

Regressão

Objetivo

Estudar a relação entre duas variáveis quantitativas.

Exemplos:

Idade e altura das crianças

Tempo de prática de esportes e ritmo cardíaco

Tempo de estudo e nota na prova

Taxa de desemprego e taxa de criminalidade

Expectativa de vida e taxa de analfabetismo

A presença ou ausência de **relação linear** pode ser investigada sob dois pontos de vista:

a) Quantificando a força dessa relação:

correlação.

b) Explicitando a forma dessa relação:

regressão.

Representação gráfica de duas variáveis quantitativas: **Diagrama de dispersão**

Exemplo 1: Um psicólogo está investigando a relação entre o tempo que um indivíduo leva para reagir a um estímulo visual (Y) com o sexo (W), idade (X) e acuidade visual (Z, medida em porcentagem).

X : idade

Y : tempo de reação

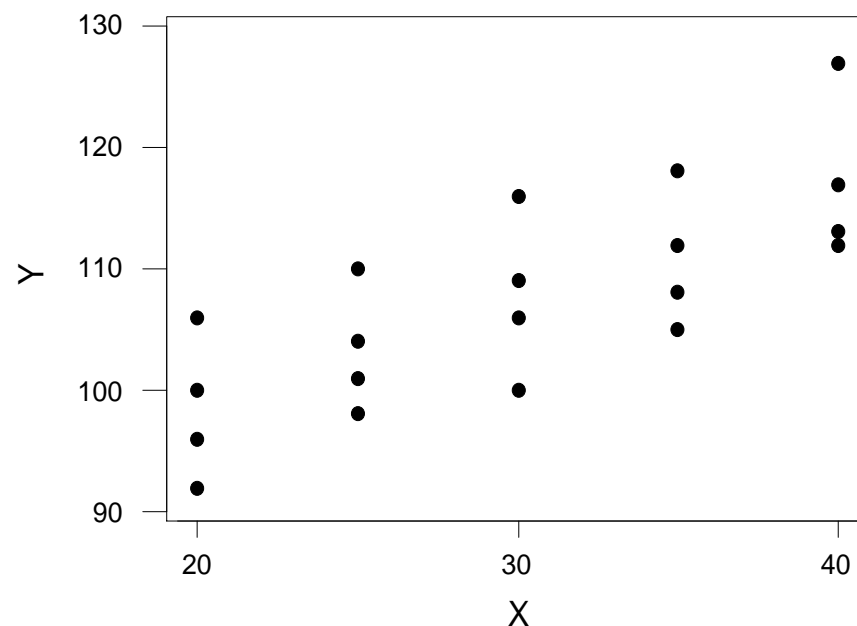
Analizamos utilizando uma ANOVA



Pelo gráfico: média de Y aumenta conforme as pessoas envelhecem



Modelo de regressão



Correlação entre Y e X = 0,768

Análise de regressão: metodologia estatística que estuda (modela) a relação entre duas ou mais variáveis

1. Tempo de reação \Rightarrow variável dependente ou resposta
idade \Rightarrow variável independente



modelo de regressão linear simples

2. Tempo de reação \Rightarrow variável dependente ou resposta
sexo, idade, acuidade visual \Rightarrow var. independentes



modelo de regressão linear múltipla

Modelo de regressão linear simples:

$$y_i = \alpha + \beta x_i + e_i, \quad i=1, \dots, n$$

em que

y_i : valor da variável resposta para o i -ésimo elemento da amostra

x_i : valor (conhecido) da variável independente ou preditora para o i -ésimo elemento da amostra

α e β são parâmetros desconhecidos

e_i : erro aleatório (independentes com média 0 e variância σ^2)

Dessa forma, fixado x_i , as variáveis y_i são independentes e $y_i \sim N(\alpha + \beta x_i, \sigma^2)$

Interpretação dos parâmetros:

α : (intercepto) valor esperado para a variável dependente y_i quando x_i é igual a zero

β : (coeficiente angular) variação esperada na variável resposta, quando a variável independente aumenta uma unidade

Estimação de α e $\beta \Rightarrow$ Método de mínimos quadrados

minimizar a soma dos quadrados dos erros

$$SQ(\alpha, \beta) = \sum (y_i - \alpha - \beta x_i)^2$$

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \quad \text{e} \quad \hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$

Reta ajustada: $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$

No exemplo:

$$n=20, \Sigma y_i= 2150, \Sigma x_i=600, \Sigma x_i y_i=65400, \Sigma x_i^2=19000$$

$$\hat{\beta} = \frac{65400 - 20 \cdot 30 \cdot 107,5}{11000 - 20 \cdot 30^2} = 0,90$$

$$\hat{\alpha} = 107,50 - 0,90 \cdot 30 = 80,50$$

$$\hat{y}_i = 80,50 + 0,90 x_i$$

Interpretação: Para um aumento de 1 ano na idade, o tempo médio de reação aumenta 0,90.

Podemos prever, por exemplo, o tempo médio de reação para pessoas de 20 anos $\Rightarrow \hat{y}(20) = 80,50 + 0,90 \cdot 20 = 98,50$

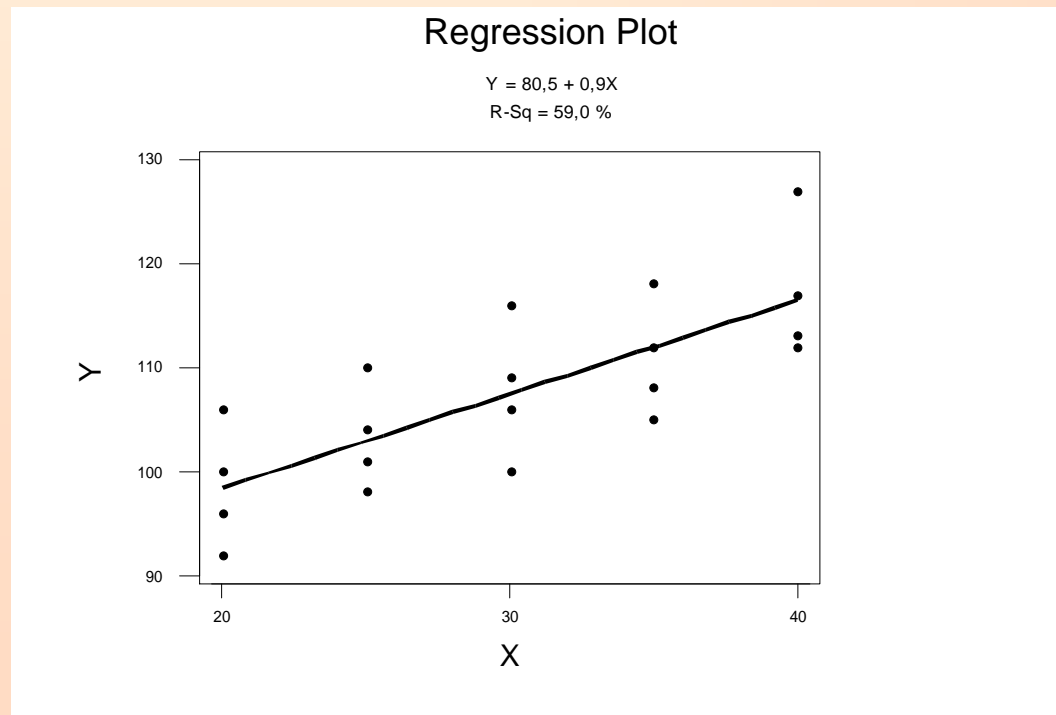
$\hat{y}(25) = 103$ $\hat{y}(30) = 107,50$ $\hat{y}(35) = 112$ $\hat{y}(40) = 116,5$
Valores próximos aos obtidos pela análise de variância



Vantagem: permite estimar o tempo médio de reação para idades não observadas



$$\hat{y}(33) = 80,50 + 0,90.33 = 110,20$$



Resíduos

Resíduo é a diferença entre o valor observado e o valor ajustado pela reta, isto é, $Y - \hat{Y}$

Para verificar a adequação do ajuste deve-se fazer uma *análise dos resíduos*.

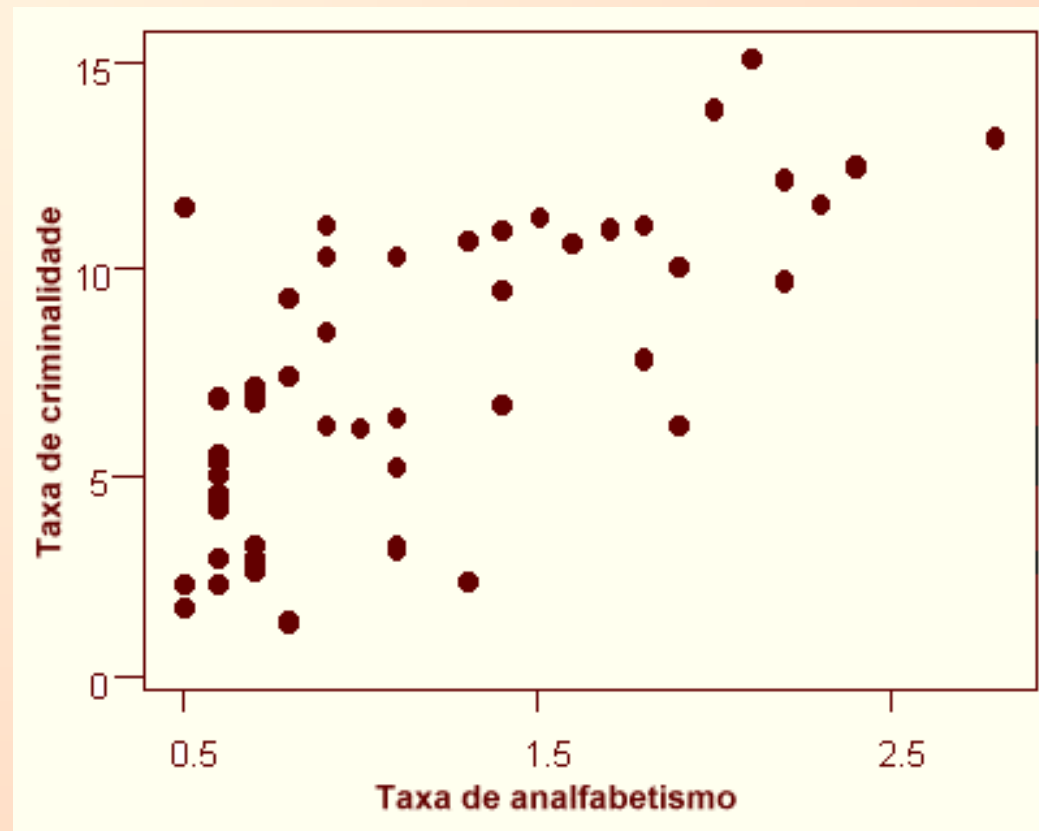
Exemplo 2: criminalidade e analfabetismo

Considere as duas variáveis observadas em 50 estados norte-americanos.

Y: taxa de criminalidade

X: taxa de analfabetismo

Diagrama de dispersão



Podemos notar que, conforme aumenta a taxa de analfabetismo (X), a taxa de criminalidade (Y) tende a aumentar. Nota-se também uma tendência linear.

Correlação entre X e Y: 0,702

a reta ajustada é:

$$\hat{Y} = 2,397 + 4,257 X$$

^

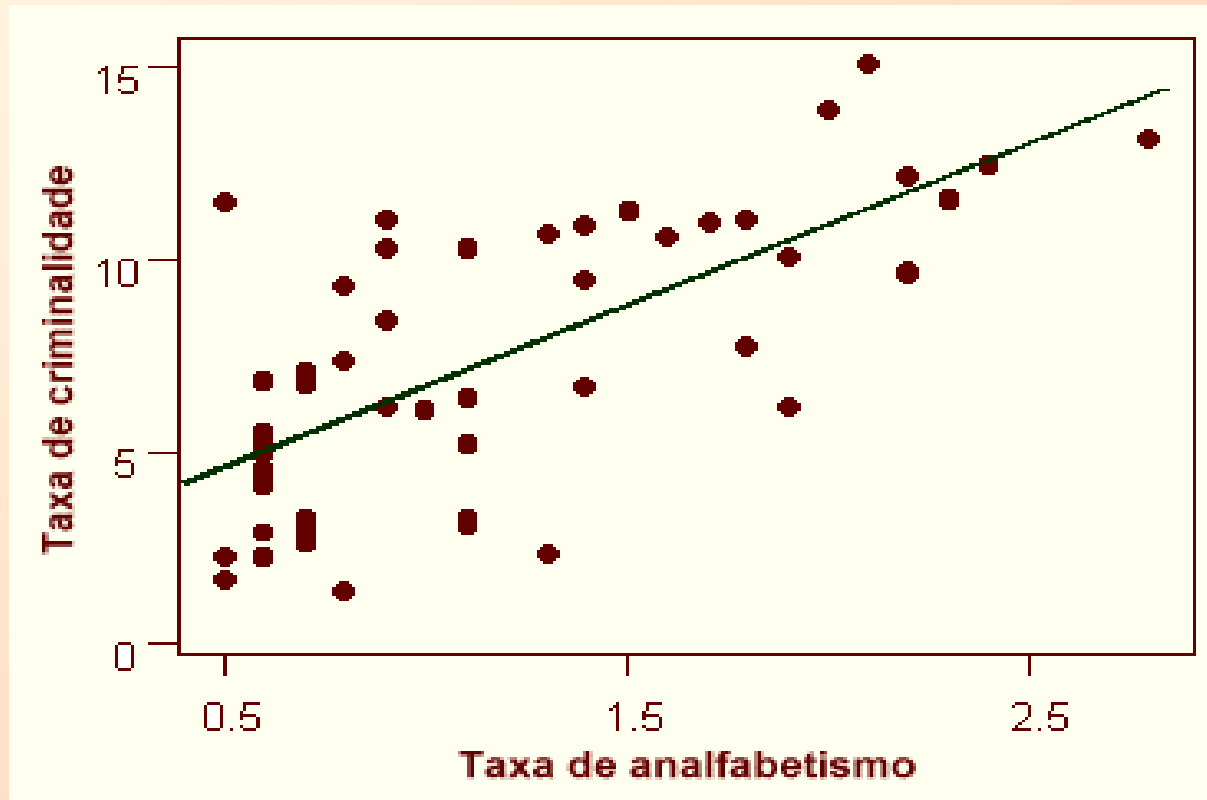
Y : valor predito para a taxa de criminalidade

X : taxa de analfabetismo

Interpretação de b:

Para um aumento de uma unidade na taxa do analfabetismo (X), a taxa de criminalidade (Y) aumenta, em média, 4,257 unidades.

Graficamente, temos



Como desenhar a reta no gráfico?

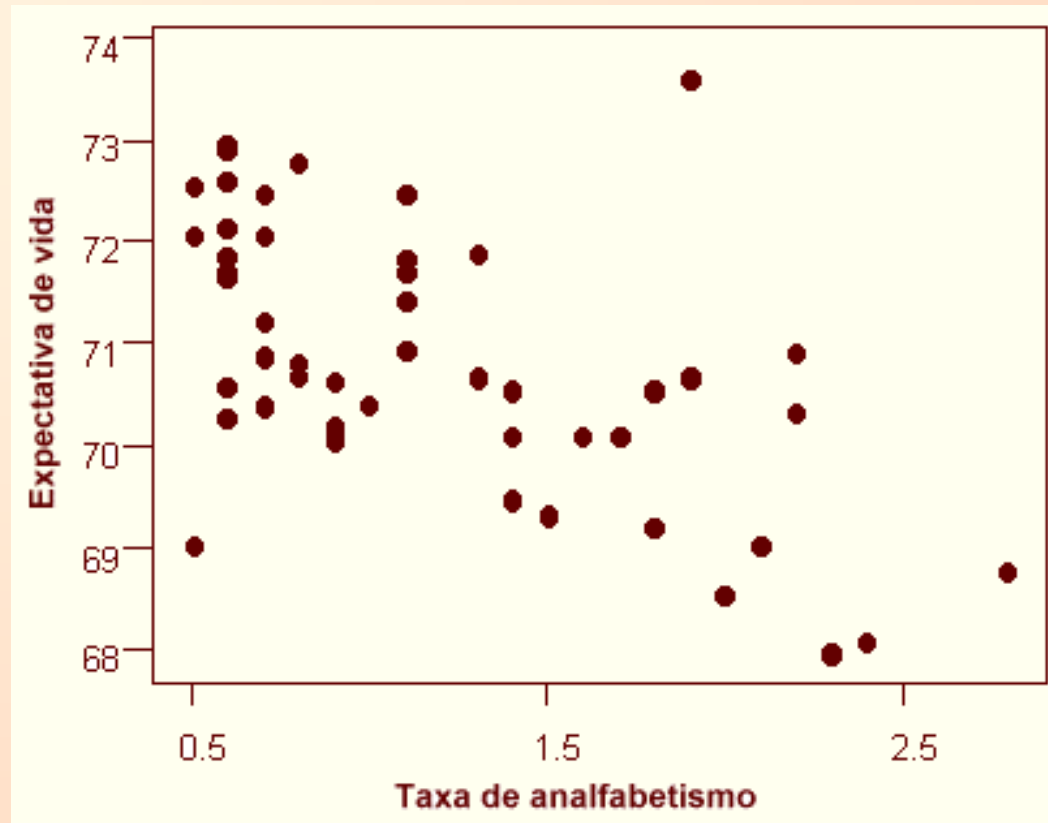
Exemplo 3: expectativa de vida e analfabetismo

Considere as duas variáveis observadas em 50 estados norte-americanos.

Y: expectativa de vida

X: taxa de analfabetismo

Diagrama de dispersão



Podemos notar que, conforme aumenta a taxa de analfabetismo (X), a expectativa de vida (Y) tende a diminuir. Nota-se também uma tendência linear.

Correlação entre X e Y:- 0,59

a reta ajustada é:

$$\hat{Y} = 72,395 - 1,296 X$$

^

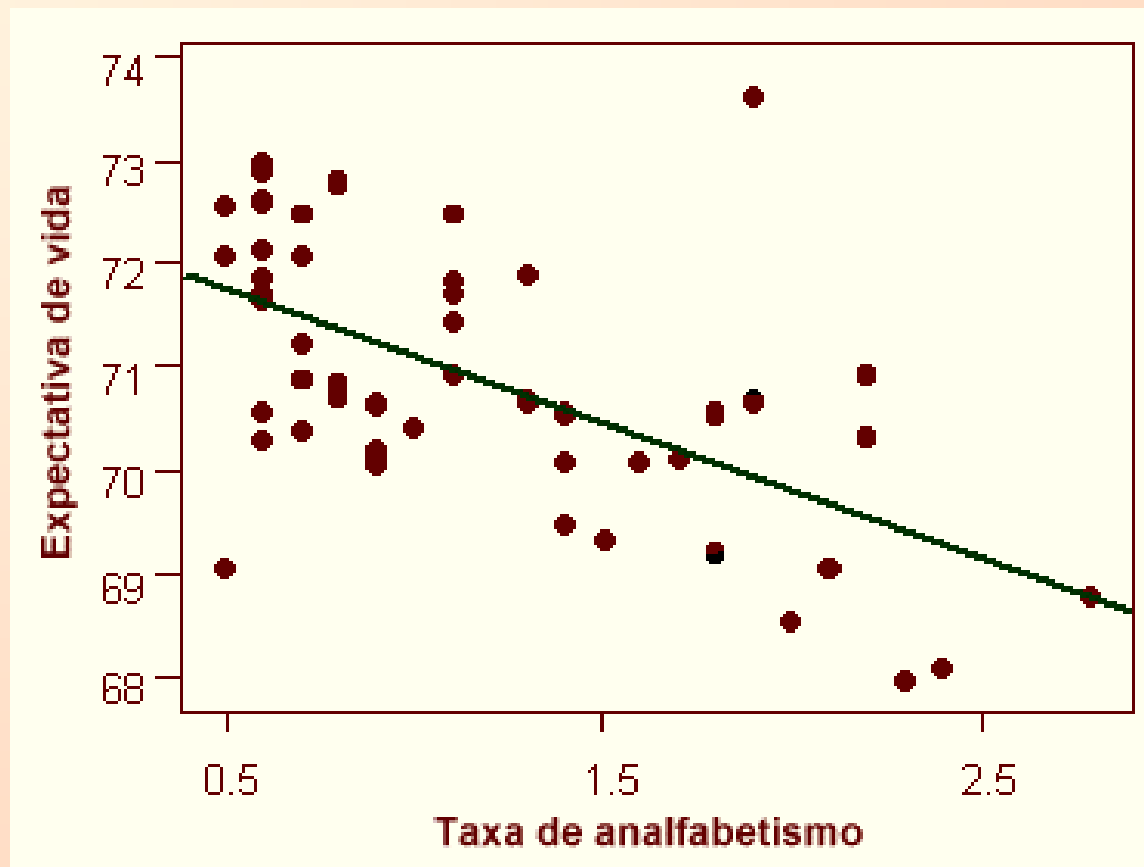
Y : valor predito para a expectativa de vida

X : taxa de analfabetismo

Interpretação de b:

Para um aumento de uma unidade na taxa do analfabetismo (X), a expectativa de vida (Y) diminui, em média, 1,296 anos.

Graficamente, temos



This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.