
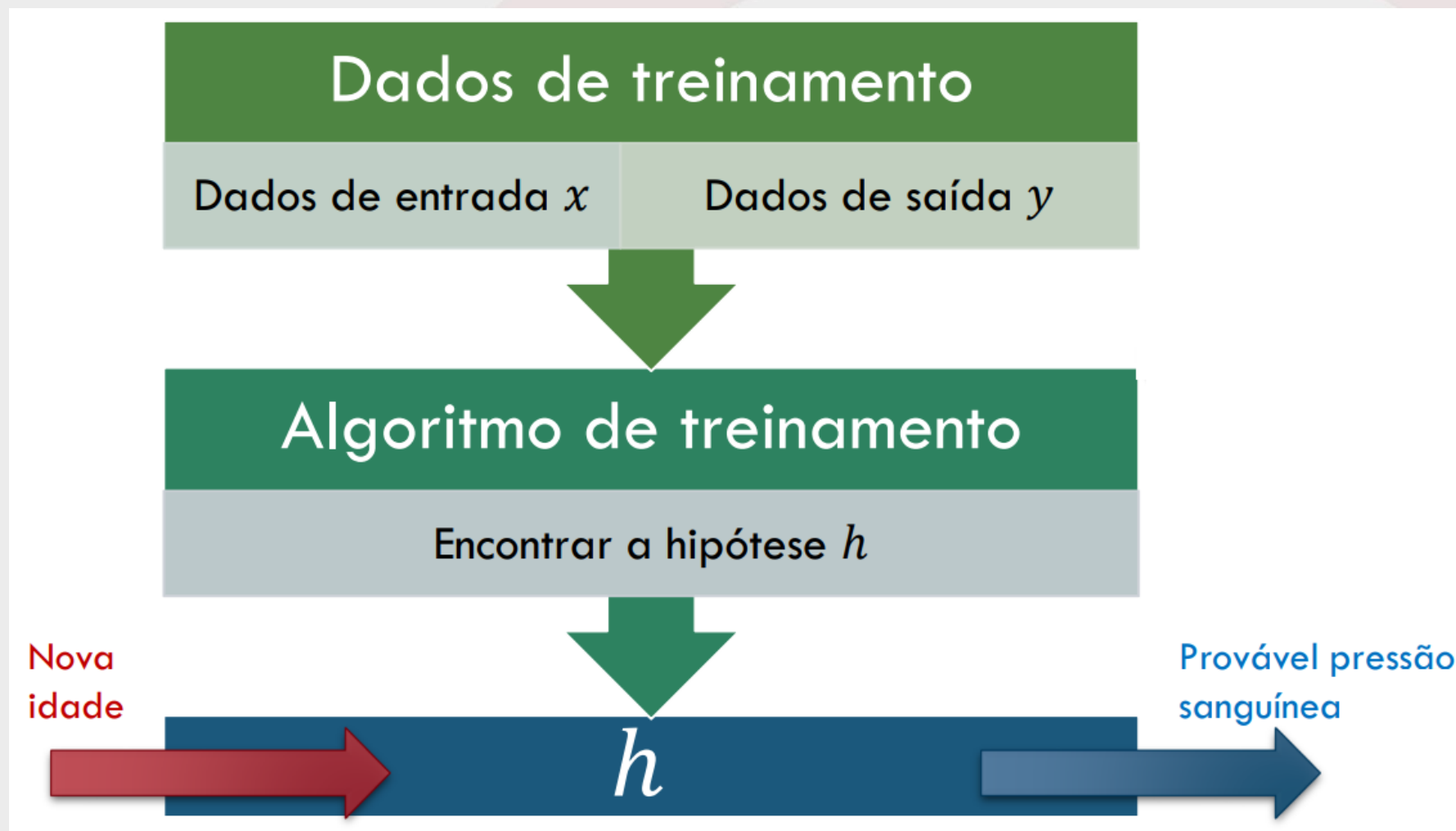


Regressão

- Regressão refere-se a prever a saída de uma variável numérica (dependente) a partir de um conjunto de uma ou mais variáveis independentes.
- Uma equação de regressão é usada em estatística para descobrir qual relação existe, e se existe, entre conjuntos de dados.
- Em Aprendizado de Máquina, essa equação é obtida através de um algoritmo de treinamento utilizando os m dados $x(i)$, $y(i)$.

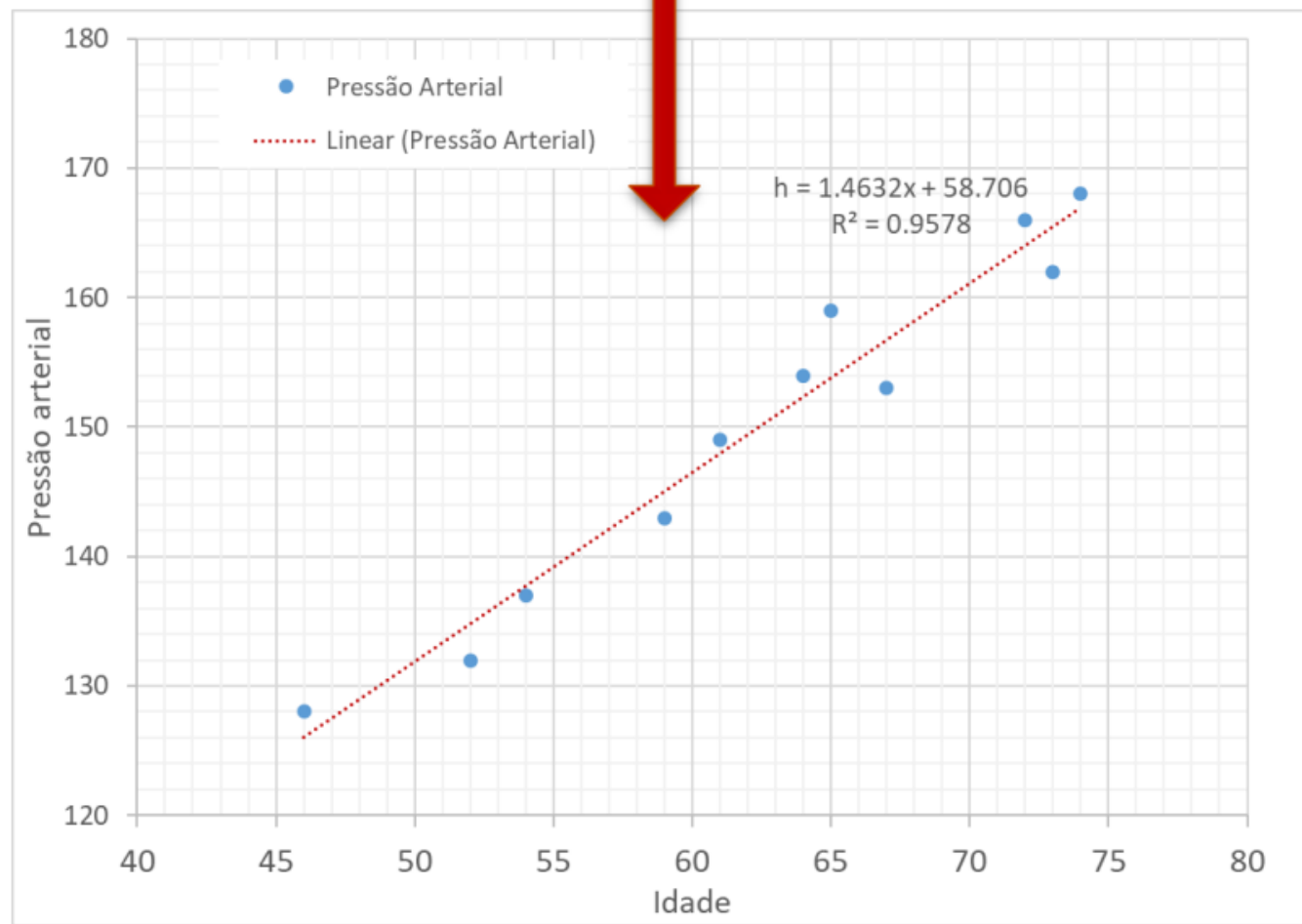
FEATURES		TARGET
		
x: Variável independente		y: variável dependente
Idade	Peso (Kg)	Pressão Arterial
52	78.5	132
59	83.5	143
67	88.0	153
73	95.7	162
64	88.9	154
74	99.8	168
54	85.3	137
61	85.3	149
65	93.9	159
46	75.7	128
72	98.4	166

Regressão



Regressão

Hipótese h mapeia x em y .

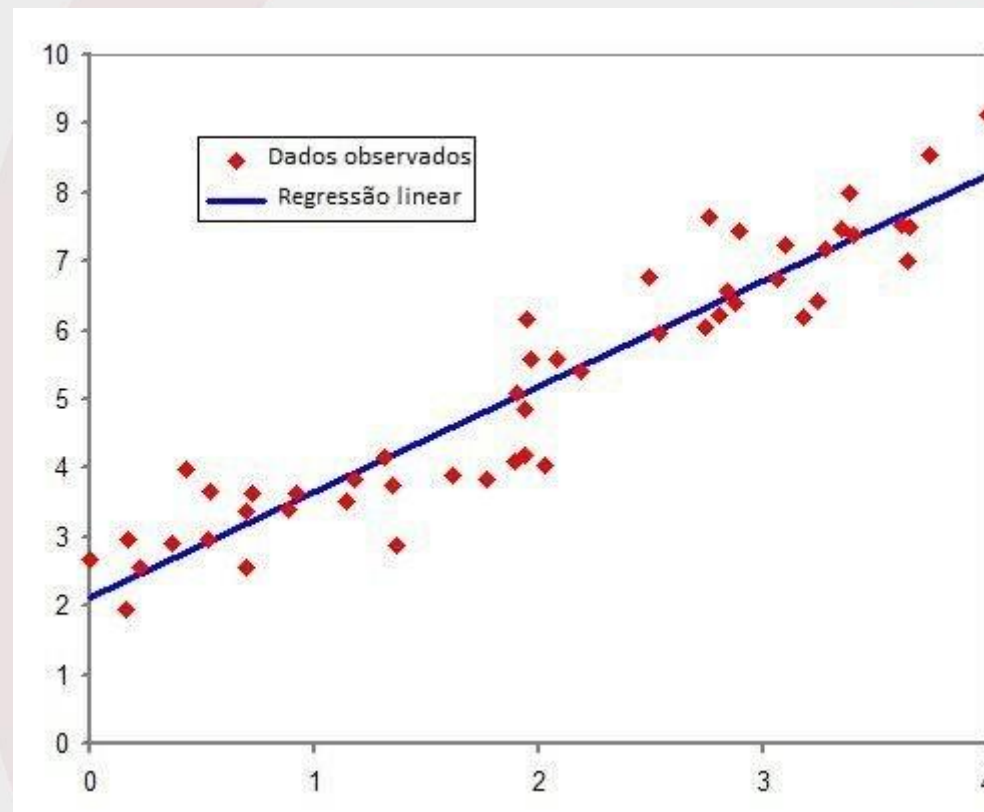


Regressão Linear

- É denominada dessa forma por ser uma reta traçada a partir de uma relação em um diagrama de dispersão.
- Essa reta resume uma relação entre os dados de duas variáveis e também **pode ser utilizada para realizar previsões**.
- Sua origem vem da correlação linear, que é a verificação da existência de um relacionamento entre duas variáveis.
 - Ou seja, **dado X e Y, quanto que X explica Y**.
- O resultado da regressão linear é sempre um número.

Regressão Linear

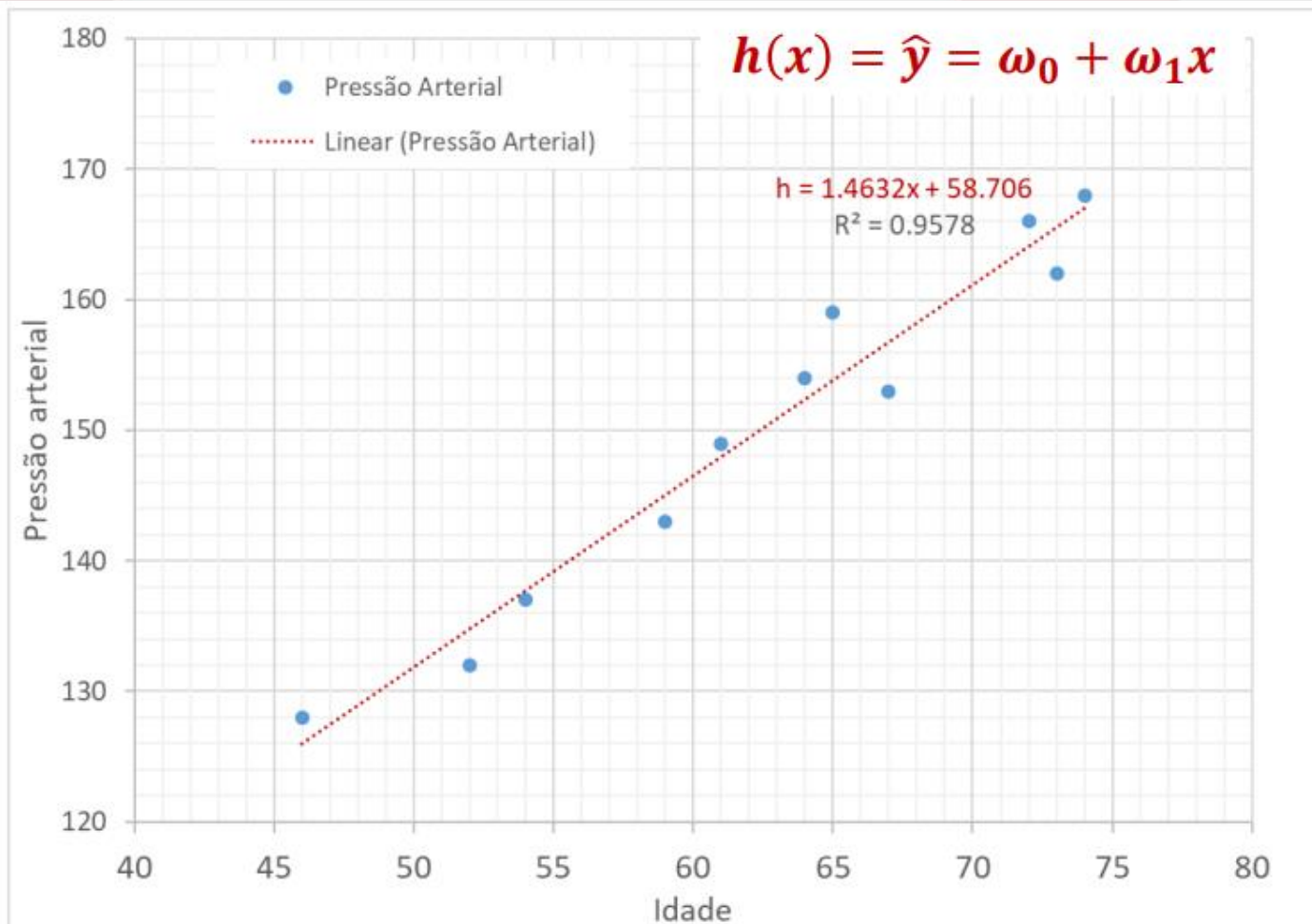
- É utilizada adequadamente quando o dataset apresenta algum tipo de tendência de crescimento/descrescimento constante.



Regressão Linear

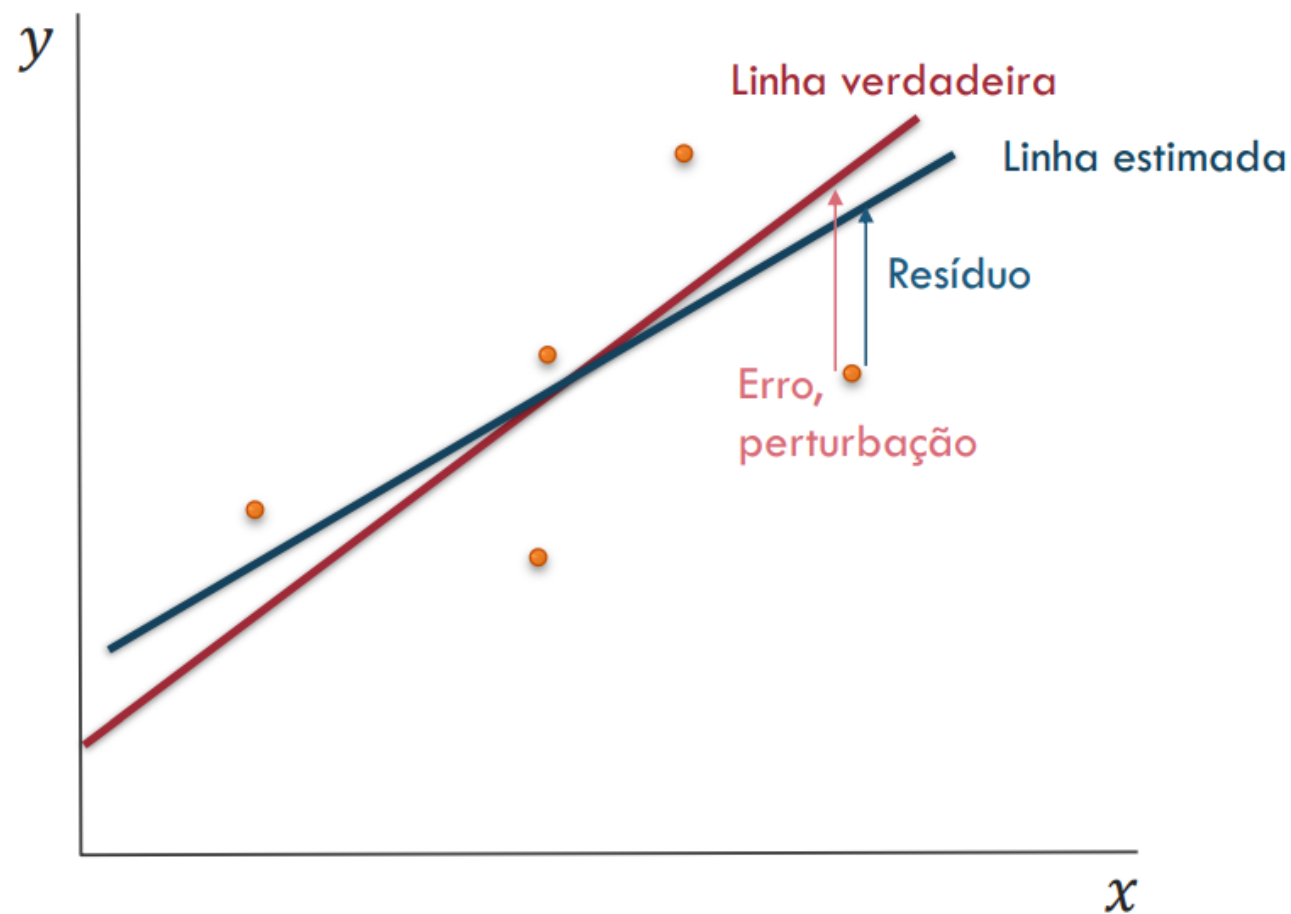
Tenho $m = 11$ dados de treinamento (\bullet), onde x é a variável única de entrada ($n = 1$) e y é a variável de saída.

m	Idade (variável de entrada)	Pressão Arterial (variável de saída)
1	52	132
2	59	143
3	67	153
4	73	162
5	64	154
6	74	168
7	54	137
8	61	149
9	65	159
10	46	128
11	72	166



Regressão Linear

- Resíduo e Erro:



○ **resíduo** é calculado após a execução do modelo de regressão e é a diferença entre os valores observados e os valores estimados.

○ **erro** do conjunto de dados é a diferença entre os valores observados e os valores verdadeiros, não observáveis.

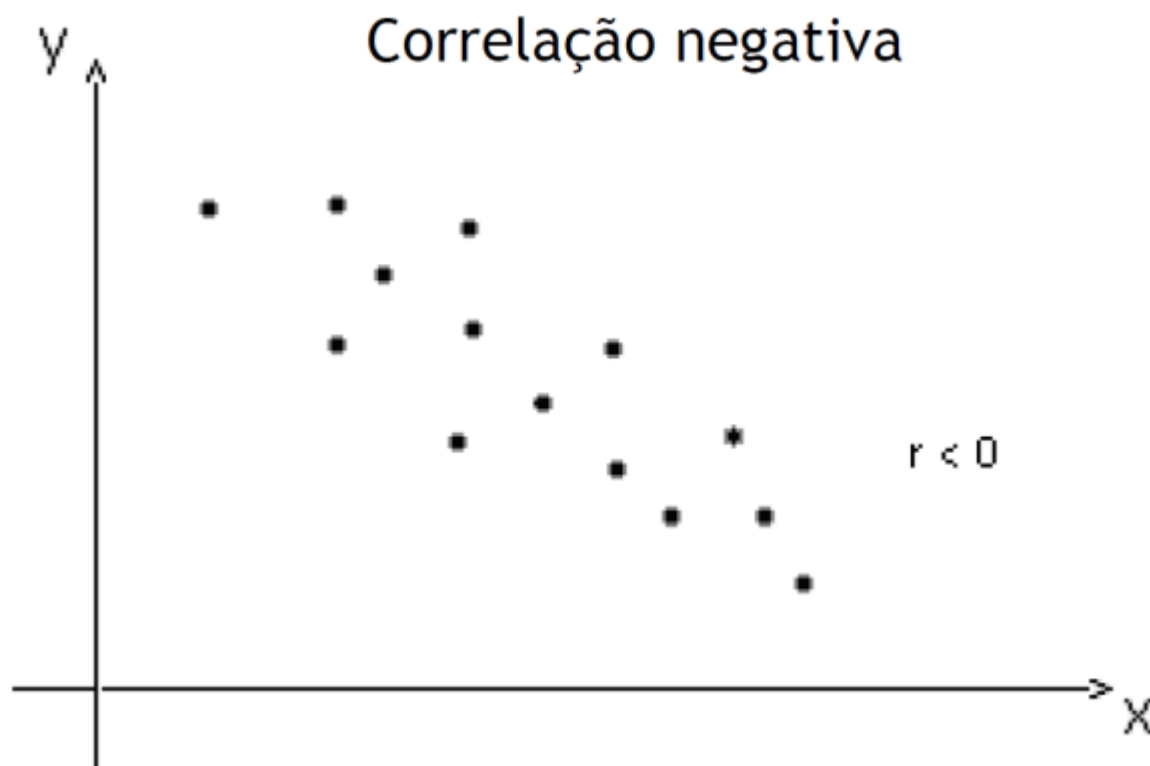
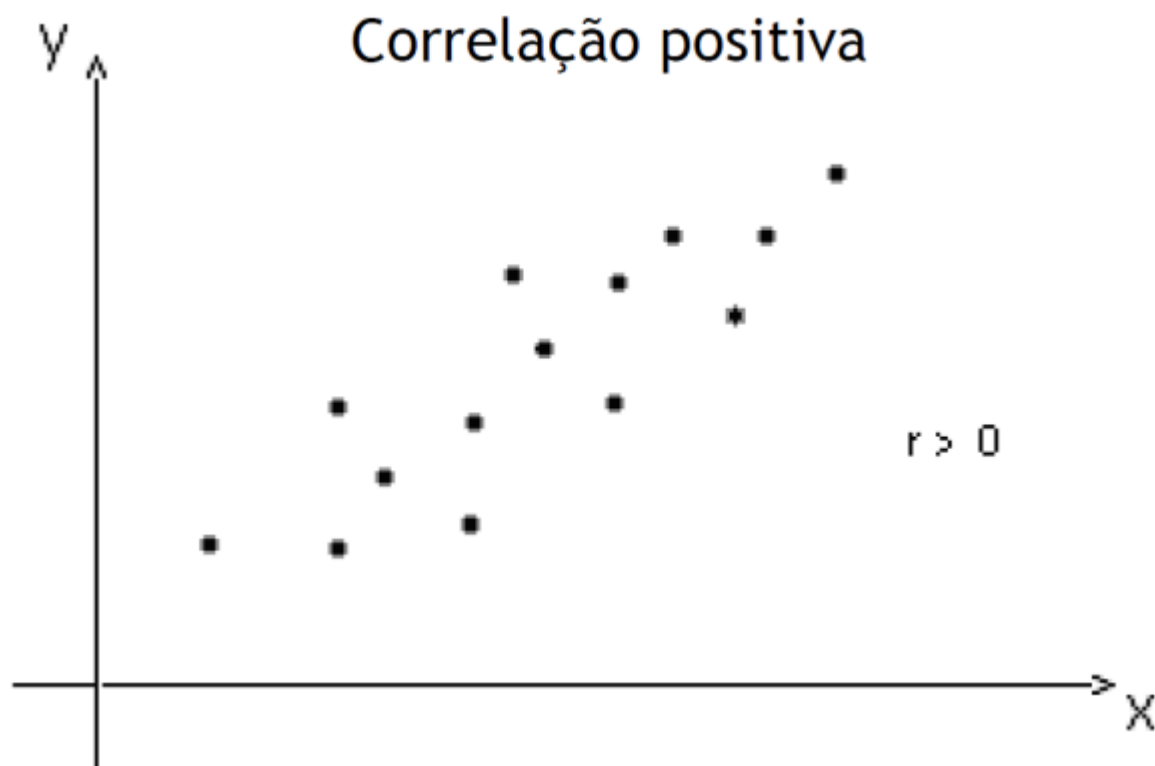
Regressão Linear

- A regressão linear pode ser de dois tipos:
 - **Regressão linear simples**, onde é utilizada apenas uma variável independente,
 - **Regressão linear múltipla**, onde múltiplas variáveis independentes são definidas.
- Para quantificar a relação entre duas variáveis quantitativas utiliza-se o coeficiente de correlação linear de Pearson:

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

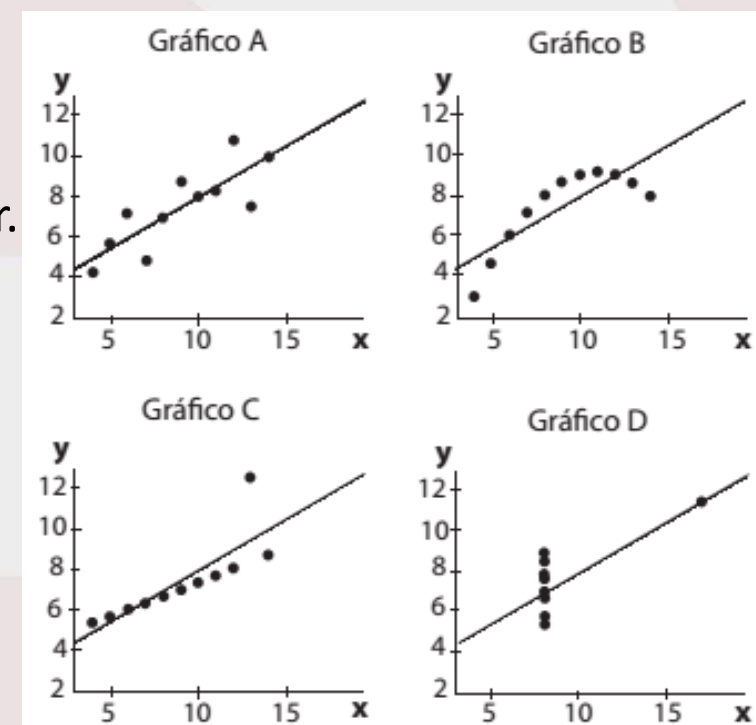
Coeficiente	Classificação
$0,9 < r \leq 1,0$	Ótima
$0,8 < r \leq 0,9$	Boa
$0,7 < r \leq 0,8$	Razoável
$0,6 < r \leq 0,7$	Medíocre
$0,5 < r \leq 0,6$	Péssima
$ r \leq 0,5$	Imprópria

Regressão Linear



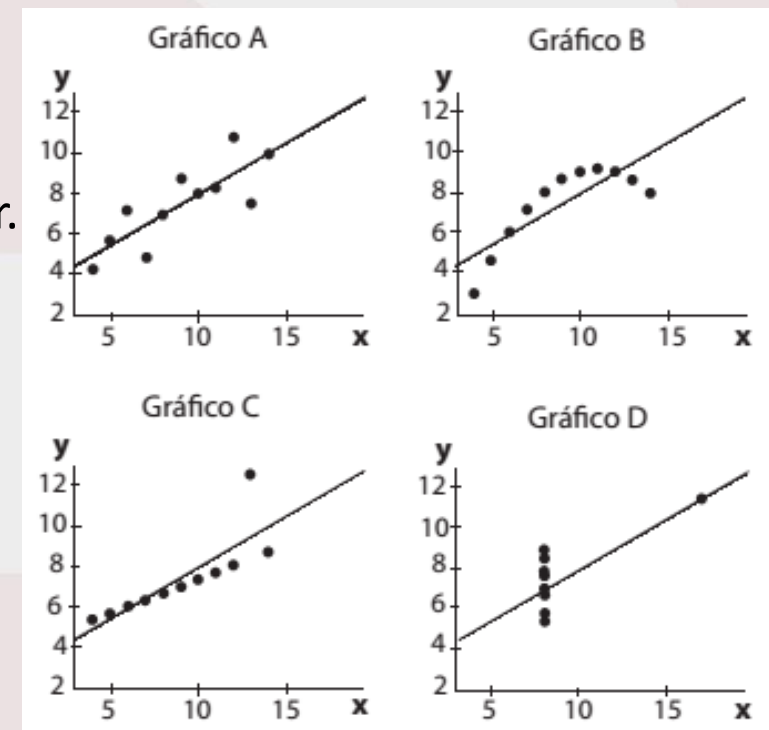
Regressão Linear

- Prova: CESGRANRIO - 2018 - Banco do Brasil - Escriturário
- Para ilustrar a importância da análise gráfica em análises de regressão linear, F. J. Anscombe produziu quatro conjuntos de pares (x, y) a partir das mesmas estatísticas suficientes, como: coeficientes linear e angular; soma dos quadrados dos resíduos e da regressão; e número de observações. Os diagramas de dispersão para as quatro bases de dados, juntamente com a reta da regressão ($y = 4 + 0,5x$), encontram-se abaixo.
- Com base nesses gráficos, considere as seguintes afirmativas:
- I – O gráfico B mostra um valor influente para gerar uma regressão linear.
- II – O gráfico C mostra uma possível observação outlier na regressão linear.
- III – O gráfico D mostra uma possível observação outlier na regressão linear.
- Está correto SOMENTE o que se afirma em
- A II e III
- B I e III
- C I
- D II
- E III



Regressão Linear

- Prova: CESGRANRIO - 2018 - Banco do Brasil - Escriturário
- Para ilustrar a importância da análise gráfica em análises de regressão linear, F. J. Anscombe produziu quatro conjuntos de pares (x, y) a partir das mesmas estatísticas suficientes, como: coeficientes linear e angular; soma dos quadrados dos resíduos e da regressão; e número de observações. Os diagramas de dispersão para as quatro bases de dados, juntamente com a reta da regressão ($y = 4 + 0,5x$), encontram-se abaixo.
- Com base nesses gráficos, considere as seguintes afirmativas:
- I – O gráfico B mostra um valor influente para gerar uma regressão linear.
- II – O gráfico C mostra uma possível observação outlier na regressão linear.
- III – O gráfico D mostra uma possível observação outlier na regressão linear.
- Está correto SOMENTE o que se afirma em
- A II e III
- B I e III
- C I
- **D II**
- E III



Regressão Linear

- Ano: 2013 Banca: Quadrix Órgão: DATAPREV
- Assinale a alternativa que contém uma justificativa para a utilização de um modelo de regressão linear múltipla em substituição a um modelo de regressão linear simples para a análise de dados.
- A Quando existe um número excessivo de dados a serem analisados.
- B Quando o resíduo é muito grande.
- C Quando a regressão é ausente.
- D Quando se necessita de mais de uma variável independente no modelo de regressão.
- E Quando a regressão é fracamente positiva.

Regressão Linear

- Ano: 2013 Banca: Quadrix Órgão: DATAPREV
- Assinale a alternativa que contém uma justificativa para a utilização de um modelo de regressão linear múltipla em substituição a um modelo de regressão linear simples para a análise de dados.
- A Quando existe um número excessivo de dados a serem analisados.
- B Quando o resíduo é muito grande.
- C Quando a regressão é ausente.
- D Quando se necessita de mais de uma variável independente no modelo de regressão.
- E Quando a regressão é fracamente positiva.

Regressão Linear

- Prova: FMP Concursos - 2012 - PROCEMPA - Analista de Logística
- Com um conjunto de dados, apresentados na forma de diagrama de dispersão, onde são identificadas duas variáveis (exemplo: renda e produtos vendidos), pode-se utilizar as seguintes ferramentas estatísticas:
- I) regressão linear simples.
- II) correlação.
- III) regressão linear múltipla.
- As alternativas corretas são:
- A apenas I e II.
- B apenas I e III.
- C apenas II e III.
- D I, II e III.
- E Nenhuma das alternativas está correta.

Regressão Linear

- Prova: FMP Concursos - 2012 - PROCEMPA - Analista de Logística
- Com um conjunto de dados, apresentados na forma de diagrama de dispersão, onde são identificadas duas variáveis (exemplo: renda e produtos vendidos), pode-se utilizar as seguintes ferramentas estatísticas:
- I) regressão linear simples.
- II) correlação.
- III) regressão linear múltipla.
- As alternativas corretas são:
- A apenas I e II.
- B apenas I e III.
- C apenas II e III.
- D I, II e III.
- E Nenhuma das alternativas está correta.