REGRESSÃO LINEAR

Supomos que temos um conjunto de N pontos experimentais (x_i, y_i), $i = 1, 2, \ldots, N$, para os quais acreditamos que a melhor descrição é dada por uma função linear,

$$y = a + bx\,. \tag{1}$$

Nem sempre um conjunto de dados apresenta uma relação linear, mas supomos que este é o caso aqui. A questão é: quais os valores de a e b em (1) que nos dão a melhor função linear? Para isso definimos,

$$\chi(a,b) = \sum_{i=1}^{N}(a + bx_i - y_i)^2\,, \tag{2}$$

e calculamos a e b de modo que χ apresente o menor valor possível. Notemos que χ é uma medida de o quanto os pontos experimentais y_i se afastam da função y.

O menor valor de χ é dado por a e b que satisfazem as condições de extremo da função,

$$\begin{aligned}
\frac{\partial\chi}{\partial a} &= 2\sum_{i=1}^{N}(a + bx_i - y_i) = 0\,,\\
\frac{\partial\chi}{\partial b} &= 2\sum_{i=1}^{N}(a + bx_i - y_i)x_i = 0\,.
\end{aligned} \tag{3}$$

Definindo agora,

$$\begin{aligned}
A &= \sum_{i=1}^{N} x_i\,, \qquad B = \sum_{i=1}^{N} y_i\,,\\
C &= \sum_{i=1}^{N} x_i^2\,, \qquad D = \sum_{i=1}^{N} x_i y_i\,,
\end{aligned} \tag{4}$$

podemos escrever o sistema (3) como,

$$\begin{aligned}
Na + Ab &= B\,,\\
Aa + Cb &= D\,.
\end{aligned} \tag{5}$$

Resolvendo para a e b temos, finalmente,

$$a = \frac{BC - AD}{NC - A^2}\,, \qquad b = \frac{ND - AB}{NC - A^2}\,. \tag{6}$$

Os valores acima para a e b nos dão a função (1) que melhor representa os pontos experimentais x_i, y_i. Vamos considerar alguns exemplos simples.

(a) A taxa metabólica de alguns animais em função da massa é dada na tabela (1). Supomos que a função correspondente é da forma,

$$R = R_0 M^n \, . \tag{7}$$

Vamos calcular R_0 e n que melhor representam os pontos da tabela 1 segundo a equação (7). Notemos que essa equação não é linear (fig. 1). Contudo, se tomamos o logaritmo dos dois lados temos,

$$\log R = \log R_0 + n \log M \, . \tag{8}$$

Essa função é linear, se a equação (7) de fato é válida (fig. 2). Vamos calcular as constantes $\log R_0$ e n na expressão acima usando as equações (6). Para os pontos da tabela 1 temos,

$$\begin{aligned} &A = 3,71\,, \qquad B = 5,31\,, \\ &C = 5,35\,, \qquad D = 5,88\,, \\ &\log R_0 = 0,507\,, \qquad n = 0,748\,. \end{aligned} \tag{9}$$

Substituindo $\log R_0$ e n na equação (8) obtemos os valores de $\log R$ na tabela 2, para os valores dados de $\log M$. Vemos que os valores são bastante próximos da segunda coluna da tabela 1. Podemos então usar a equação (7), com $n \cong 0,75 = 3/4$. Essa relação é chamada em alguns livros de *regra de Kleiber*.

R(kcal/h)	$\log R$	M(kg)	$\log M$
2,5	0,4	0,7	-0,15
5,4	0,73	2,0	0,3
7,3	0,86	3,0	0,48
24,3	1,39	15,0	1,18
85,5	1,93	80,0	1,9

Tabela 1. Alguns valores experimentais para a taxa metabólica em função da massa.

$\log M$	$\log R$
-0,15	0,39
0,3	0,73
0,48	0,87
1,18	1,39
1,9	1,93

Tabela 2. Valores de $\log R$ usando (8) e (9).

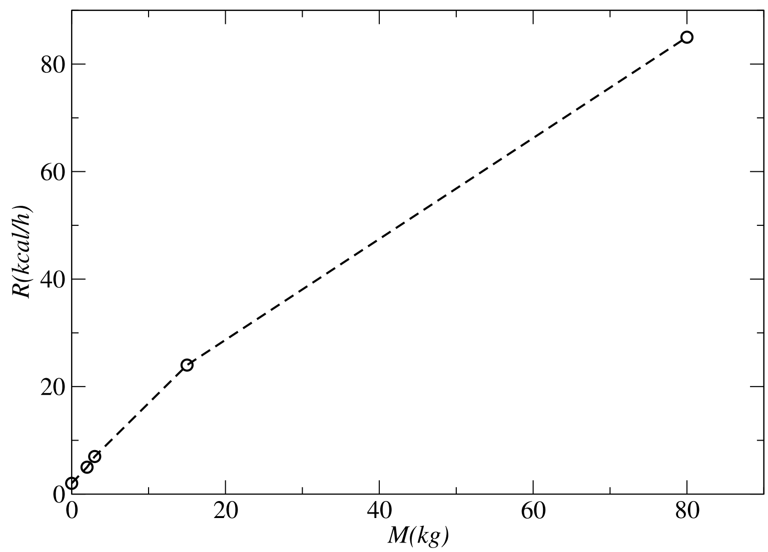

Fig. 1. Os dados da tabela 1 em um gráfico $R \times M$.

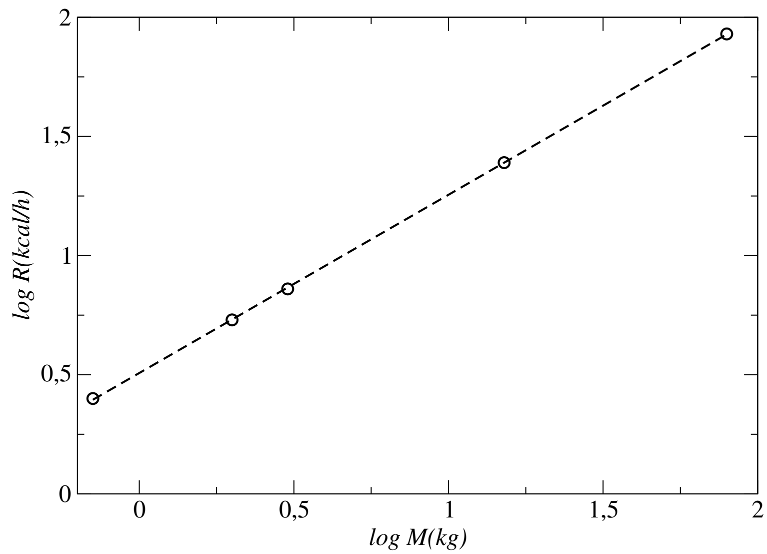

Fig. 2. Os dados da tabela 1 em um gráfico $\log R \times \log M$. A linha tracejada representa o ajuste linear.

(b) Um organismo unicelular se reproduz por divisão binária. Se inicialmente há duas bactérias e cada uma se divide em um intervalo de 20 minutos, temos o resultado mostrado na tabela 3. Fazendo um gráfico $N \times t$ temos a figura 3.

t(min)	N	$\log N$
0	2	0,301
20	4	0,602
40	8	0,903
60	16	1,204
80	32	1,505
...	...	...

Tabela 3. Número de células em função do tempo para uma divisão binária.

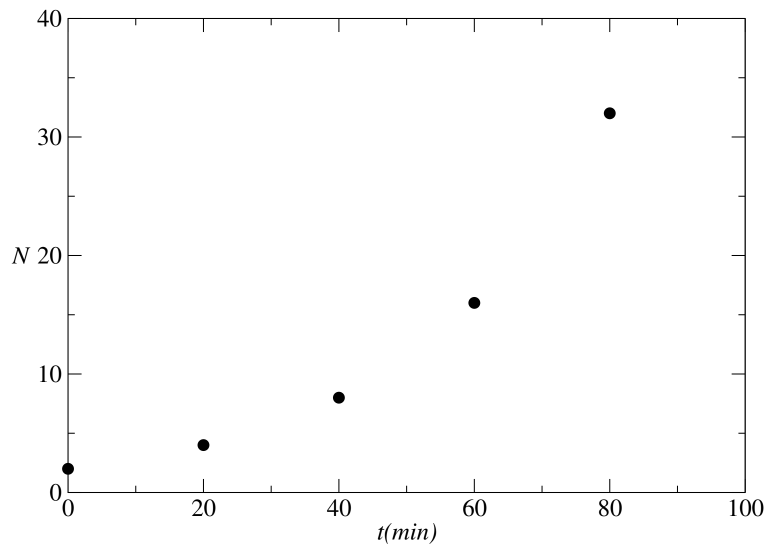

Fig. 3. Os dados da tabela 3 em um gráfico $N \times t$.

Podemos ver que há um crescimento exponencial, logo podemos escrever,

$$N(t) = N_0 e^{at}\,, \tag{10}$$

com $N_0 = N(0) = 2$ e a uma constante a ser determinada. Calculando o logaritmo decimal de ambos os lados da equação acima temos,

$$\log N = \log N_0 + (a\,\log e)t\,. \tag{11}$$

Fazendo o gráfico $\log N \times t$ obtemos uma reta (fig. 4), com inclinação $(a\,\log e)$. Não precisamos fazer uma regressão linear porque os dados na tabela 3 são exatamente da forma,

$$N(t) = N_0 2^{t/T}\,, \tag{12}$$

com $N_0 = 2$ e $T = 20$ min. Comparando (10) e (12) temos,

$$\begin{aligned}
e^{at} &= 2^{t/T}\,,\\
at &= \frac{t}{T}\ln 2\,,\\
a &= \frac{1}{T}\ln 2\,.
\end{aligned}$$

Assim, para $T = 20$ min temos $a = 0,346$. As dimensões de a são tempo^{-1}, ou frequência. A inclinação da reta em (11) é portanto $a \log e = 0,015$.

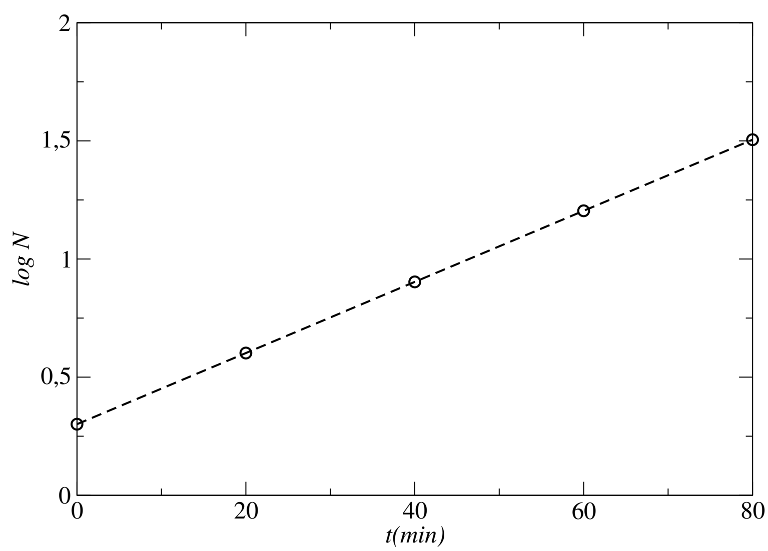

Fig. 4. Os dados da tabela 3 em um gráfico $\log N \times t$.

(c) Uma substância com átomos radioativos de tecnécio (^{99}Tc) mostra uma atividade dada na tabela 4. A atividade A de uma amostra radioativa é a taxa com que os núcleos dos átomos decaem. A figura 5 mostra esses dados.

t(h)	$A(t)$	$\log A$
0	1	0
2	0,79	-0,102
4	0,63	-0,2
6	0,50	-0,301
8	0,40	-0,398
10	0,32	-0,495
12	0,25	-0,602
...	...	...

Tabela 4. Atividade de uma amostra radioativa de tecnécio (^{99}Tc).

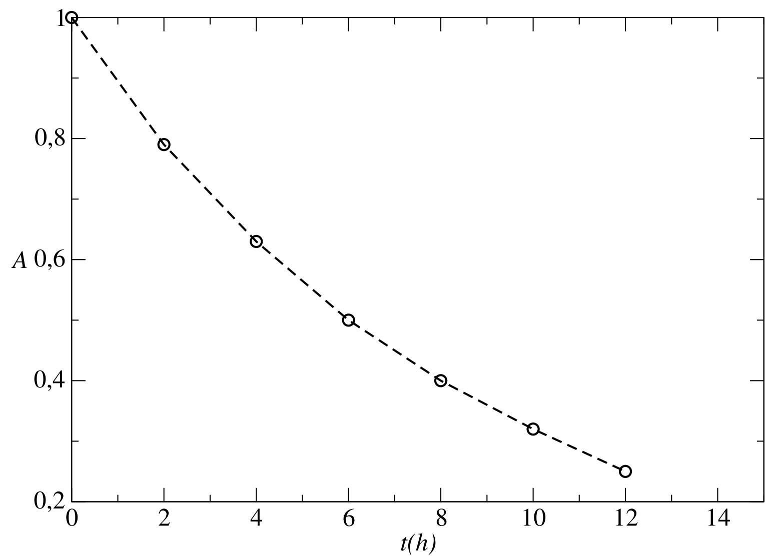

Fig. 5. Os dados da tabela 4 em um gráfico $A \times t$.

Escrevemos a relação funcional para A na forma,

$$A(t) = A_0 e^{-at} \,. \tag{13}$$

Vemos que $A_0 = 1$, vamos calcular a constante a. Calculando o logaritmo decimal da expressão acima temos,

$$\log A(t) = \log A_0 - at \log e \,. \tag{14}$$

A figura 6 mostra essa relação. Vamos fazer a regressão linear e calcular as constantes $\log A_0$ e a. Aplicando as equações (4) para os dados da tabela 4 temos (não confundir a constante A com a atividade),

$$
\begin{aligned}
&A = 43\,, \qquad B = -2,1\,, \\
&C = 364\,, \qquad D = -18,17\,, \\
&\log A_0 = -7,86 \times 10^{-4}\,, \\
&a \log e = -0,05\,, \qquad a = -0,115\,.
\end{aligned} \tag{15}
$$

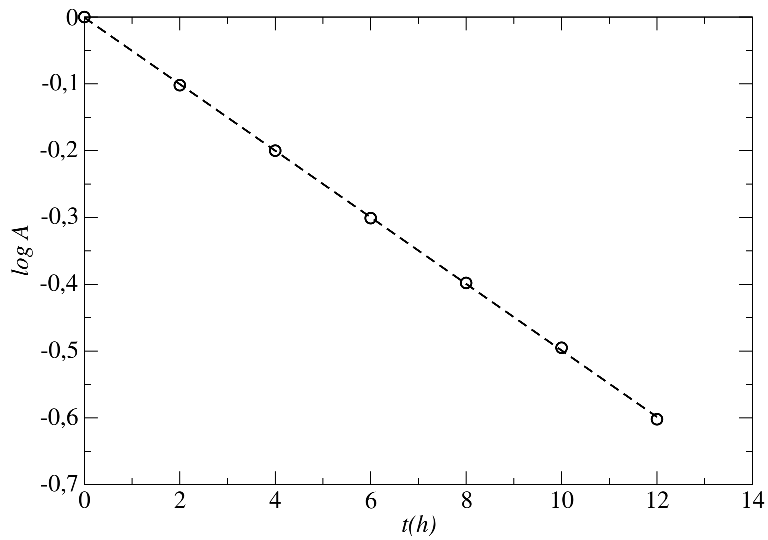

Fig. 6. Os dados da tabela 4 em um gráfico $\log A \times t$. A linha tracejada mostra o ajuste linear.