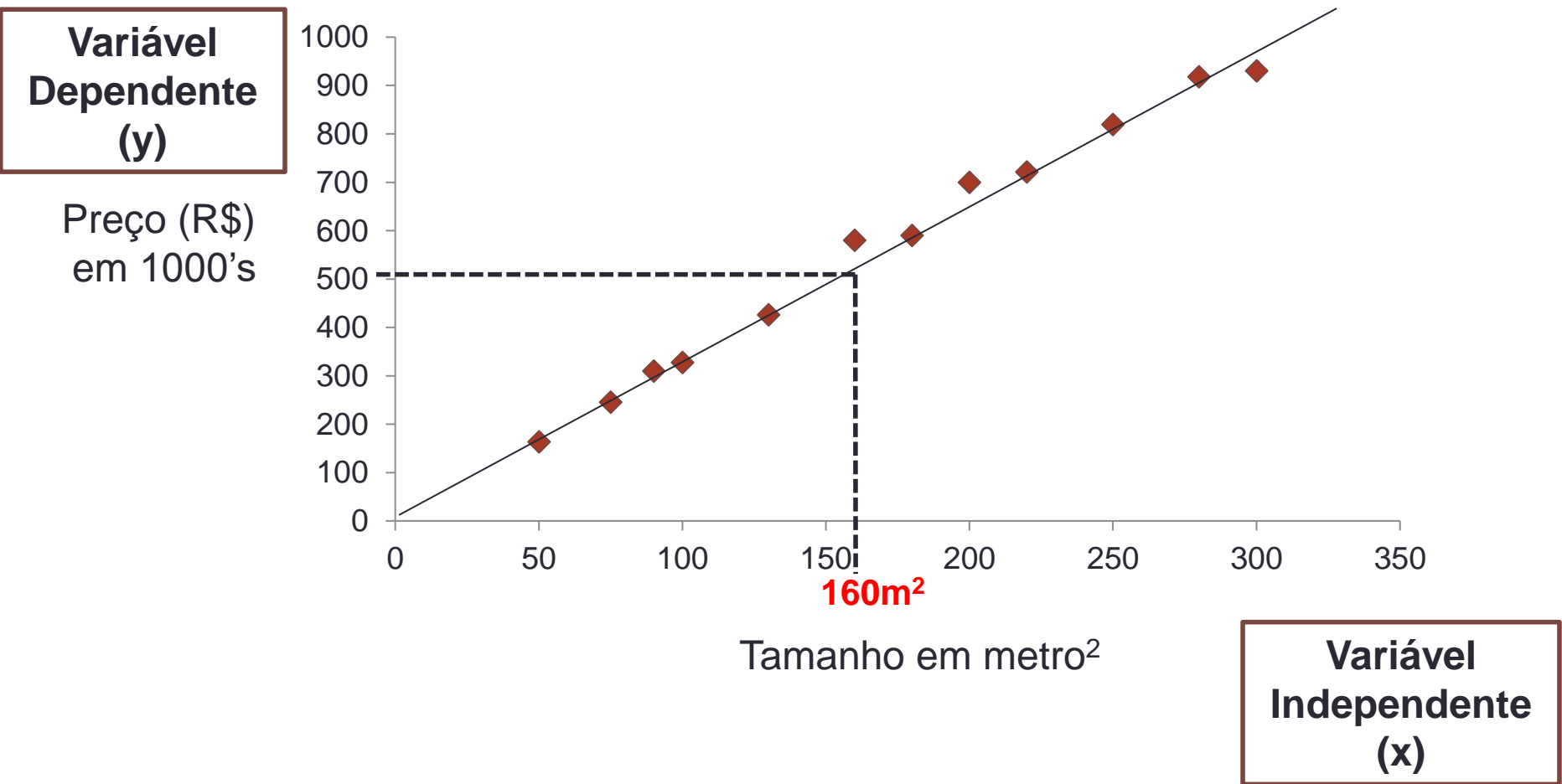
2. Regressão Linear e Análise de Correlação

Correlação e regressão estão intimamente relacionados.

- A **Correlação** resume as relações entre 2 variáveis.
- A **Regressão** é utilizada para prever os valores de uma variável dados os valores da outra.
 - **Prever** o valor de uma variável dependente com base no valor de, pelo menos, uma variável independente.
 - **Explicar o impacto das mudanças** em uma variável independente (x) sobre a variável dependente (y).

2. Regressão Linear e Análise de Correlação

Prever o Preço de Imóveis



2. Regressão Linear

Diagram illustrating the components of the linear regression equation:

Labels and their corresponding parts in the equation $y = \beta_0 + \beta_1 x$:

- Variável Dependente (previsão) points to y .
- Ponto onde a reta intercepta o eixo y points to β_0 .
- Coeficiente Angular points to β_1 .
- Variável Independente points to x .

Onde,

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2}$$

Modelo de Regressão Linear

