MAE-212: Introdução à Probabilidade e à Estatística II

Aula 05

Inferência Estatística

Estimadores de Mínimos Quadrados

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

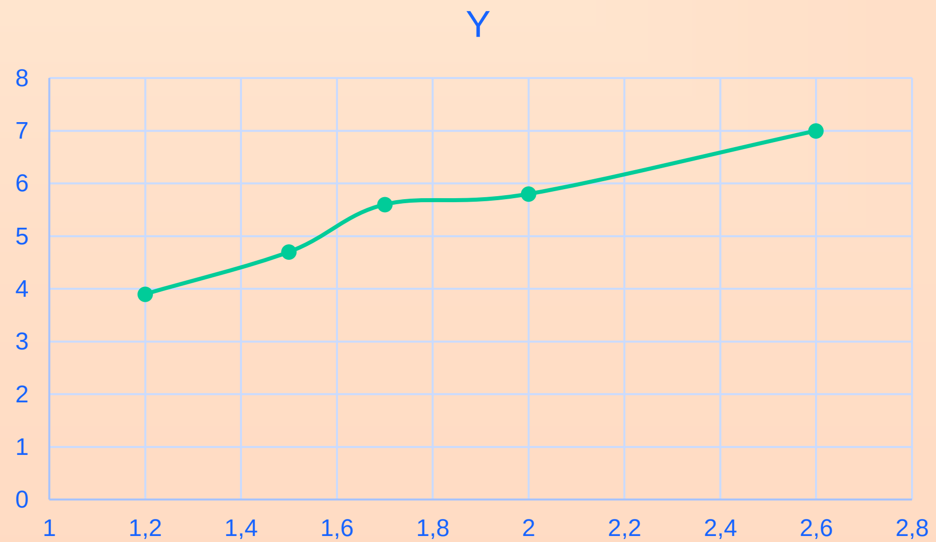
Expectativa de vida e taxa de analfabetismo

Exemplo 11.9 Um engenheiro está estudando a resistência Y de uma fibra em função de seu diâmetro X e notou que as variáveis são aproximadamente proporcionais, isto é, elas obedecem à relação

$$Y \approx \theta X, \tag{11.25}$$

em que θ é o coeficiente de proporcionalidade. Agora ele deseja estimar o parâmetro θ, baseado numa amostra de cinco unidades, que, submetidas a mensuração e testes, produziram os resultados:

$$X: \quad 1{,}2 \quad 1{,}5 \quad 1{,}7 \quad 2{,}0 \quad 2{,}6 \qquad \bar{X} = 1{,}8;$$
$$Y: \quad 3{,}9 \quad 4{,}7 \quad 5{,}6 \quad 5{,}8 \quad 7{,}0, \qquad \bar{Y} = 5{,}4.$$



5,4/1,8 = 3

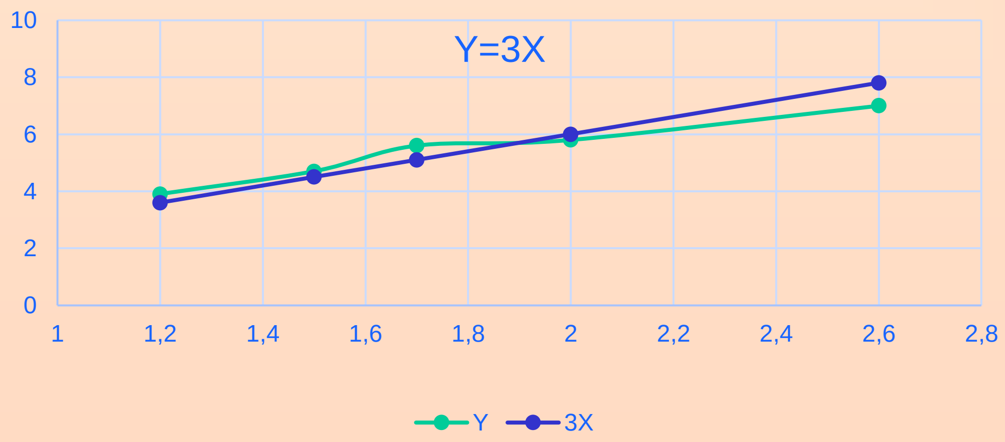
Inspecionando os resultados, conclui-se que $\hat{\theta} = 3$ parece ser um valor razoável. Como verificar a qualidade dessa estimativa? Podemos utilizar o modelo $\hat{Y} = 3X$ e ver como esse prevê os valores de Y, para os dados valores de X, e como são as discrepâncias entre os valores observados e os estimados pelo modelo. Essa análise está resumida na Tabela 11.1.

Tabela 11.1 Análise do modelo $\hat{Y} = 3X$.

X	Y	$3X$	$Y - 3X$	$(Y - 3X)^2$
1,2	3,9	3,6	0,3	0,09
1,5	4,7	4,5	0,2	0,04
1,7	5,6	5,1	0,5	0,25
2,0	5,8	6,0	–0,2	0,04
2,6	7,0	7,8	–0,8	0,64
		Total	0	1,06

erro

erro quadrático total



Os valores da coluna $(Y-3X)$ medem a inadequação do modelo para cada observação da amostra, enquanto o valor $\sum_{i=1}^{5}(Y_i-3X_i)^2=1{,}06$ é uma tentativa de medir "o erro quadrático total da amostra". Como em situações anteriores, elevou-se ao quadrado para evitar o problema do sinal. Quanto menor for o erro quadrático total, melhor será a estimativa. Isso nos sugere procurar a estimativa que torne mínima essa soma de quadrados. Matematicamente, o problema passa a ser o de encontrar o valor de θ que minimize a função

$$S(\theta)=\sum_{i=1}^{5}(Y_i-\theta X_i)^2. \qquad (11.26)$$

O mínimo da função é obtido derivando-a em relação a θ, e igualando o resultado a zero (ver Morettin et al., 2005), o que resulta

$$\frac{dS(\theta)}{d\theta}=\sum_{i=1}^{5}\left(Y_i-\hat{\theta}X_i\right)(-2X_i)=0.$$

Resolvendo essa equação, obtemos

$$\hat{\theta}_{MQ}=\frac{\sum_{i=1}^{5}X_iY_i}{\sum_{i=1}^{5}X_i^2}.$$

Usando os dados acima encontramos $\hat{\theta}_{MQ}=2{,}94$, que conduz a um valor mínimo para $S(\theta)$ de 0,94. Observe que esse valor é realmente menor do que o observado para $\theta=3$, ou seja, 1,06.

É comum supor que ε tem a mesma distribuição, para todo valor x da variável explicativa X. Desse modo, é comum escrever

$$Y = \theta x + \varepsilon,$$

com ε seguindo a distribuição $f_\varepsilon(.)$, com média zero. Como ilustração, poderíamos supor que $\varepsilon \sim N(0, \sigma^2)$, para todo x. Quanto menor for a variância σ^2, melhor será a "previsão" de Y como função de x. Assim, parece razoável escolher θ que torna mínima a soma dos quadrados do erros:

$$\sum_{i=1}^{5} \varepsilon_i^2 = \sum_{i=1}^{5} \left(Y_i - \theta X_i\right)^2 .$$

O modelo acima pode ser generalizado, de modo a envolver outras funções do parâmetro θ, resultando no modelo

$$Y = g(X; \theta) + \varepsilon, \qquad (11.27)$$

e devemos procurar o valor de θ que minimize a função

$$S(\theta) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} \left(Y_i - g\left(X_i;\theta\right)\right)^2, \qquad (11.28)$$

para uma amostra $(X_1, Y_1), \ \ldots, (X_n, Y_n)$ das variáveis X e Y. A solução $\hat{\theta}_{MQ}$ é chamada de estimador de mínimos quadrados (EMQ) de θ.

Regressão

Representação gráfica de duas variáveis quantitativas: Diagrama de dispersão

Modelo

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

Reta ajustada:

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

O que são a e b?

a: intercepto

b: inclinação

<u>Interpretação de b</u>:

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:

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

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

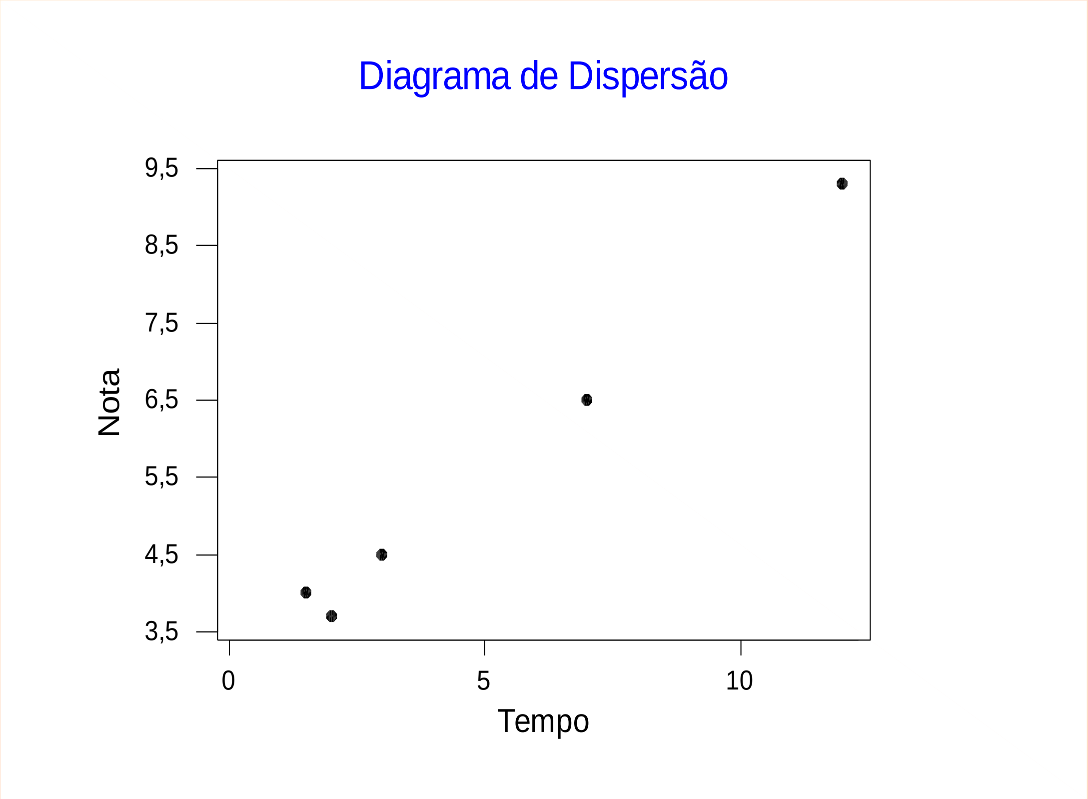
Exemplo 1: nota da prova e tempo de estudo

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



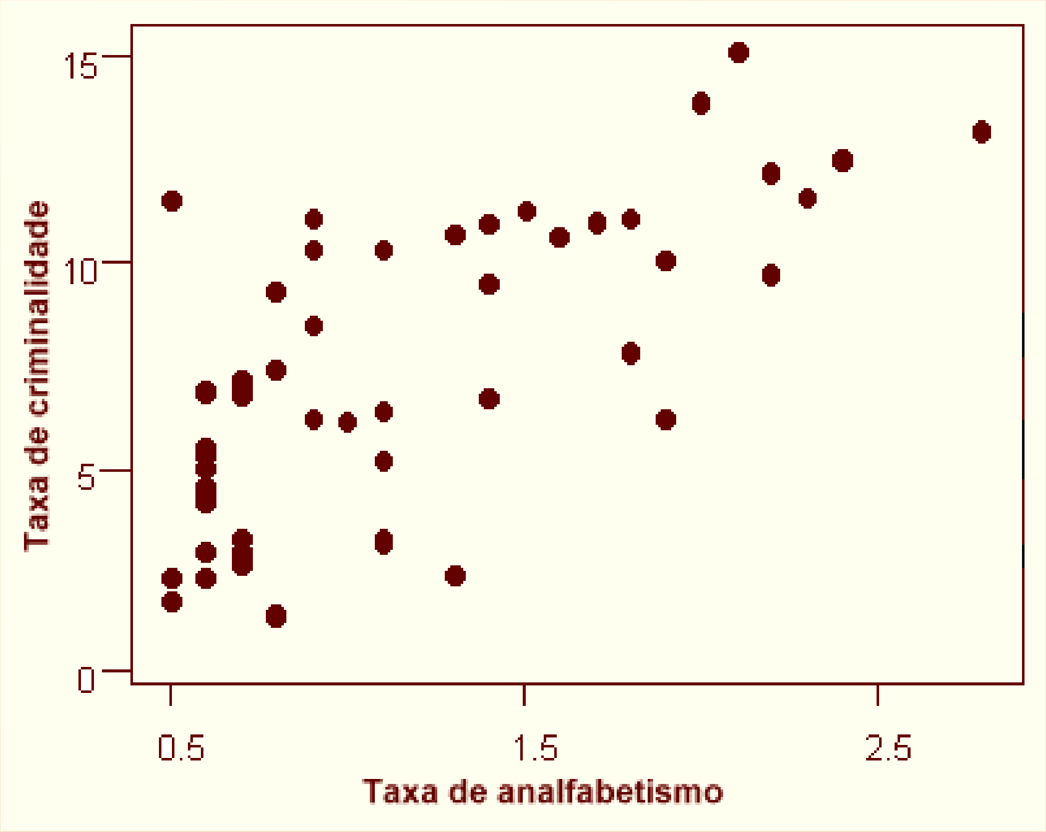
Exemplo 2: criminalidade e analfabetismo

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 <u>reta</u> <u>ajustada</u> é:

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

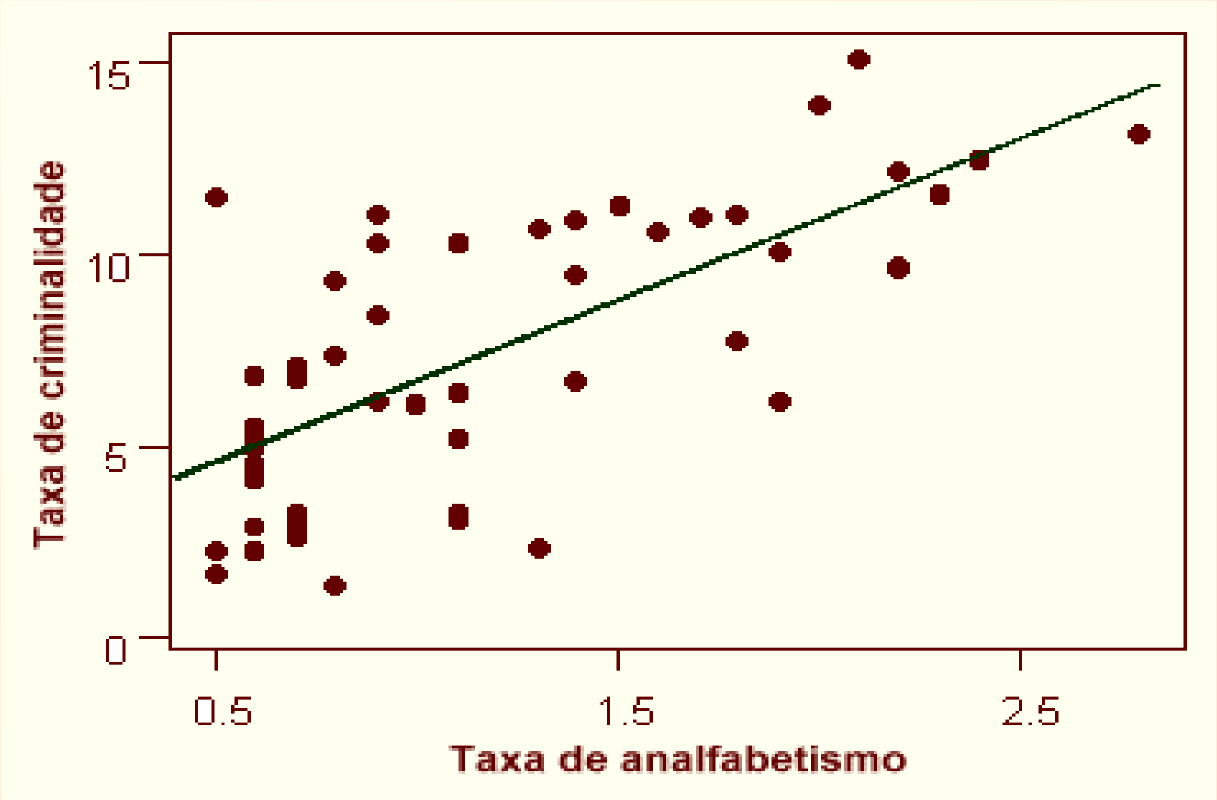
$\hat{Y}$: valor predito para a taxa de criminalidade

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?

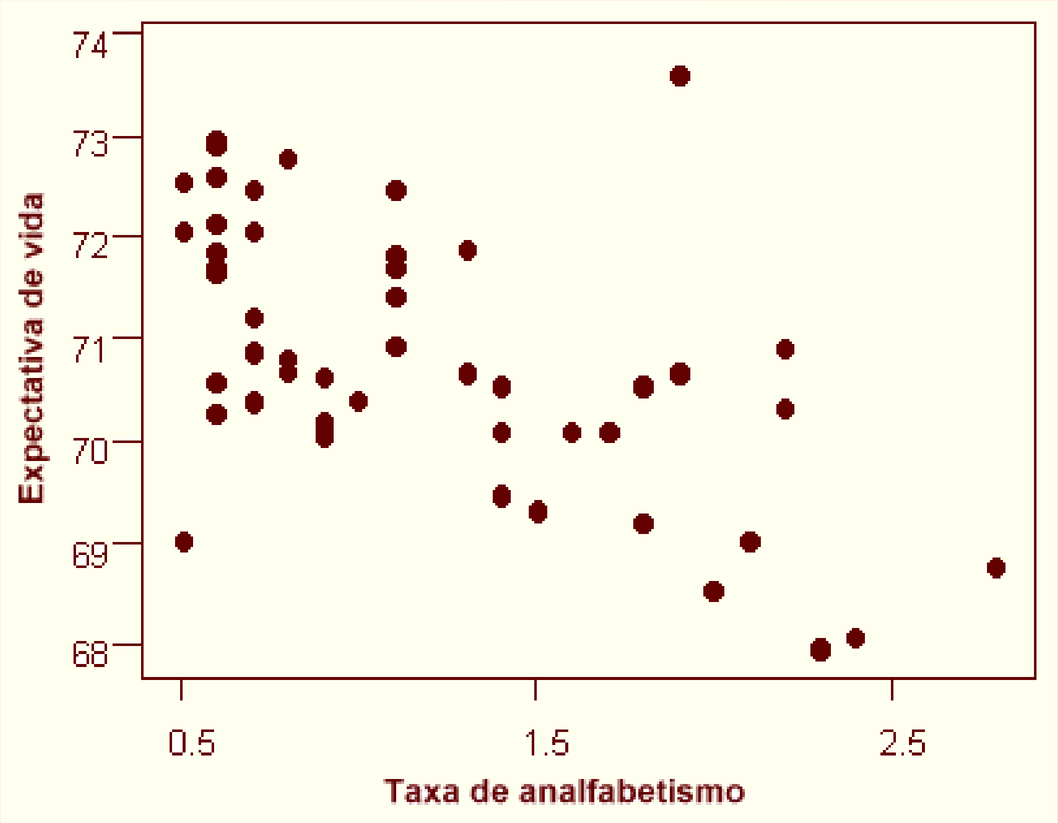
Exemplo 3: expectativa de vida e analfabetismo

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.

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

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

$\Sigma X_i Y_i = 4122,8$

No exemplo 3,

a <u>reta</u> <u>ajustada</u> é:

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

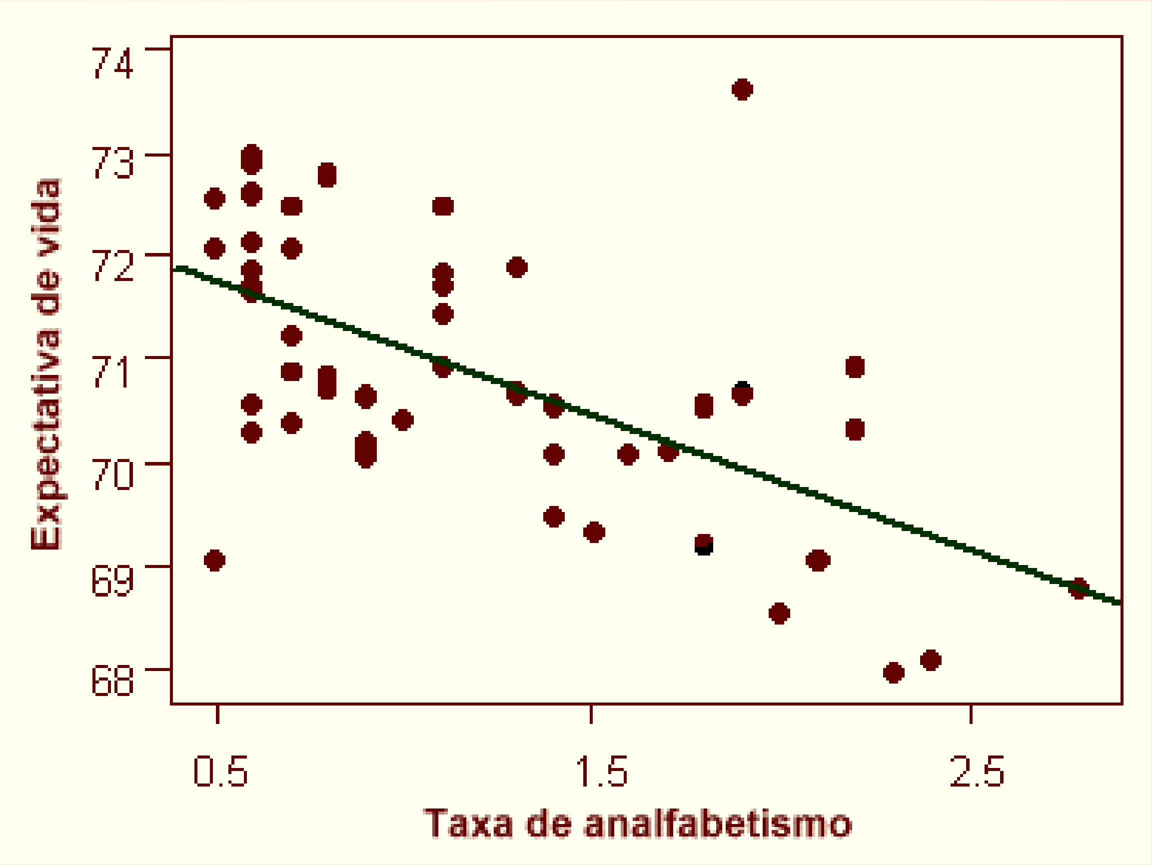
$\hat{Y}$: valor predito para a expectativa de vida

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.

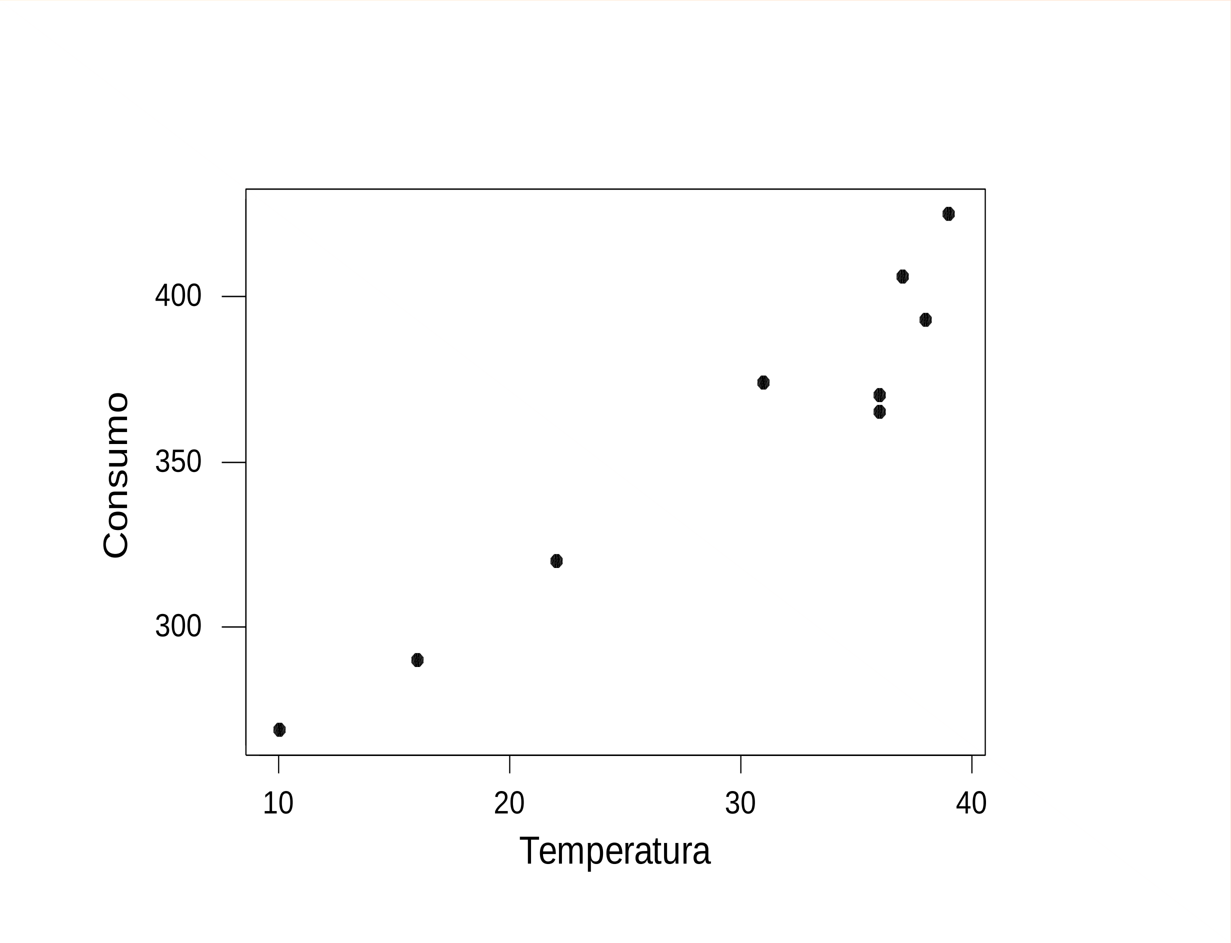
X: temperatura máxima (em ºC).

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

Dados:

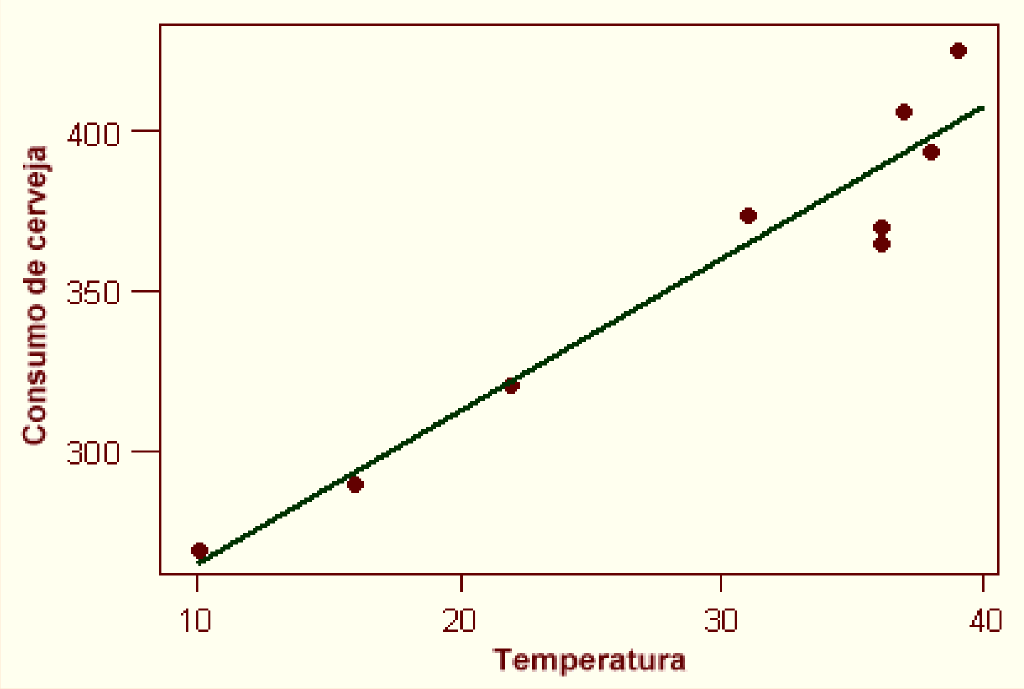
Localidade	Temperatura (X)	Consumo (Y)
1	16	290
2	31	374
3	38	393
4	39	425
5	37	406
6	36	370
7	36	365
8	22	320
9	10	269

Diagrama de dispersão



A <u>correlação</u> 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 25ºC?

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