

Estatística Aplicada



Correlação e regressão linear Cap. 12 – Prof. Simões



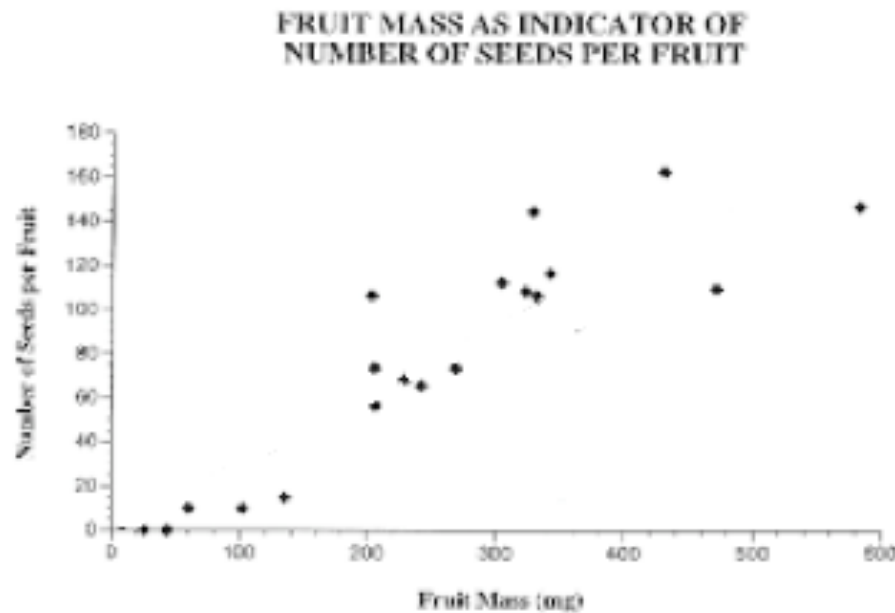
Correlação

- Muitas vezes queremos investigar se duas ou mais variáveis estão ou não relacionadas
 - adição de álcool à gasolina e consumo por quilômetro rodado
 - distância de freagem de um carro e quantidade de um componente na borracha
 - horas de trabalho e rendimento do trabalho
 - quantidade de chuva e produtividade agrícola
 - quantidade de fertilizantes e safra de cereais
 - quantidade de veículos de uma cidade e poluição
 - volume de um composto químico e tempo de reação
- Chamamos de correlação à medida da dependência de uma variável em relação à outra

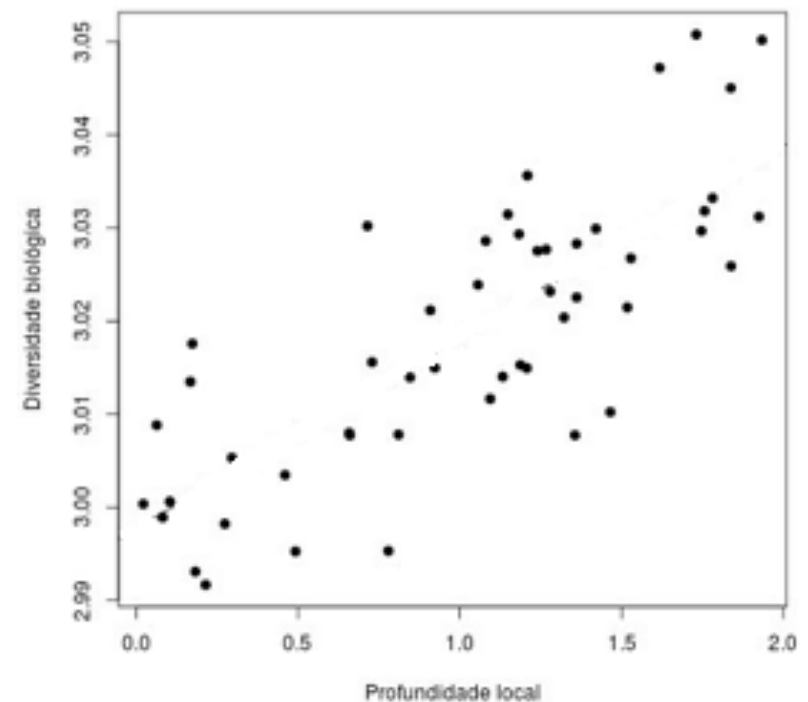
Correlação

- O valor varia de 0 a 1, sendo que 0 significa nenhuma correlação, e um total correlação. Referências:
 - 0 a 0,3 – baixa correlação
 - 0,3 a 0,7 – correlação moderada
 - 0,7 a 1,0 – forte correlação

Cuidado: o fato de haver uma correlação estatística, não implica em uma correlação causal. Exemplo: consumo de água e aumento da taxa de criminalidade.

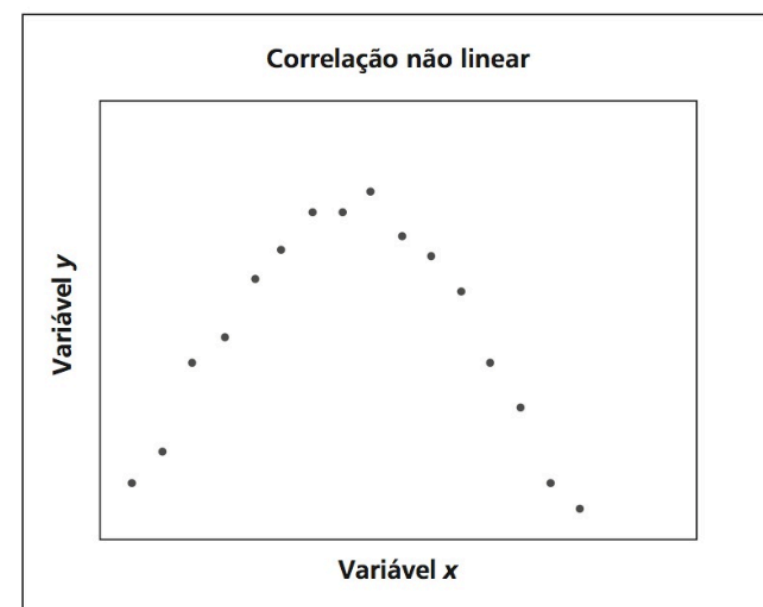
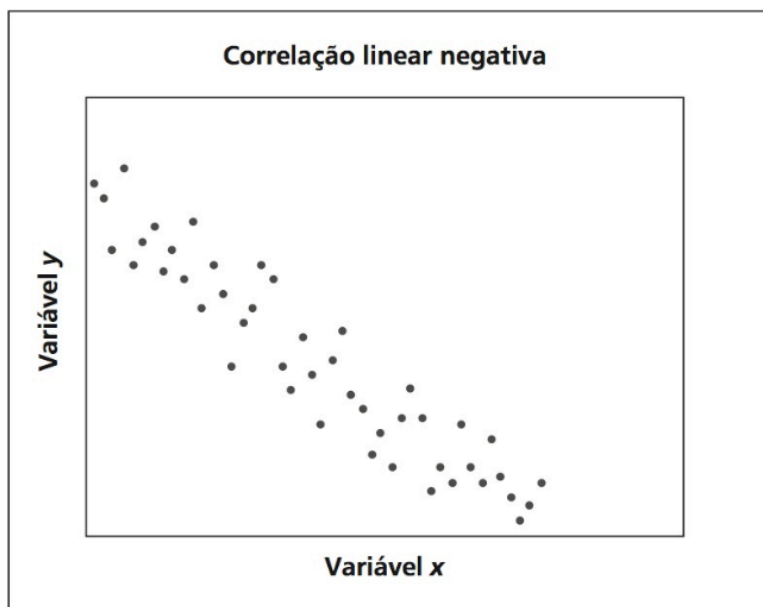
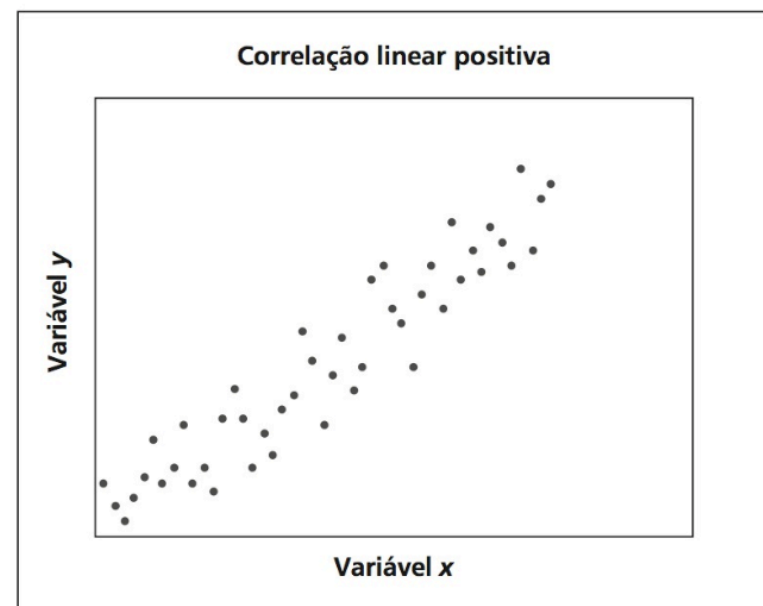
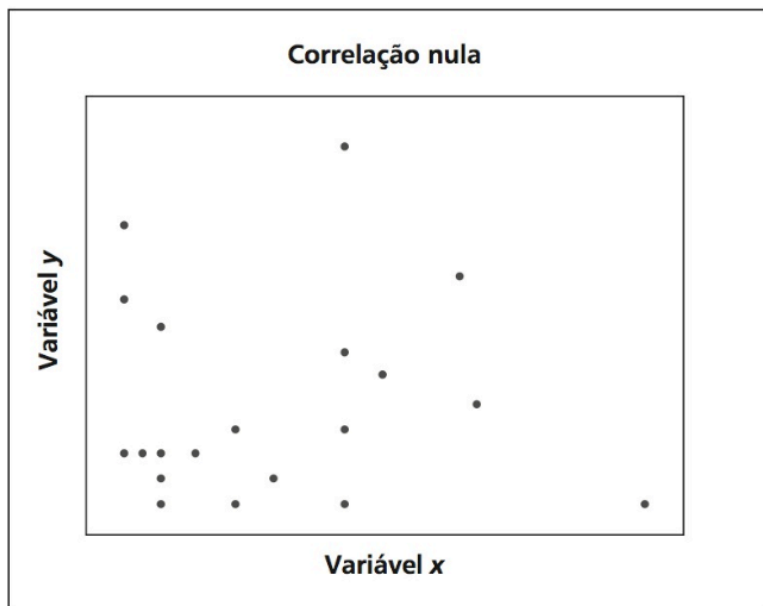


Moderada a forte



Fraca a moderada

Tipos de correlação



Coeficiente de correlação linear

- A fórmula de Karl Pearson é:

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

- E o coeficiente de determinação é:

$$C_d = r^2$$

O coeficiente de determinação informa o quanto de uma variável pode ser explicado pela outra. Por exemplo, suponhamos que o valor de r para gastos com propaganda versus aumento de vendas seja $r = 0,9$. Então

$$C_d = r^2 = 0,9^2 = 0,81$$

Isso significa que 81% do aumento das vendas pode ser explicado pelos gastos com a propaganda.

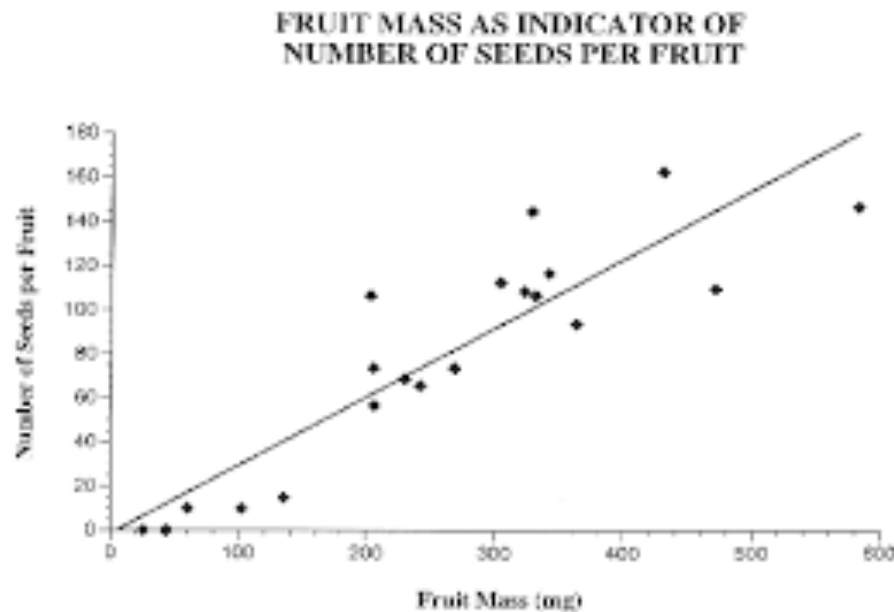
Regressão linear

- Consiste em um conjunto de cálculos que permite encontrar a **equação da reta** de melhor ajuste aos pontos encontrados empiricamente.

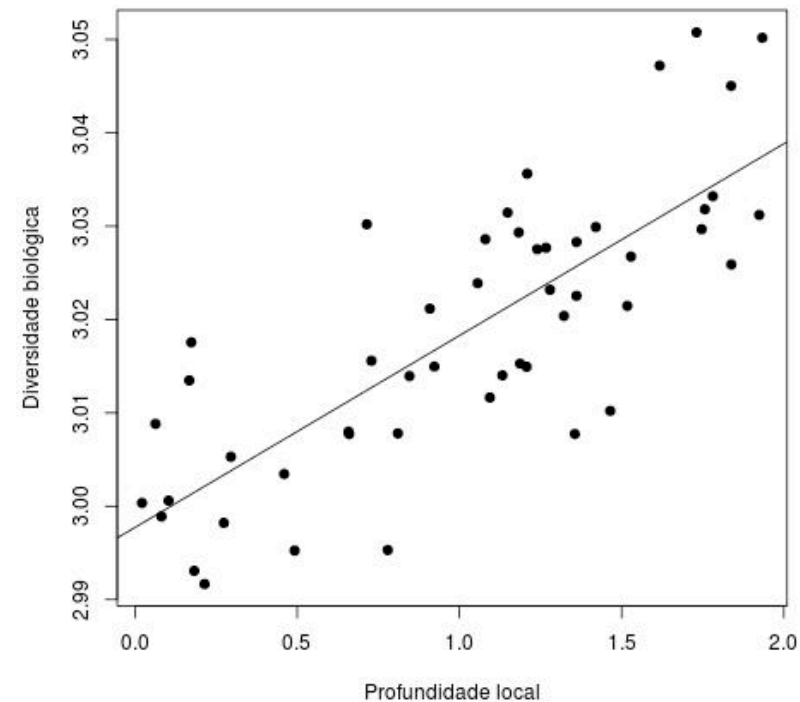
$$y = a + b \cdot x$$

$a \Rightarrow$ valor inicial

$b \Rightarrow$ taxa de variação



$$y = 0,30 \cdot x$$

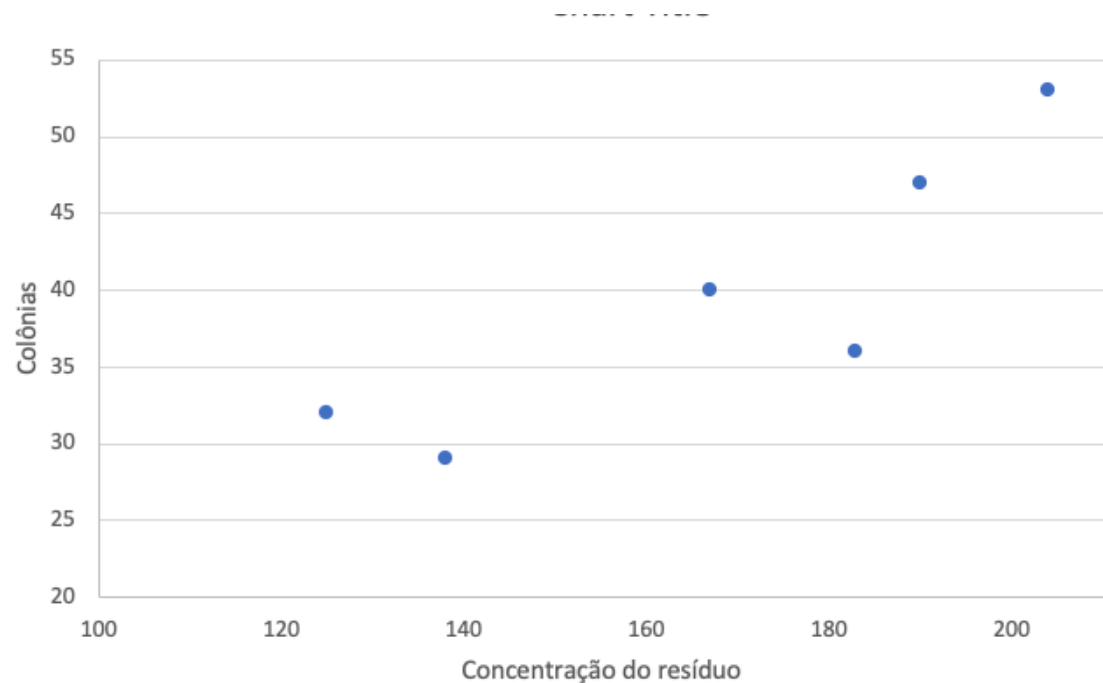


$$y = 2,995 + 0,02 \cdot x$$

Exemplo

- Suspeita-se que a presença de um resíduo contribua para a reprodução de uma bactéria. Foram feitos experimentos com várias concentrações do resíduo, avaliando-se a quantidade de colônias de bactérias observadas, com os resultados abaixo. Estabeleça uma correlação entre as variáveis e faça uma previsão da quantidade de colônias possíveis para uma concentração de 200 unidades do resíduo.

Resíduo (ppm)	Colônias de bactérias
138	29
167	40
204	53
183	36
125	32
190	47



Resolução, correlação

Solução:	<i>x</i>	<i>y</i>	<i>xy</i>	<i>x</i> ²	<i>y</i> ²
<i>n</i> = 6 {	138	29	4002	19044	841
	167	40	6680	27889	1600
	204	53	10812	41616	2809
	183	36	6588	33489	1296
	125	32	4000	15625	1024
	190	47	8930	36100	2209
	$\Sigma x = 1007$	$\Sigma y = 237$	$\Sigma xy = 41012$	$\Sigma x^2 = 173763$	$\Sigma y^2 = 9779$

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

$$r = \frac{6 \cdot 41012 - 1007 \cdot 237}{\sqrt{[6 \cdot 173763 - (1007)^2] \cdot [6 \cdot 9779 - (237)^2]}} = + 0,877$$

$$r^2 = (0,877)^2 = 0,7691 \text{ (76,91\%).}$$

Conclusão: a correlação é forte e positiva, sendo que aprox. 77% do surgimento das bactérias pode ser explicado pelo aumento da concentração do resíduo.

Resolução, equação linear

- Coeficiente angular b

$$b = \frac{n \cdot \sum xy - (\sum x) \cdot (\sum y)}{n \cdot \sum x^2 - (\sum x)^2} \Rightarrow b = \frac{7413}{28529} \Rightarrow \boxed{b = 0,260}$$

- Coeficiente linear a

$$a = \frac{\sum y}{n} - b \cdot \frac{\sum x}{n} \Rightarrow a = \frac{237}{6} - 0,260 \cdot \frac{1007}{6} = 39,5 - 43,637$$

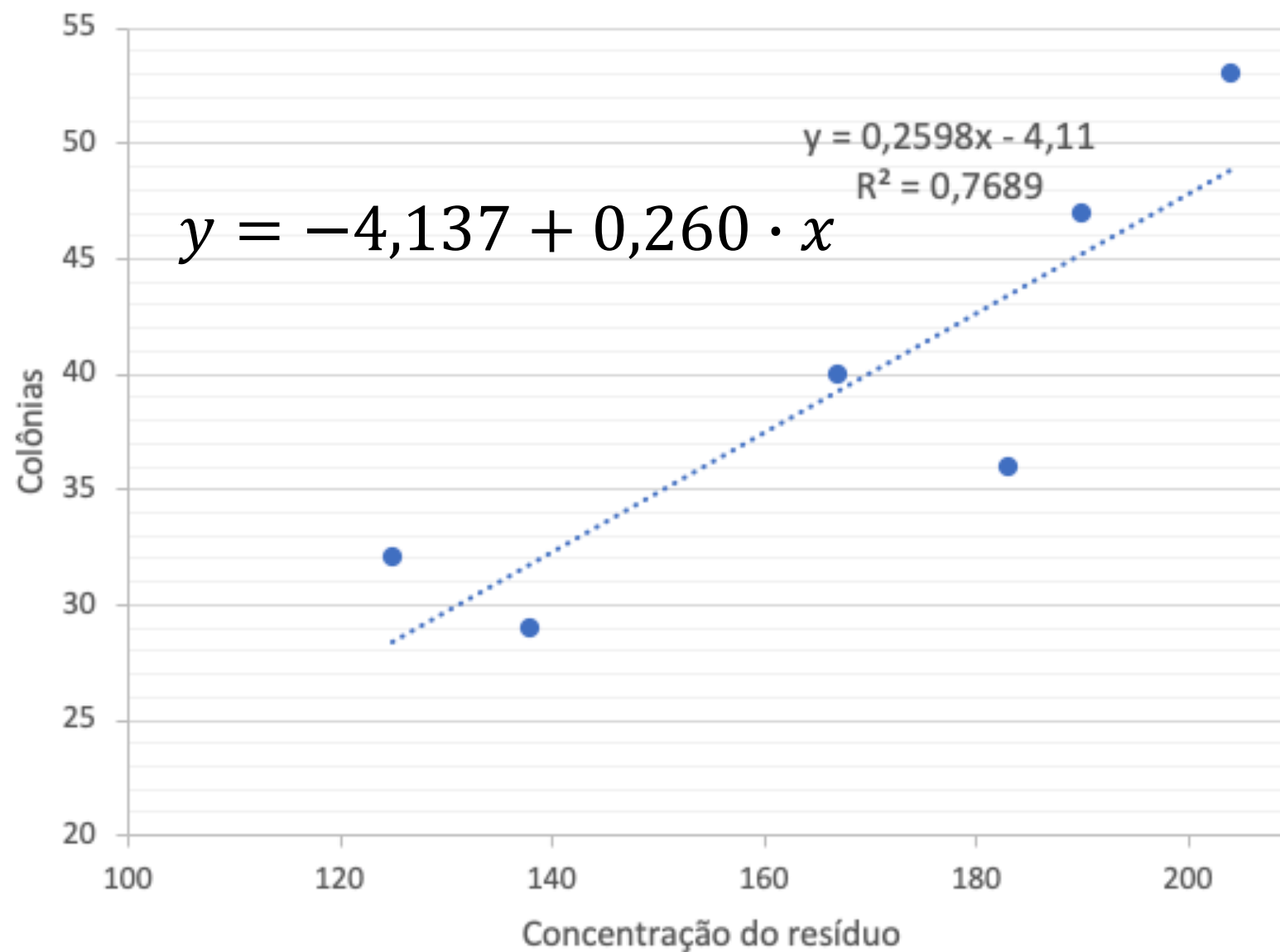
$$\boxed{a = -4,137}$$

- Equação da reta

$$y = -4,137 + 0,260 \cdot x$$

Resolução, equação linear

- Equação da reta



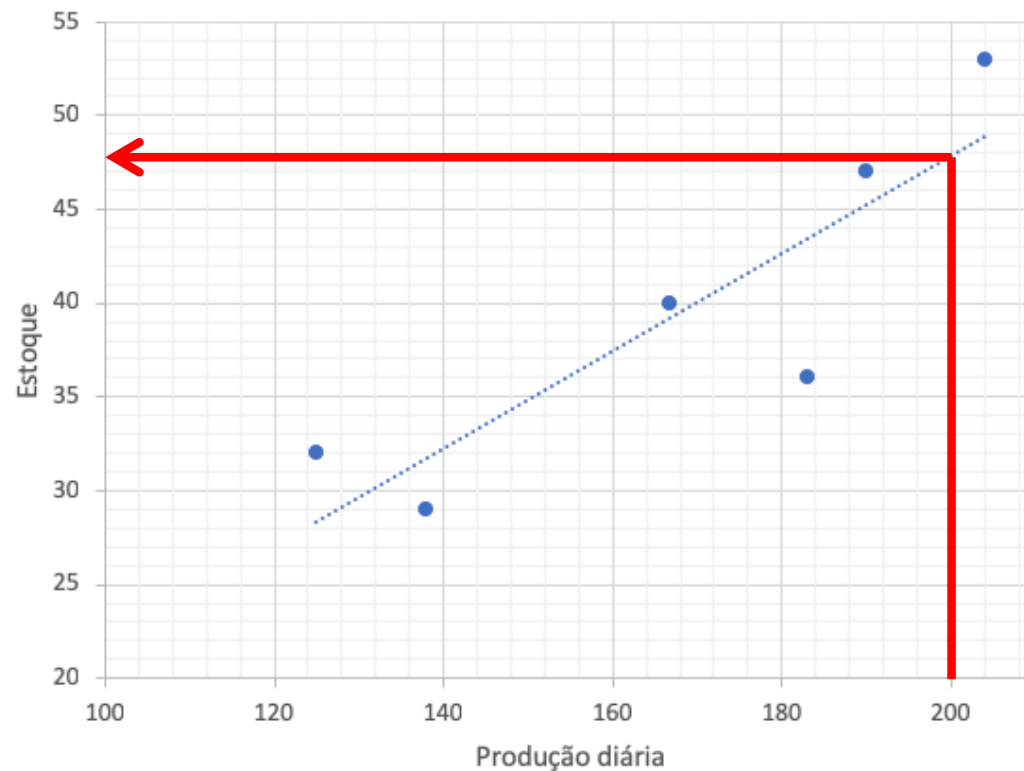
Interpolação

- Qual a quantidade de colônias esperadas caso a concentração seja de 200 ppm?

$$y = -4,137 + 0,260 \cdot x$$

$$y = -4,137 + 0,260 \cdot 200$$

$$y = 48 \text{ colônias}$$



Regressão linear com a Cassio fx ou HP10s



- **MODE 3 1:** Vai para o modo de regressão linear
- **SHIFT CLR 1 = :** limpa os registradores estatísticos
- **138 , 29 M+ :** armazena o primeiro par de dados (atenção, não é o ponto decimal); deve aparecer $n=1$ no visor
- **167 , 40 M+ :** etc até $n=6$
- **SHIFT S-VAR ►► 1 = :** A
- **SHIFT S-VAR ►► 2 = :** B
- **SHIFT S-VAR ►► 3 = :** r

$$y = A + Bx$$

Exercício

- Um estudo busca estabelecer a relação entre um aditivo no combustível e a presença de fuligem nos gases de escape de um motor. Foram testadas várias concentrações do aditivo entre os valores recomendados pelo fabricante, e a fuligem avaliada de acordo com o protocolo do ensaio. Os resultados obtidos foram:

Aditivo	Fuligem
26,0	2,8
34,0	1,3
19,0	7,5
14,0	8,1
22,0	7,2
30,0	3,6
28,0	4,1

$$y = -0,3589x + 13,814; R^2 = 0,8637; 35,8 \text{ u}$$

Com base nesses valores:

- Estabeleça a correção entre a concentração do aditivo e a fuligem e interprete
- Calcule o coeficiente de determinação e interprete
- Supondo que o máximo de fuligem seja de 1,0 unidade, estime a concentração necessária de aditivo.

Fazer os
exercícios 1 a
3 da pg. 267

