Computação Científica

prof. Marco Villaça

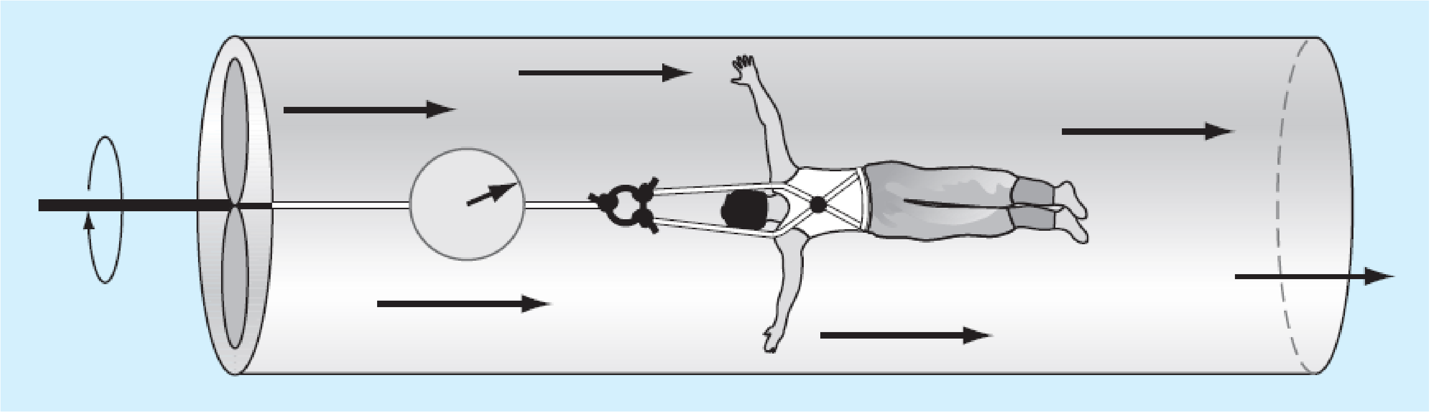
AJUSTE DE CURVAS: MÉTODO DOS MÍNIMOS QUADRADOS, MÉTODO POLINOMIAL E LINEARIZAÇÃO

PARTE I – REGRESSÃO LINEAR

Regressão Linear

O objetivo da regressão linear é determinar a “melhor” reta que se ajuste a um conjunto de dados.

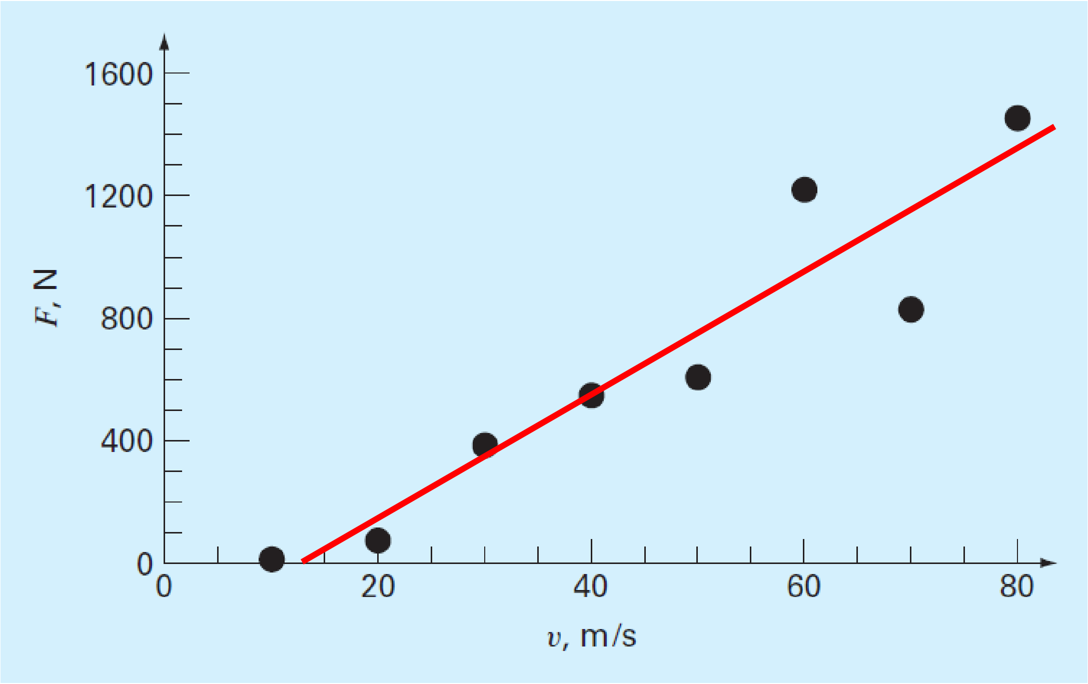
Exemplo: Experimento do túnel de vento para medir como a força da resistência do ar depende da velocidade



REGRESSÃO LINEAR

v, m/s	10	20	30	40	50	60	70	80
F, N	25	70	380	550	610	1220	830	1450

Dados experimentais para a força e a velocidade em um experimento de túnel de vento e gráfico correspondente



REGRESSÃO LINEAR

- Linear em estatística significa que a resposta deve ser uma combinação linear de variáveis explicativas.
- O problema básico é encontrar a melhor linha reta $y = a_0 + a_1 x$, dado que $i \in \{1, \ldots, N\}$ e os pares (y_i, x_i) são observados.
- O método é facilmente generalizado para encontrar o melhor ajuste da forma

$$y = a_1 f_1(x) + \cdots + a_k f_k(x)$$

- Não é necessário que as funções f_k sejam lineares em x:
  - ✓ Tudo o que é necessário é que y seja uma combinação linear dessas funções.

Método dos mínimos quadrados

- Determinar uma função aproximada que ajuste a tendência geral dos dados sem necessariamente passar pelos pontos individuais.
- O exemplo mais simples é o ajuste de uma reta a um conjunto de pares de observação $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$

$$y = a_0 + a_1 x + e \qquad \text{(i)}$$

onde: a_0 – intersecção com o eixo y;

a_1 – inclinação da reta

e = erro ou resíduo entre o modelo e a observação

Método dos mínimos quadrados

- O erro ou resíduo é, portanto, descrito por

$$e = y - (a_0 + a_1 x)$$

- O ajuste dos parâmetros pelo MMQ consiste em determinar os valores de a_0 e a_1 que minimizam a soma dos quadrados dos resíduos:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \qquad \text{(ii)}$$

Método dos mínimos quadrados

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

- Para determinar os parâmetros a_0 e a_1 deriva-se (ii) com relação a estes:

$$\frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)\, x_i$$

Método dos mínimos quadrados

- Para obter o mínimo, iguala-se as derivadas a zero:

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} a_1 x_i = 0$$

$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} a_0 x_i - \sum_{i=1}^{n} a_1 {x_i}^2 = 0$$

$$\frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)\, x_i$$

Método dos mínimos quadrados

- Prosseguindo e reorganizando:

$$na_0 + \left(\sum_{i=1}^{n} x_i\right) a_1 = \sum_{i=1}^{n} y_i \qquad \text{(iii)}$$

$$\left(\sum_{i=1}^{n} x_i\right) a_0 + \left(\sum_{i=1}^{n} {x_i}^2\right) a_1 = \sum_{i=1}^{n} x_i y_i \qquad \text{(iv)}$$

- Resolvendo o sistema para a_1 :

$$a_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \ \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad \text{(v)}$$

Método dos mínimos quadrados

- Substituindo (v) em (iii), obtém-se a_0 a partir de:

$$a_0 = \bar{y} - a_1 \bar{x} \qquad \text{(vi)}$$

onde $\bar{x}$ e $\bar{y}$ são as médias de x e y

EXEMPLO 1

Ajuste uma reta aos valores tabela ao lado.

v (m/s)	F (N)
10	25
20	70
30	380
40	550
50	610
60	1220
70	830
80	1450

EXEMPLO 1

Nesse caso F = f (v). Reorganizando os dados e totalizando os somatórios necessários, constrói-se a tabela ao lado.

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	10	25	100	250	625
2	20	70	400	1400	4900
3	30	380	900	11400	144400
4	40	550	1600	22000	302500
5	50	610	2500	30500	372100
6	60	1220	3600	73200	1488400
7	70	830	4900	58100	688900
8	80	1450	6400	116000	2102500
$\sum$	360	5135	20400	312850	5104325

EXEMPLO I

- A inclinação e a intersecção com o eixo y podem ser calculadas com as equações (v) e (vi):

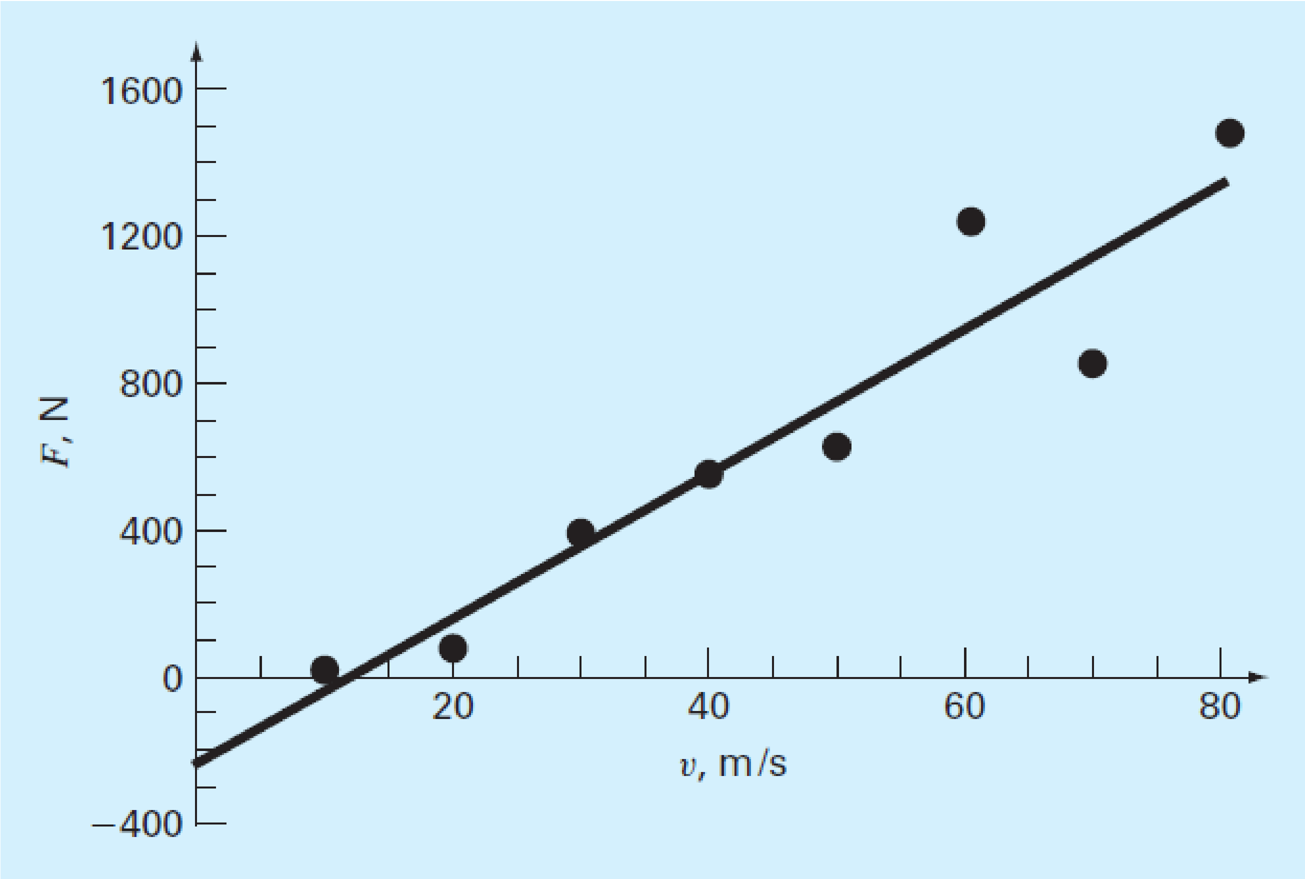
$$\boldsymbol{a_1} = \frac{\mathrm{n}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i y_i} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}\ \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y_i}}{\mathrm{n}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}^2 - \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}\right)^2} = \frac{8 \cdot 312 - 360 \cdot 5135}{8 \cdot 20400 - 360^2} = \mathbf{19{,}47024}$$

$$\boldsymbol{a_0} = \bar{y} - a_1\,\bar{x} = 641{,}875 - 19{,}47024 \cdot 45 = \mathbf{-\,234{,}2875}$$

- Utilizando F no lugar de y e v no lugar de x, o ajuste pelo MMQ é

$$\boxed{\mathrm{F} = -234{,}29 + 19{,}47\, v} \ (\text{para v} \geq 12{,}03)$$

EXEMPLO I – Ajuste de uma reta por mínimos quadrados



Quantificação do erro - Regressão linear

- A soma dos quadrados dos resíduos é definida por

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

- Similar a soma dos quadrados dos resíduos entre os pontos dados e a média

$$S_t = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

- Utilizada em estatística para calcular o desvio padrão:

$$S_y = \sqrt{\frac{S_t}{n-1}}$$

- Por analogia podemos determinar um "desvio padrão" para a reta de regressão:

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

Quantificação do erro - Regressão linear

- A diferença entre o módulo do erro residual antes da regressão

$$S_t = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

e o erro que permanece depois da regressão

$$S_r = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

normalizada, é chamada de coeficiente de determinação

$$r^2 = \frac{S_t - S_r}{S_t}$$

- Para um ajuste perfeito, $S_r = 0$ e $r^2 = 1$, ou seja a reta explica 100% da variação dos dados

Método dos mínimos quadrados

- Uma equação alternativa para r^2, conveniente para cálculo computacional é:

$$r^2 = \left(\frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \ \sum_{i=1}^{n} y_i}{\sqrt{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \sqrt{n \sum_{i=1}^{n} {y_i}^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} \right)^2$$

- Utilizando a equação acima, calcula-se o coeficiente de determinação do ajuste realizado no Exemplo 1:

$$r^2 = \left(\frac{8 \cdot 312850 - 360 \cdot 5135}{\sqrt{8 \cdot 20400 - 360^2} \sqrt{8 \cdot 5104325 - 5135^2}} \right)^2 = 0{,}8805$$

$$a_1 = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} {x_i}^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

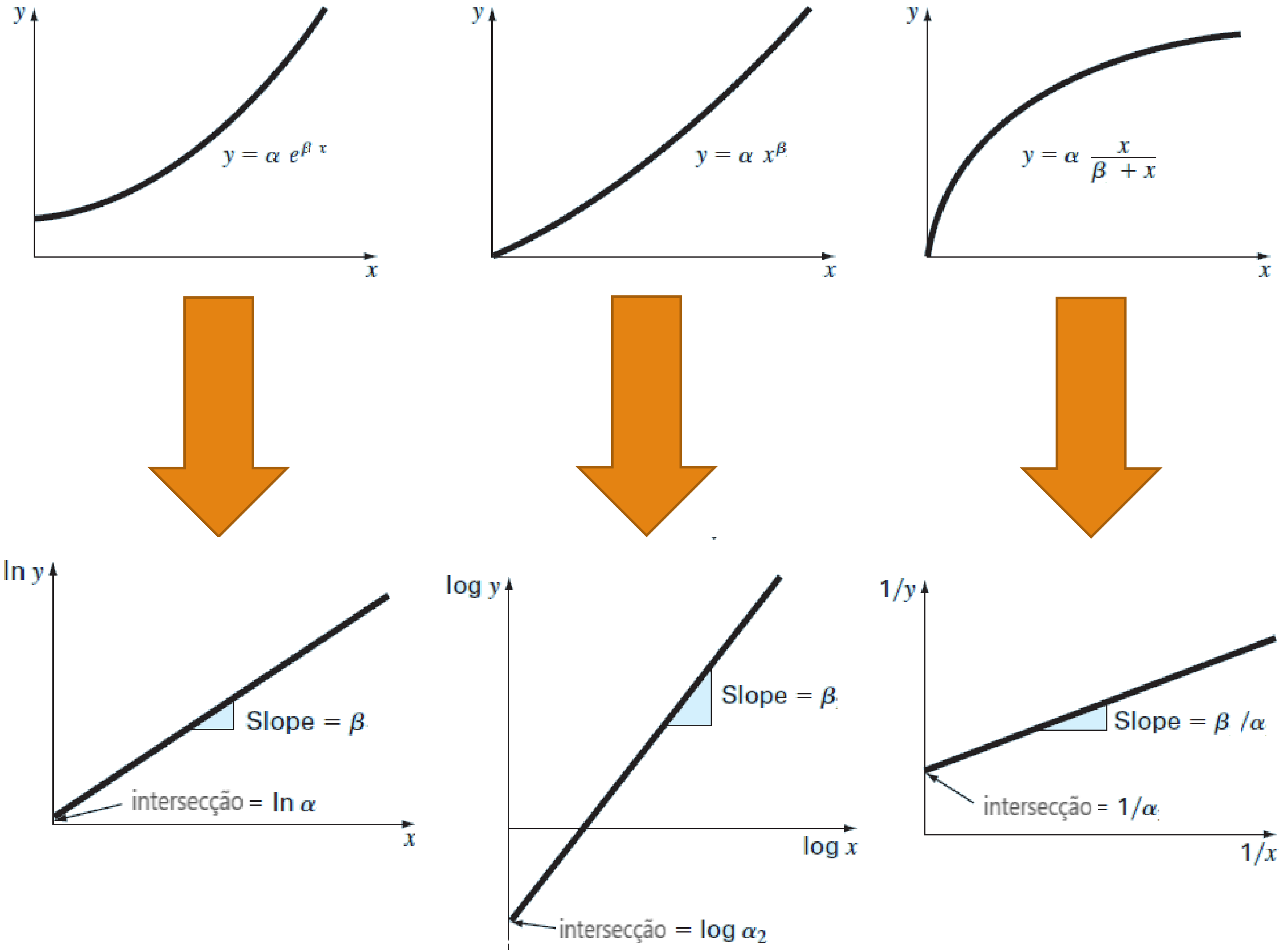
- O que indica que 88,05% da incerteza original foi explicada pelo ajuste linear.

Linearização de relações não lineares

- Nem sempre a relação entre a variável dependente e a independente é linear.
- Nesses casos, podem ser usadas transformações para expressar os dados em uma forma que seja compatível com a regressão linear:
  - ✓ Equação exponencial: $y = \alpha e^{\beta x}$
  - ✓ Equação de potência simples: $y = \alpha x^{\beta}$
  - ✓ Equação da taxa de crescimento de saturação: $y = \alpha \frac{x^m}{\beta + x^m}$

    onde m é a ordem do ajuste

Linearização de relações não lineares



Linearização de relações não lineares

- Equação exponencial:
  - $y = \alpha\, e^{\beta x}$ ➔ $\ln y = \ln \alpha + \beta x$
- Equação de potência simples:
  - $y = \alpha\, x^{\beta}$ ➔ $\log y = \log \alpha + \beta \log x$
- Equação da taxa de crescimento de saturação:
  - ✓ $y = \alpha \dfrac{x^m}{\beta + x^m}$ ➔ $\dfrac{1}{y} = \dfrac{1}{\alpha} + \dfrac{\beta}{\alpha}\dfrac{1}{x^m}$

EXEMPLO 2

Ajuste a equação $y = \alpha\, x^{\beta}$ aos dados da tabela ao lado usando uma transformação de potência simples

v (m/s)	F (N)
10	25
20	70
30	380
40	550
50	610
60	1220
70	830
80	1450

EXEMPLO 2

Nesse caso F = f (v). Reorganizando os dados e totalizando os somatórios necessários, constrói-se a tabela ao lado.

i	x_i	y_i	$log\ x_i$	$log\ y_i$	$(log\ x_i)^2$	$log\ x_i log\ y_i$	$(log\ y_i)^2$
1	10	25	1,0000	1,3979	1,0000	1,3979	1,9542
2	20	70	1,3010	1,8451	1,6927	2,4005	3,4044
3	30	380	1,4771	2,5798	2,1819	3,8107	6,6553
4	40	550	1,6021	2,7404	2,5666	4,3902	7,5096
5	50	610	1,6990	2,7853	2,8865	4,7322	7,7581
6	60	1220	1,7782	3,0864	3,1618	5,4880	9,5256
7	70	830	1,8451	2,9191	3,4044	5,3860	8,5210
8	80	1450	1,9031	3,1614	3,6218	6,0164	9,9942
$\sum$	360	5135	12,606	20,515	20,516	33,622	55,322

EXEMPLO 2

- A inclinação e a intersecção com o eixo y podem ser calculadas com as equações (iii) e (v):

$$\boldsymbol{a_1} = \frac{\mathrm{n}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i y_i} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y_i}}{\mathrm{n}\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}^2 - \left(\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i}\right)^2} = \frac{8 \cdot 33{,}622 - 12{,}606 \cdot 20{,}515}{8 \cdot 20{,}516 - 12{,}606^2} = \mathbf{1{,}9842}$$

$$\boldsymbol{a_0} = \bar{y} - a_1\,\bar{x} = 2{,}5644 - 1{,}9842 \cdot 1{,}5757 = \mathbf{-0{,}5620}$$

- Utilizando F no lugar de y e v no lugar de x, o ajuste pelo MMQ é

$$\log \mathrm{F} = -0{,}5620 + 1{,}9842 \log \mathrm{v}$$

$$y = \alpha\, x^{\beta} \quad \rightarrow \quad \log y = \log \alpha + \beta \log x$$

EXEMPLO 2

- Para exibir o ajuste usando coordenadas não transformadas:

$$a_0 = log\, \alpha \quad \rightarrow \quad \alpha = 10^{a_0}$$

$$a_1 = \beta$$

- Substituindo os valores:

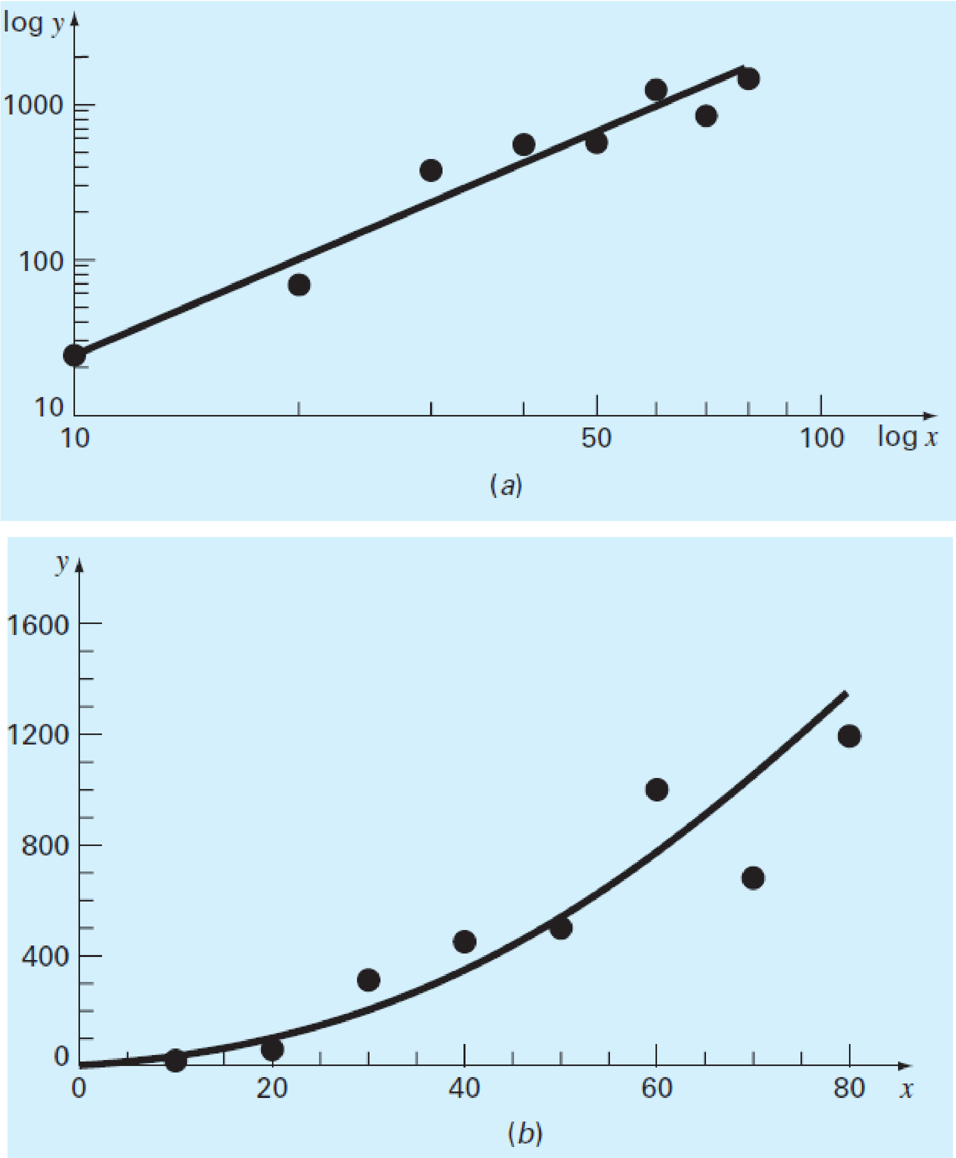
$$\alpha = 10^{-0,562} = 0,2741$$

$$a_1 = 1,9842$$

- O ajuste por mínimos quadrados é

$$F = 0,27\, v^{1,98}$$

com coeficiente de determinação (calcule) $r^2 = 0,9481$



EXEMPLO 2 – Ajuste com equação de potência simples

A Figura ao lado representa o ajuste por mínimos quadrados de um modelo de potência simples aos dados sendo:

(a) o ajuste dos dados transformados e

(b) o ajuste da equação de potência juntamente com os dados

EXEMPLO 3

A tabela ao lado contém os registros efetuados dos valores médios da radiação solar em uma localidade para alguns meses do ano.

Ajuste o modelo

$$M(x) \;=\; a_1\,(x-6)^4 \;+\; a_0\,sen\left(\frac{\pi}{12}x\right)$$

aos valores da tabela, no sentido dos mínimos quadrados, e use o modelo encontrado para prever a radiação média no mês de Agosto.

Um langley (Ly) é uma unidade de distribuição de energia por área Um langley é definido como uma caloria termoquímica por cm². Em unidades SI, 1 langley vale 41,840 J/m².

Mês	Radiação (Ly/dia)
1	122
3	188
6	270
10	160
12	120

EXEMPLO 3

- Olhando para a prova do Método dos Mínimos Quadrados, percebe-se que não é essencial que $y = a_1 x + a_o$; poderia ser $y = a_1 f(x) + a_o g(x)$. Nesse caso:

$$a_1 = \frac{\sum_{i=1}^{n} g(x_i)^2 \sum_{i=1}^{n} f(x_i) y_i - \sum_{i=1}^{n} f(x_i) g(x_i) \sum_{i=1}^{n} g(x_i) y_i}{\sum_{i=1}^{n} f(x_i)^2 \sum_{i=1}^{n} g(x_i)^2 - \left(\sum_{i=1}^{n} f(x_i) g(x_i)\right)^2}$$

$$a_0 = \frac{\sum_{i=1}^{n} g(x_i) y_i - a_1 \sum_{i=1}^{n} f(x_i) g(x_i)}{\sum_{i=1}^{n} g(x_i)^2}$$

$$r^2 = \left(\frac{\sum_{i=1}^{n} g(x_i)^2 \sum_{i=1}^{n} f(x_i) y_i - \sum_{i=1}^{n} f(x_i) g(x_i) \sum_{i=1}^{n} g(x_i) y_i}{\sqrt{\sum_{i=1}^{n} f(x_i)^2 \sum_{i=1}^{n} g(x_i)^2 - \left(\sum_{i=1}^{n} f(x_i) g(x_i)\right)^2} \sqrt{\sum_{i=1}^{n} g(x_i)^2 \sum_{i=1}^{n} {y_i}^2 - \left(\sum_{i=1}^{n} g(x_i) y_i\right)^2}}\right)^2$$

EXEMPLO 3

Trata-se de um modelo linear não polinomial, com

$$f(x) = (x-6)^4 \qquad \text{e} \qquad g(x) = sen\left(\frac{\pi}{12}x\right)$$

Reorganizando os dados e totalizando os somatórios necessários, constrói-se a tabela abaixo.

x_i	y_i	$f(x_i)$	$g(x_i)$	$f(x_i)^2$	$g(x_i)^2$	$f(x_i)g(x_i)$	$f(x_i)\,y_i$	$g(x_i)\,y_i$	${y_i}^2$
1	122	**625**	0,25882	390625	0,0670	161,7623	76250	31,5760	14884
3	188	**81**	0,70711	6561	0,5000	57,2758	15228	132,9363	35344
6	270	**0**	1	0	1	0	0	270	72900
10	160	**256**	0,49999	65536	0,2500	127,9986	40960	79,9992	25600
12	120	**1296**	-0,00001	1679616	0,0000	-0,0095	155520	-0,0009	14400
$\sum$		2258	2,4659	2142338	1,8170	347,0271	287958	514,5106	163128

EXEMPLO 3

- Calculando a inclinação e a intersecção com o eixo y, resulta :

$$a_1 = 0{,}0914$$

$$a_0 = 265{,}7164$$

- E o modelo obtido é

$$M(x) = 0{,}0914\,(x-6)^4 + 265{,}7\,sen\left(\frac{\pi}{12}x\right)$$

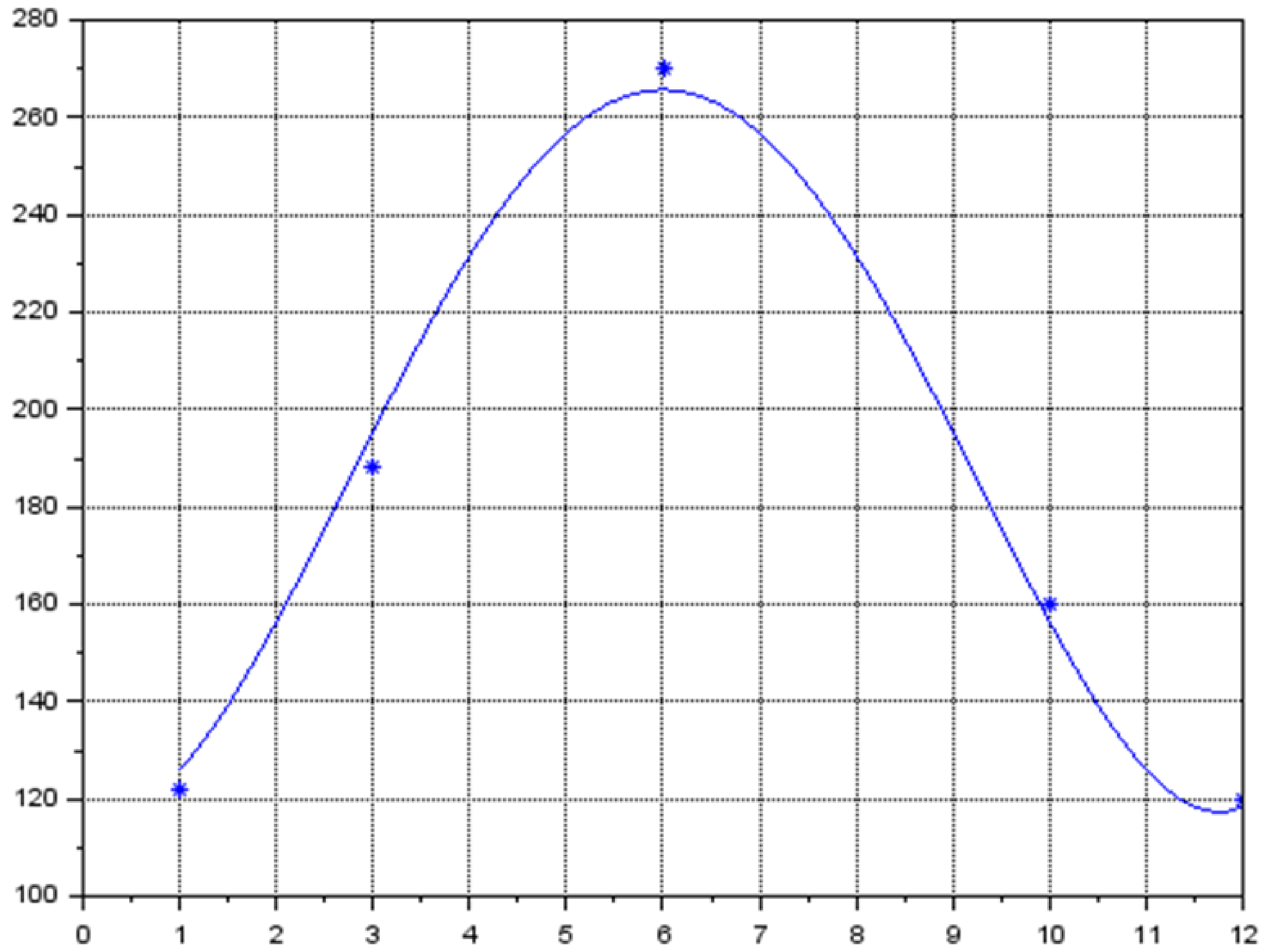
com coeficiente de determinação $r^2 = 0{,}9941$.

- Com $x = 8$, obtém-se a radiação média no mês de agosto:

$$M(x) = 0{,}0914\,(2)^4 + 265{,}7\,sen\left(\frac{2\pi}{3}\right) = 231{,}7\ Ly/dia$$

EXEMPLO 3 – Modelo MMQ generalizado

A Figura ao lado representa o ajuste por mínimos quadrados com o modelo generalizado juntamente com os dados



1) A concentração de bactéria E. Coli em uma área de prática de natação é monitorada após uma tempestade:

t *(h)*	4	8	12	16	20	24
c(*UFC/100 ml*)	1600	1320	1000	890	650	560

O tempo é medido em horas seguindo o fim da tempestade, e a unidade UFC é uma Unidade Formadora de Colônia. Use esses dados para estimar (a) a concentração no fim da tempestade t = 0 e (b) o instante de tempo em que a concentração alcança 200 *UFC/100 ml*.

Exercícios

2) Um pesquisador relatou os dados tabulados a seguir para uma experiência a fim de determinar a taxa de crescimento da bactéria k (por dia) como uma função da concentração de oxigênio c (*mg/l*). Sabe-se que tais dados podem ser modelados pela seguinte equação:

$$y = k_m \frac{c^2}{c_s + c^2}$$

onde c_s e k_m são parâmetros Use uma transformação para linearizar esta equação. A seguir use a regressão linear para estimar c_s e k_m e prever a taxa de crescimento em c = 2 *mg/l* .

c	0,5	0,8	1,5	2,5	4
k	1,1	2,5	5,3	7,6	8,9

Exercícios

Função Scilab

Função Scilab

```
function [alfa, betha, r]=reglinear(x, y, e, m)
/* x vetor com os pontos da variável independente
 y vetor com os pontos da variávei dependente
 e: 'r' - reta, 'e' - exp, 'p' - potencia, 's' - saturação
 se for s tem que passar a potencia m */
   a=[0 0];
   n = length(x)
   if n~=length(y) then
      error("inconsistencia nos dados");
   end
   plot(x,y,'o')
   xx = linspace(min(x),max(x),100);
```

Função Scilab

```
if (e == 'e') then
   y = log(y);
elseif (e == 'p') then
   x = log10(x);
   y = log10(y);
elseif (e == 's') then
   x = 1 ./ x .^m;
   y = 1 ./ y;
elseif (e == 'r')
   // não necessita ajuste nos eixos
else
   error("opção inexistente");
end
```

Função Scilab

```
somax=sum(x); somay=sum(y); somaxy=sum(x.*y);

somaxx=sum(x.^2); somayy=sum(y.^2);

A = n*somaxy - somax*somay;

B = n*somaxx - somax^2;

C = n*somayy - somay^2;

a(2) = A/B;     // a1

a(1) = somay/n - a(2)*somax/n;  // a0

r = (A/(sqrt(B)*sqrt(C)))^2;
```

Função Scilab

```
if (e == 'e') then
   betha = a(2);
   alfa = exp(a(1));
   yy =alfa*exp(betha*xx);
elseif (e == 'p') then
   betha = a(2);
   alfa = 10^(a(1))
   yy = alfa*xx.^betha
elseif (e == 's') then
   alfa = 1/a(1);
   betha = alfa*a(2);
   yy = alfa*xx.^m ./ (betha + xx.^m);
```

Função Scilab

```
    elseif (e == 'r')
        yy = a(1)+a(2)*xx;
        alfa = a(1);
        betha = a(2);
    end
    plot(xx,yy)
    xgrid;
endfunction
```

Teste a sua função conferindo com os resultados obtidos nos Exemplos 1 e 2 e nos Exercícios 1 e 2.

Bibliografia e crédito das figuras

CHAPRA, Steven. **Applied numerical methods with MATHLAB for engineers and scientists.** McGrawHill, 2012.

CHAPRA, Steven e CANALE, Raymond. **Numerical methods for engineers.** McGrawHill, 2010.