

Estimadores de Mínimos Cuadrados

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

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

Modelo

$$Y = \alpha + \beta X + \varepsilon$$

Reta ajustada:

$$\hat{Y} = a + bX$$

O que são **a** e **b**?

a: intercepto

b: inclinação

Interpretação de b:

Para cada aumento de uma unidade em X, temos um aumento médio de b unidades em Y.

Reta ajustada (método de mínimos quadrados)

Os coeficientes a e b são calculados da seguinte maneira:

$$b = \frac{\sum_{i=1}^n X_i Y_i - n.\bar{X}.\bar{Y}}{(n-1).S_X^2}$$

$$a = \bar{Y} - b.\bar{X}$$

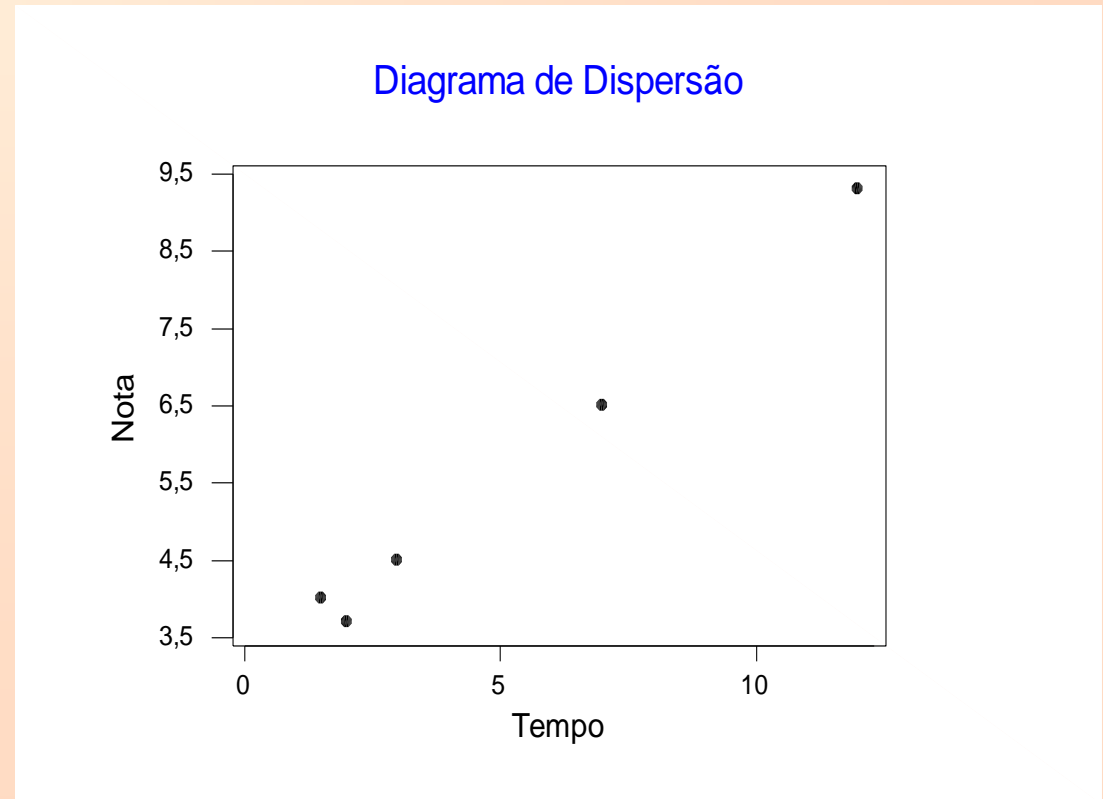
Exemplo 1: nota da prova e tempo de estudo

X : tempo de estudo (em horas)

Y : nota da prova

Pares de observações (X_i, Y_i) para cada estudante

Tempo (X)	Nota (Y)
3,0	4,5
7,0	6,5
2,0	3,7
1,5	4,0
12,0	9,3



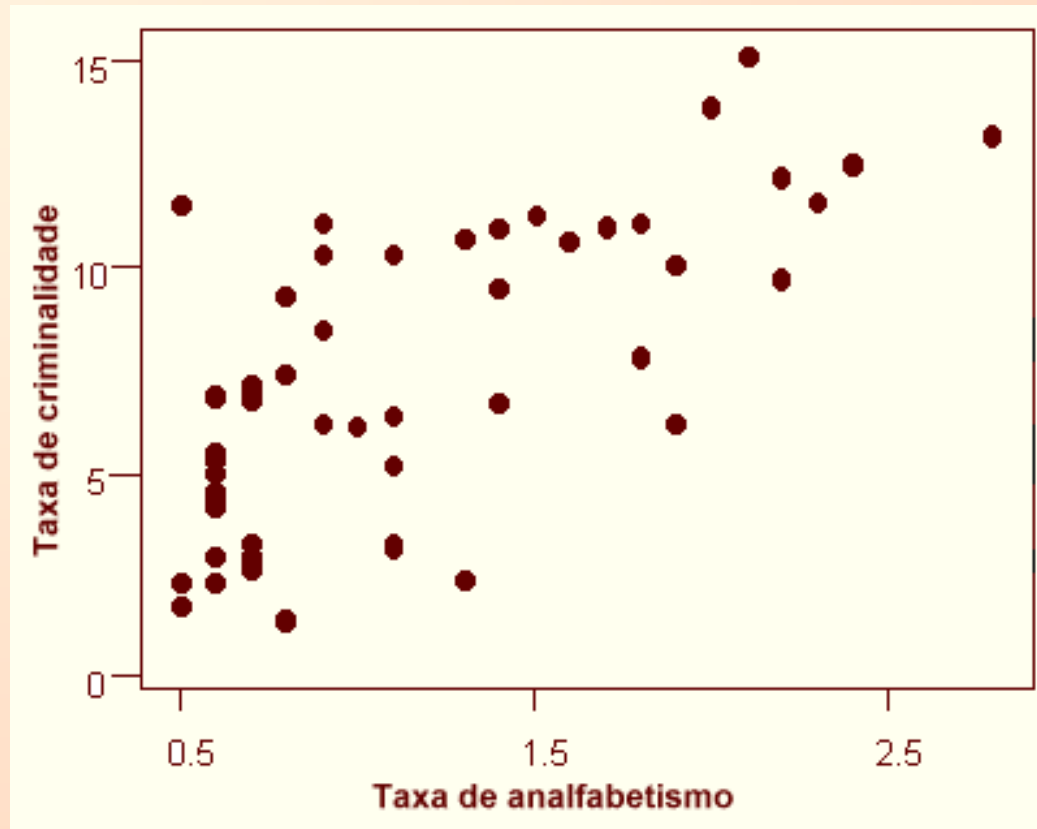
Exemplo 2: criminalidade e analfabetismo

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.

$\bar{Y} = 7,38$ (média de Y) e $S_Y = 3,692$ (desvio padrão de Y)

$\bar{X} = 1,17$ (média de X) e $S_X = 0,609$ (desvio padrão de X)

$\Sigma X_i Y_i = 509,12$

No exemplo 2,

a reta ajustada é:

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

^

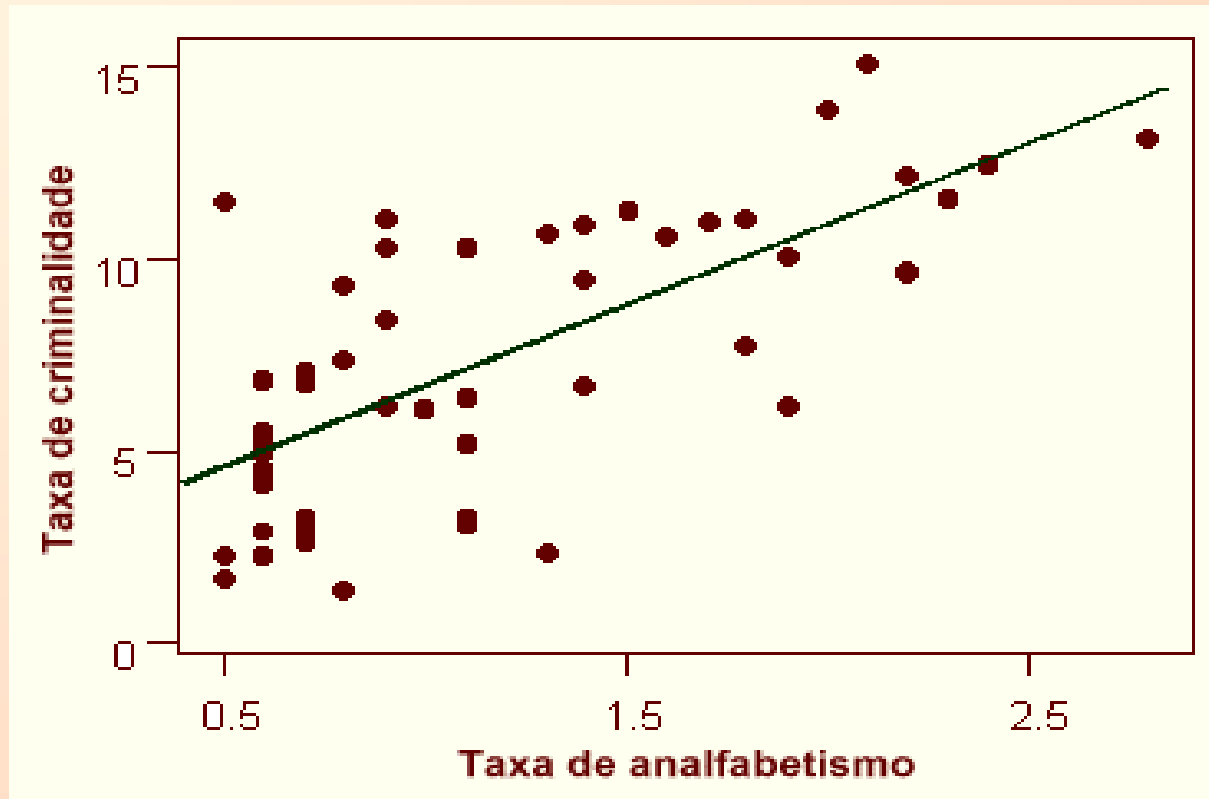
Y : valor predito para a taxa de criminalidade

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?

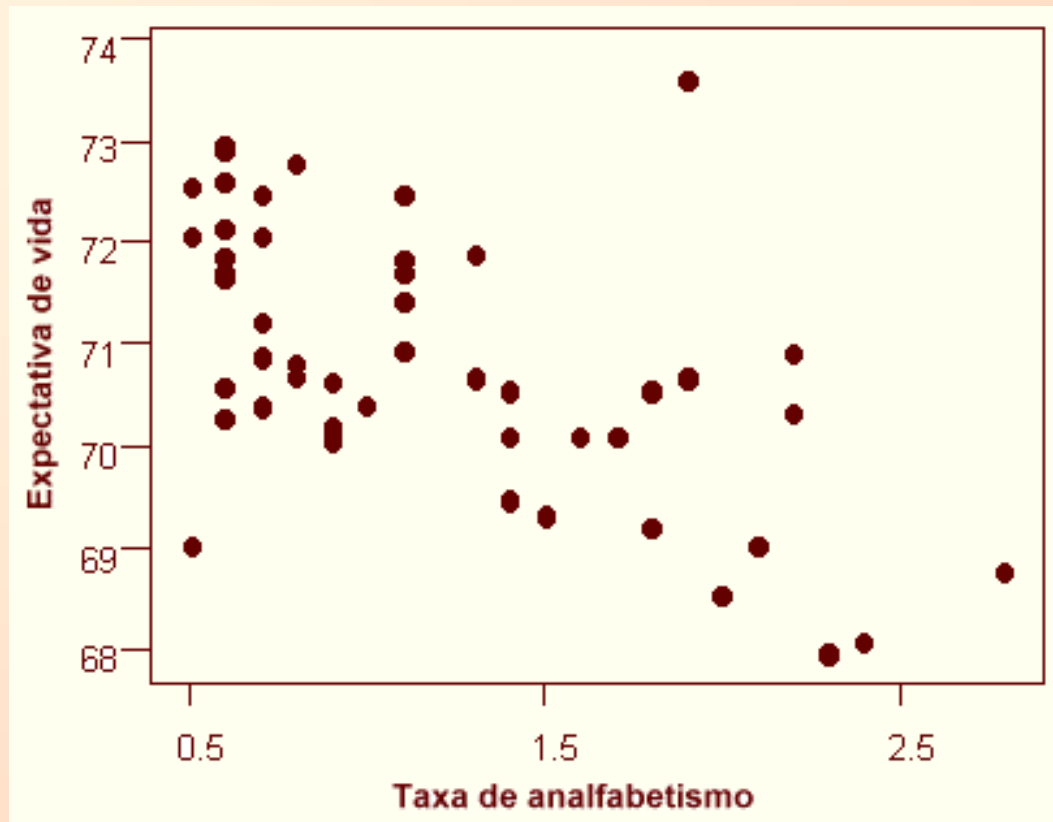
Exemplo 3: expectativa de vida e analfabetismo

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.

$\bar{Y} = 70,88$ (média de Y) e $S_Y = 1,342$ (desvio padrão de Y)

$\bar{X} = 1,17$ (média de X) e $S_X = 0,609$ (desvio padrão de X)

$\Sigma X_i Y_i = 4122,8$

No exemplo 3,

a reta ajustada é:

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

^

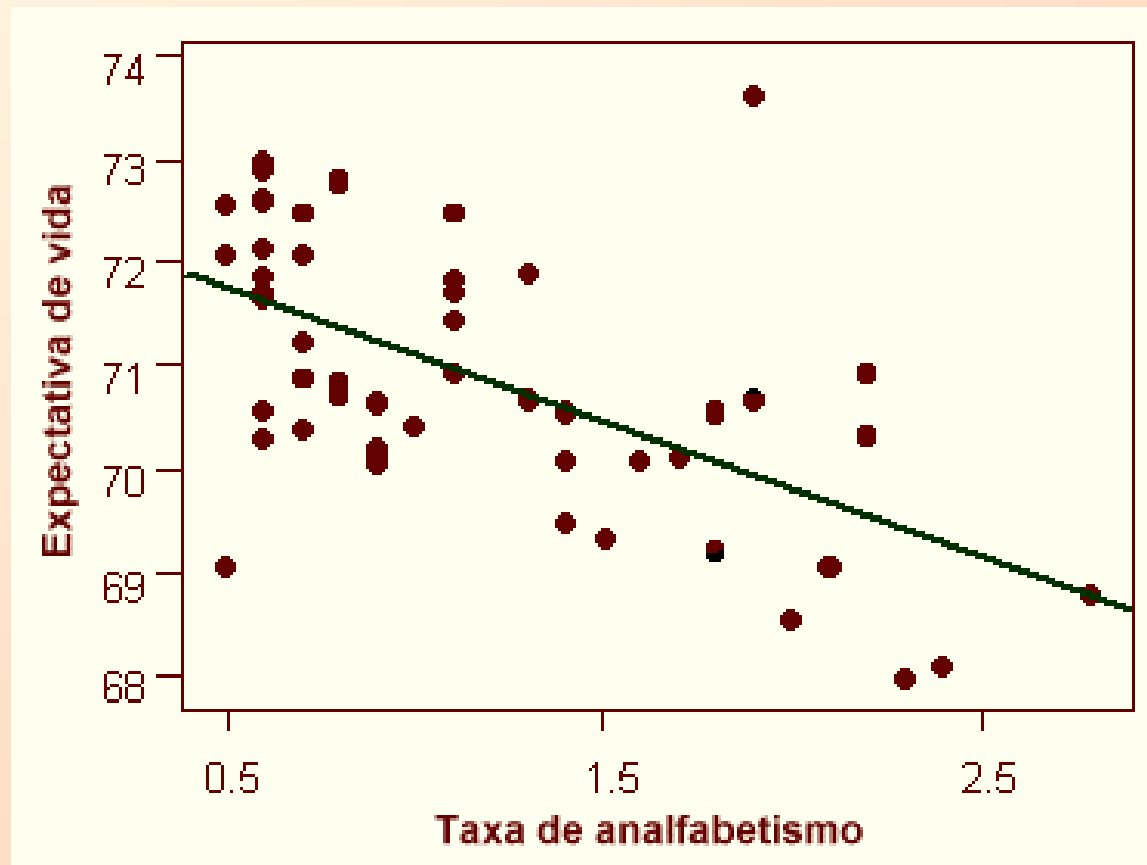
Y : valor predito para a expectativa de vida

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



Exemplo 4: consumo de cerveja e temperatura

Y: consumo de cerveja diário por mil habitantes, em litros.

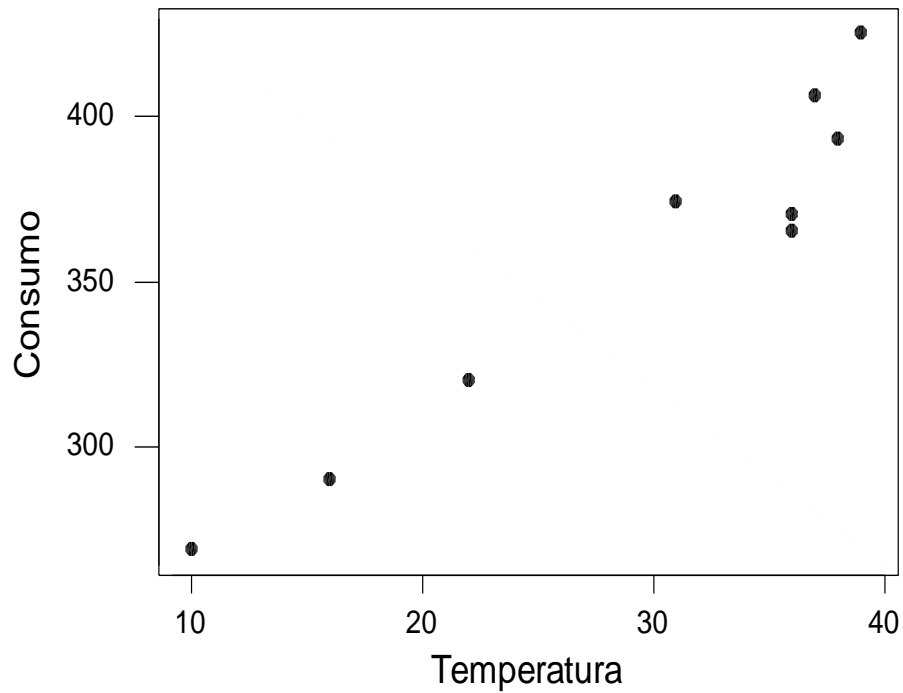
X: temperatura máxima (em °C).

As variáveis foram observadas em nove localidades com as mesmas características demográficas e sócio-econômicas.

Dados:

Localidade	Temperatura (X)	Consumo (Y)
1	16	290
2	31	374
3	38	393
4	39	425
5	37	406
6	36	370
7	36	365
8	22	320
9	10	269

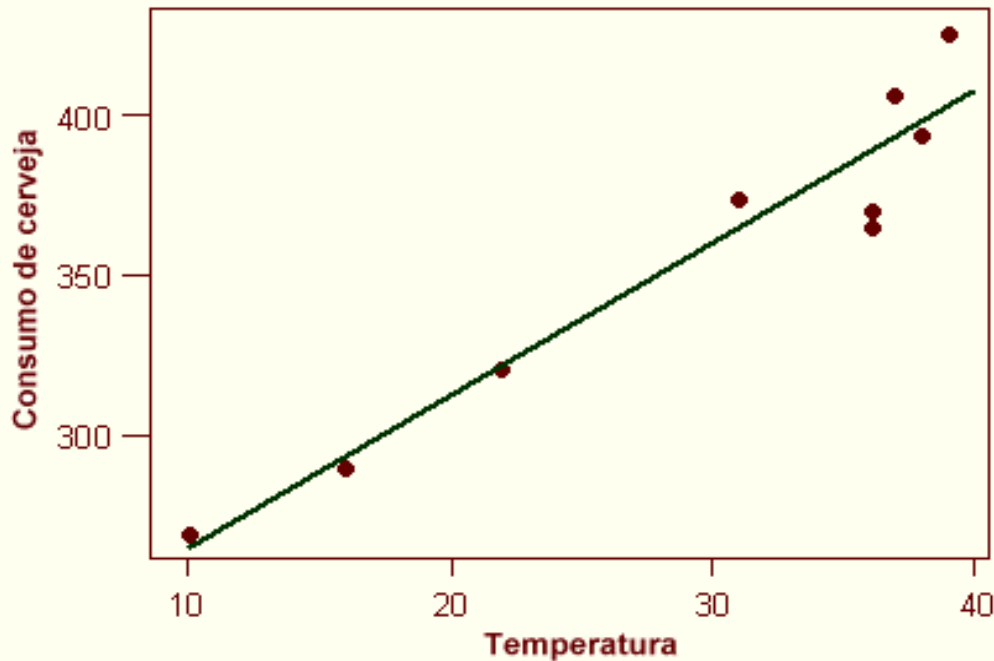
Diagrama de dispersão



A correlação entre X e Y é $r = 0,962$.

A reta ajustada é:

$$\hat{Y} = 217,37 + 4,74 X$$



Qual a interpretação de b ?
Aumentando-se um grau de temperatura (X), o consumo de cerveja (Y) aumenta, em média, 4,74 litros por mil habitantes.

Qual o consumo previsto para uma temperatura de 25°C?

$$\hat{Y} = 217,37 + 4,74 \cdot 25 = 335,87 \text{ litros}$$