

CCI-22



Matemática Computacional

Carlos Alberto Alonso Sanches
Juliana de Melo Bezerra

CCI-22



6) Ajuste de Curvas

Método dos Mínimos Quadrados, Regressão Linear

CCI-22



- Introdução
- Método dos Mínimos Quadrados
- Ajuste a um polinômio
- Regressão linear
- Ajuste a outras curvas
- Qualidade do ajuste

CCI-22



- **Introdução**
- Método dos Mínimos Quadrados
- Ajuste a um polinômio
- Regressão linear
- Ajuste a outras curvas
- Qualidade do ajuste

Definição

- Situações em que a interpolação não é aconselhável:
 - Quando se deseja obter um valor aproximado da função em algum ponto fora do intervalo de tabelamento, ou seja, quando se quer extrapolar
 - Quando os valores tabelados são resultados de algum experimento físico ou de alguma pesquisa, e por isso podem conter erros inerentes e não previsíveis
- Nesses casos, convém ajustar a função tabelada a uma função f^* que seja uma "boa aproximação" para os valores obtidos, e que permita a extrapolação

Possíveis critérios de qualidade

- Dados m pontos experimentais $\{(x_1, y_1), \dots, (x_m, y_m)\}$, uma questão importante é estabelecer uma medida de qualidade para a função de ajuste f^*
- Sejam $R_i = f^*(x_i) - y_i$, $1 \leq i \leq m$, os resíduos ou erros desse ajuste. Possibilidades:
 - Fazer com que os resíduos tendam a zero
 - Equivaleria à interpolação...
 - Minimizar a soma dos resíduos
 - Não é um bom critério: pode haver soma nula, mas com valores grandes
 - Minimizar a soma dos módulos dos resíduos
 - É difícil encontrar seu mínimo, pois não é diferenciável no ponto zero
 - Critério de Tschebycheff: minimizar $\max \{|R_i|\}$
 - Solução difícil e não recomendada para cálculos manuais
 - Critério dos mínimos quadrados: minimizar $\sum_i R_i^2$
 - É o mais largamente utilizado

CCI-22



- Introdução
- Método dos Mínimos Quadrados
- Ajuste a um polinômio
- Regressão linear
- Ajuste a outras curvas
- Qualidade do ajuste

Método dos Mínimos Quadrados

- A técnica dos Mínimos Quadrados consiste no cálculo de constantes c_j , $0 \leq j \leq n$, de uma função f^* que aproxima f :
 - $f^*(x) = c_0\Phi_0(x) + c_1\Phi_1(x) + \dots + c_n\Phi_n(x)$
- As funções $\Phi_j(x)$, $0 \leq j \leq n$, que podem ser não lineares em x , são escolhidas de acordo com a natureza dos dados experimentais
- Considerando os resíduos $R_i = f^*(x_i) - y_i$, $1 \leq i \leq m$, seja $R = \sum_i R_i^2 = \sum_i (c_0\Phi_0(x_i) + c_1\Phi_1(x_i) + \dots + c_n\Phi_n(x_i) - y_i)^2$
- R é uma função dos c_j , $0 \leq j \leq n$, e passará por um mínimo quando suas $n+1$ derivadas parciais se anularem simultaneamente:
 - $\partial R / \partial c_j = 2 \sum_i [f^*(x_i) - y_i] \cdot \partial f^*(x_i) / \partial c_j = 0$, $0 \leq j \leq n$
- Como $\partial f^*(x_i) / \partial c_j = \Phi_j(x_i)$, isso é equivalente a:
 - $\sum_i (c_0\Phi_0(x_i) + c_1\Phi_1(x_i) + \dots + c_n\Phi_n(x_i) - y_i) \cdot \Phi_j(x_i) = 0$, $0 \leq j \leq n$

Equações normais

- Condições para que $R = \sum_i R_i^2$ seja mínimo:
 - $\sum_i (c_0 \Phi_0(x_i) + c_1 \Phi_1(x_i) + \dots + c_n \Phi_n(x_i) - y_i) \cdot \Phi_j(x_i) = 0, 0 \leq j \leq n$
 - $\sum_i (c_0 \Phi_0(x_i) + c_1 \Phi_1(x_i) + \dots + c_n \Phi_n(x_i)) \cdot \Phi_j(x_i) = \sum_i (y_i \Phi_j(x_i)), 0 \leq j \leq n$
- Temos então um sistema $Ac = y$ de $n+1$ equações lineares, comumente chamadas de *equações normais*, que pode ser resolvido com técnicas já apresentadas:

$$A = \begin{bmatrix} \sum_i \Phi_0(x_i) \Phi_0(x_i) & \sum_i \Phi_0(x_i) \Phi_1(x_i) & \dots & \sum_i \Phi_0(x_i) \Phi_n(x_i) \\ \sum_i \Phi_1(x_i) \Phi_0(x_i) & \sum_i \Phi_1(x_i) \Phi_1(x_i) & \dots & \sum_i \Phi_1(x_i) \Phi_n(x_i) \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i \Phi_n(x_i) \Phi_0(x_i) & \sum_i \Phi_n(x_i) \Phi_1(x_i) & \dots & \sum_i \Phi_n(x_i) \Phi_n(x_i) \end{bmatrix} \quad c = \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} \quad y = \begin{bmatrix} \sum_i y_i \Phi_0(x_i) \\ \sum_i y_i \Phi_1(x_i) \\ \vdots \\ \sum_i y_i \Phi_n(x_i) \end{bmatrix}$$

- $\Phi_0(x), \Phi_1(x), \dots, \Phi_n(x)$ também são escolhidas de tal modo que $\det(A) \neq 0$, pois assim o sistema terá solução única

CCI-22



- Introdução
- Método dos Mínimos Quadrados
- **Ajuste a um polinômio**
- Regressão linear
- Ajuste a outras curvas
- Qualidade do ajuste

Ajuste a um polinômio

- Vamos ajustar a curva f a um polinômio de grau n :
 $f^*(x) = a_0 + a_1x + \dots + a_nx^n$
- Originariamente, $f^*(x) = c_0\Phi_0(x) + c_1\Phi_1(x) + \dots + c_n\Phi_n(x)$.
Portanto, podemos definir:
 - $c_i = a_i, 0 \leq i \leq n$
 - $\Phi_i(x) = x^i, 0 \leq i \leq n$
- Desse modo, teremos o seguinte sistema:

$$\begin{bmatrix} m & \sum_i x_i & \sum_i x_i^2 & \sum_i x_i^3 & \dots & \sum_i x_i^n \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i^3 & \sum_i x_i^4 & \dots & \sum_i x_i^{n+1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_i x_i^n & \sum_i x_i^{n+1} & \sum_i x_i^{n+2} & \sum_i x_i^{n+3} & \dots & \sum_i x_i^{2n} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \\ \vdots \\ \sum_i x_i^n y_i \end{bmatrix}$$

Exemplo 1

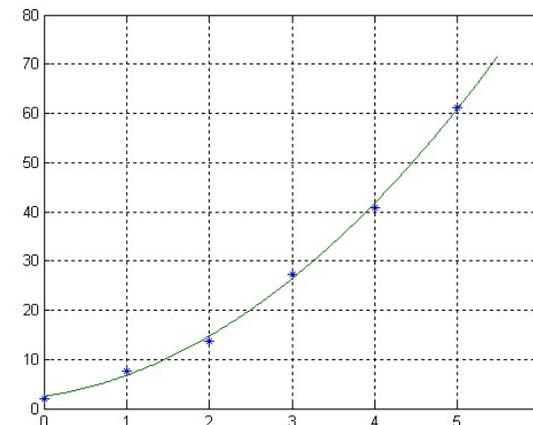
- Ajuste um polinômio de segundo grau aos dados abaixo:

x	0	1	2	3	4	5
y	2,1	7,7	13,6	27,2	40,9	61,1

- A partir desses dados, construímos o seguinte sistema:

$$\begin{bmatrix} m & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix} \Rightarrow \begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 152,6 \\ 585,6 \\ 2488,8 \end{bmatrix}$$

- $m = 6$
- $\sum x_i = 15$; $\sum y_i = 152,6$;
 $\sum x_i^2 = 55$; $\sum x_i^3 = 225$;
 $\sum x_i^4 = 979$; $\sum x_i y_i = 585,6$;
 $\sum x_i^2 y_i = 2488,8$
- $a_0 = 2,47857$; $a_1 = 2,35929$;
 $a_2 = 1,86071$



Exemplo 2

- Os dados abaixo correspondem ao volume do álcool anídrico em função da temperatura. Considerando um volume inicial de 1cm^3 a 0°C , deseje-se uma tabela do volume para temperaturas entre 20 e 40°C

t ($^\circ\text{C}$)	13,9	43,0	67,8	89,0	99,2
v (cm^3)	1,04	1,12	1,19	1,24	1,27

- Ajustaremos $v(t)$ a um polinômio de grau 2 $\overset{a_0}{\curvearrowright}$
- Considerando o volume inicial, temos $v^*(t) = 1 + a_1t + a_2t^2$
- Sistema de equações normais para as demais constantes:

$$\begin{bmatrix} \sum_i t_i^2 & \sum_i t_i^3 \\ \sum_i t_i^3 & \sum_i t_i^4 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum_i t_i v_i - \sum_i t_i \\ \sum_i t_i^2 v_i - \sum_i t_i^2 \end{bmatrix} \Rightarrow \begin{bmatrix} 24400,69 & 2,0750189 \cdot 10^6 \\ 2,0750189 \cdot 10^6 & 1,841675 \cdot 10^8 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 66,142 \\ 56610202 \end{bmatrix}$$

$$a_1 = 0,003068189$$

$$a_2 = 1,548545 \cdot 10^{-7}$$

t	13,9	20	25	30	35	40	43,0	67,8	89,0	99,2
v	1,04						1,12	1,19	1,24	1,27
v^*	1,04	1,06	1,08	1,09	1,11	1,12	1,13	1,21	1,27	1,31

CCI-22



- Introdução
- Método dos Mínimos Quadrados
- Ajuste a um polinômio
- **Regressão linear**
- Ajuste a outras curvas
- Qualidade do ajuste

Regressão linear

- *Regressão linear* é o caso particular em que a curva f é ajustada a uma reta $f^*(x) = a_0 + a_1x$

- Portanto, o sistema linear torna-se:
$$\begin{bmatrix} m & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{bmatrix}$$

- Pela regra de Cramer:

$$a_0 = \frac{\sum_i x_i^2 \cdot \sum_i y_i - \sum_i x_i \cdot \sum_i x_i \cdot y_i}{m \sum_i x_i^2 - (\sum_i x_i)^2}$$

$$a_1 = \frac{m \sum_i x_i \cdot y_i - \sum_i x_i \cdot \sum_i y_i}{m \sum_i x_i^2 - (\sum_i x_i)^2}$$

- Desde que o denominador não seja nulo, esta solução é sempre definida
- Demonstra-se que $m \sum_i x_i^2 - (\sum_i x_i)^2 = (\sum_i \sum_k (x_i - x_k)^2) / 2$
- Portanto, se os pontos x_i são distintos, a_0 e a_1 são únicos
- As expressões de a_0 e a_1 podem ser reescritas:

$$a_1 = \frac{\sum_i x_i \cdot y_i - (\sum_i x_i \cdot \sum_i y_i) / m}{\sum_i x_i^2 - (\sum_i x_i)^2 / m}$$

$$a_0 = y^* - a_1 x^* \\ y^* = \sum_i y_i / m \quad x^* = \sum_i x_i / m$$

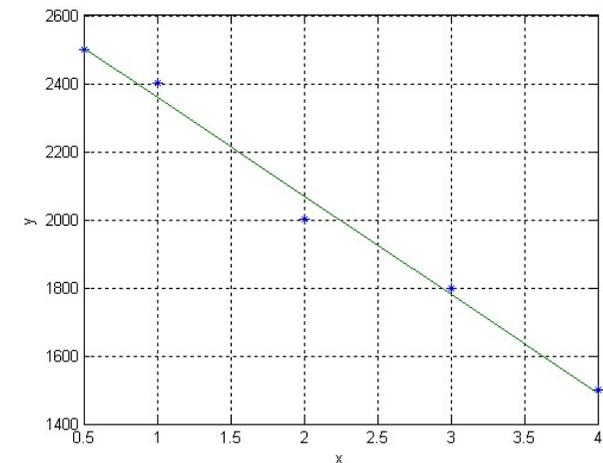
Exemplo

- A tabela abaixo mostra o desempenho de um torno de parafusos em função do seu tempo de uso. Fazer a projeção para 5 e 6 anos:

x (anos)	0,5	1	2	3	4
y (parafusos/dia)	2500	2400	2000	1800	1500

- Através de uma análise gráfica, é possível constatar que uma reta é um bom ajuste:

- $m = 5$
- $\sum_i x_i = 10,5$; $\sum_i y_i = 10200$; $\sum_i x_i y_i = 19050$;
 $\sum_i x_i^2 = 30,25$; $(\sum_i x_i)^2 = 110,25$
- $a_1 = (\sum_i x_i \cdot y_i - (\sum_i x_i \cdot \sum_i y_i) / m) / (\sum_i x_i^2 - (\sum_i x_i)^2 / m) =$
 $-2370 / 8,2 = -289,0244$
- $y^* = \sum_i y_i / m = 2040$
- $x^* = \sum_i x_i / m = 2,1$
- $a_0 = y^* - a_1 x^* = 2646,9511$
- $y = 2646,9511 - 289,0244x$
- $x = 5 \Rightarrow y = 1202$
- $x = 6 \Rightarrow y = 913$



Regressão linear múltipla

- É possível estender a regressão linear para o caso de funções lineares de múltiplas variáveis: $f(x_1, x_2, x_3, \dots, x_n)$
- Veremos um caso como exemplo, onde $n=2$:
 - $f^*(x_1, x_2) = a_0 + a_1x_1 + a_2x_2$
- Definição da função resíduo:
 - $R = \sum_i R_i^2 = \sum_i (a_0 + a_1x_{1i} + a_2x_{2i} - y_i)^2$
- Para que R seja mínimo, é necessário que:
 - $\partial R / \partial a_0 = 2 \sum_i (a_0 + a_1x_{1i} + a_2x_{2i} - y_i) = 0$
 - $\partial R / \partial a_1 = 2 \sum_i x_{1i} (a_0 + a_1x_{1i} + a_2x_{2i} - y_i) = 0$
 - $\partial R / \partial a_2 = 2 \sum_i x_{2i} (a_0 + a_1x_{1i} + a_2x_{2i} - y_i) = 0$

Regressão linear múltipla

- Temos então um sistema linear com três incógnitas (a_0 , a_1 e a_2) e três equações
- Este sistema pode ser escrito na forma matricial abaixo:

$$\begin{bmatrix} m & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} \cdot x_{2i} \\ \sum x_{2i} & \sum x_{1i} \cdot x_{2i} & \sum x_{2i}^2 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i} y_i \\ \sum x_{2i} y_i \end{bmatrix}$$

- De modo análogo ao já visto, este sistema pode ser resolvido por algum método numérico
- Determina-se assim o plano que ajusta os pontos tridimensionais

CCI-22



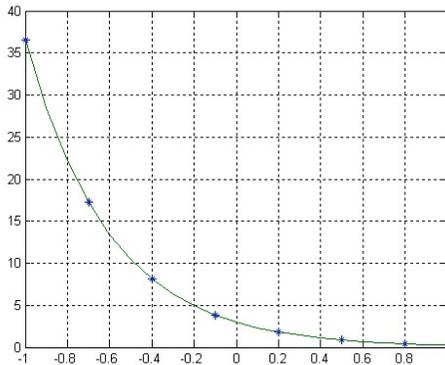
- Introdução
- Método dos Mínimos Quadrados
- Ajuste a um polinômio
- Regressão linear
- **Ajuste a outras curvas**
- Qualidade do ajuste

Ajuste à curva exponencial

- Também é possível ajustar f a uma curva exponencial, fazendo-se antes uma troca de variáveis:
 - $f^*(x) = c_1 e^{kx}$, onde c_1 e k são constantes
 - $\ln f^*(x) = \ln c_1 + kx$
 - $z^*(x) = c_2 + kx$, onde $z^*(x) = \ln f^*(x)$
 - z^* e x estão linearmente relacionadas: basta resolver a regressão linear
 - Depois de resolvido o sistema correspondente, volta-se ao problema original

Exemplo

x	-1,0	-0,7	-0,4	-0,1	0,2	0,5	0,8	1,0
f(x)	36,547	17,264	8,155	3,852	1,820	0,860	0,406	0,246

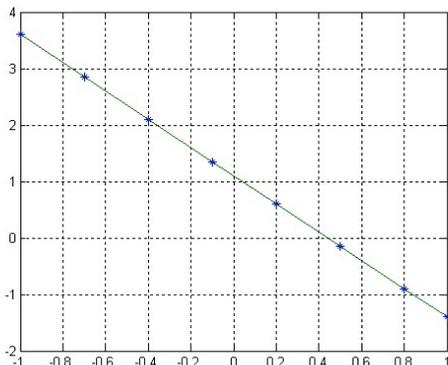


A dispersão dos dados sugere um ajuste à curva exponencial

$$z^*(x) = c_2 + kx,$$

onde $z^*(x) = \ln f^*(x)$

x	-1,0	-0,7	-0,4	-0,1	0,2	0,5	0,8	1,0
z(x) = ln f(x)	3,599	2,849	2,099	1,349	0,599	-0,151	-0,901	-1,402



$$\begin{bmatrix} m & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{bmatrix} \begin{bmatrix} c_2 \\ k \end{bmatrix} = \begin{bmatrix} \sum_i z_i \\ \sum_i z_i x_i \end{bmatrix}$$



$$\begin{bmatrix} 8 & 0,3 \\ 0,3 & 3,59 \end{bmatrix} \begin{bmatrix} c_2 \\ k \end{bmatrix} = \begin{bmatrix} 8,041 \\ -8,646 \end{bmatrix}$$

$$c_2 = 1,099$$

$$k = -2,5$$

$$\ln f^*(x) = 1,099 - 2,5x$$

$$f^*(x) = 3,001e^{-2,5x}$$

Ajuste a outras curvas

- $f^*(x) = ax^b$
 - $\ln f^*(x) = \ln a + b \cdot \ln x$
 - Sejam $z^*(x) = \ln f^*(x)$ e $t = \ln x$
 - Portanto, $z^*(x) = \ln a + bt$
 - z^* e t estão linearmente relacionadas
- $f^*(x) = ab^x$
 - $\ln f^*(x) = \ln a + x \cdot \ln b$
 - Seja $z^*(x) = \ln f^*(x)$
 - Portanto, $z^*(x) = \ln a + x \cdot \ln b$
 - z^* e x estão linearmente relacionadas

CCI-22



- Introdução
- Método dos Mínimos Quadrados
- Ajuste a um polinômio
- Regressão linear
- Ajuste a outras curvas
- Qualidade do ajuste

Qualidade da regressão linear

- Quanto melhor for a qualidade da regressão linear, menor será o valor do resíduo R , onde $R = \sum_i R_i^2 = \sum_i (f^*(x_i) - y_i)^2$
- Vamos definir o *resíduo em relação à média dos pontos experimentais* : $R_M = \sum_i (\hat{y} - y_i)^2$, onde $\hat{y} = \sum_i y_i / m$
- O valor $R_M - R$ quantifica a redução de erro decorrente da descrição dos dados em termos de uma reta, em vez de um ponto médio
- $(R_M - R) / R_M$ é o valor normalizado dessa redução
- O coeficiente de correlação r é definido como:

Critério relativo,
válido para
qualquer ajuste

$$r = \sqrt{\frac{R_M - R}{R_M}}$$

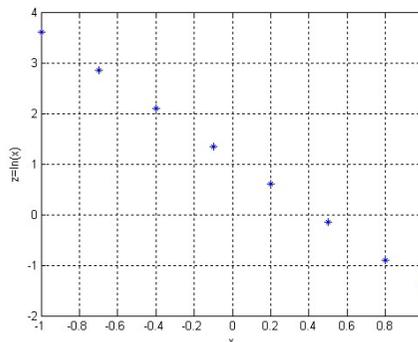
Critério absoluto,
mas usado apenas
para regressão linear

- Em um ajuste linear perfeito, $r = 1$ (pois $R = 0$)
- Portanto, quanto mais próximo de 1 for o coeficiente de correlação, melhor será o ajuste da regressão linear

Teste de alinhamento

- Há uma maneira simples de averiguar se o ajuste de uma função não linear tem boa qualidade:
 - Nos m pontos tabelados $(x_1, y_1), \dots, (x_m, y_m)$, fazer trocas de variáveis, de modo a que passem a obedecer a uma relação linear
 - Fazer o diagrama de dispersão desses novos dados
 - Verificar o alinhamento dos pontos
- Exemplo:

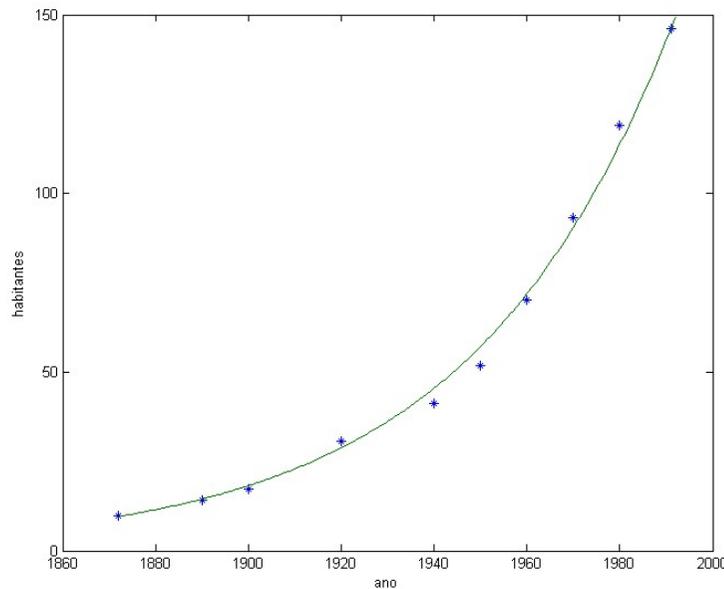
x	-1,0	-0,7	-0,4	-0,1	0,2	0,5	0,8	1,0
$y = f(x)$	36,547	17,264	8,155	3,852	1,820	0,860	0,406	0,246
$z(x) = \ln f(x)$	3,599	2,849	2,099	1,349	0,599	-0,151	-0,901	-1,402



Exemplo 1

- Ajuste uma curva à tabela abaixo, que fornece dados da evolução da população brasileira em milhões de habitantes:

x = ano	1872	1890	1900	1920	1940	1950	1960	1970	1980	1991
y = habitantes	9,9	14,3	17,4	30,6	41,2	51,9	70,2	93,1	119,0	146,2



- Suponhamos que a melhor curva seja uma exponencial
- Encontramos como resultado $y^* = a.e^{bx}$, onde $a = 2,3111.10^{-18}$ e $b = 0,0229$
- Fazendo as correspondentes trocas de variáveis para linearizar a relação, obtemos o coeficiente de correlação $r = 0,99775$

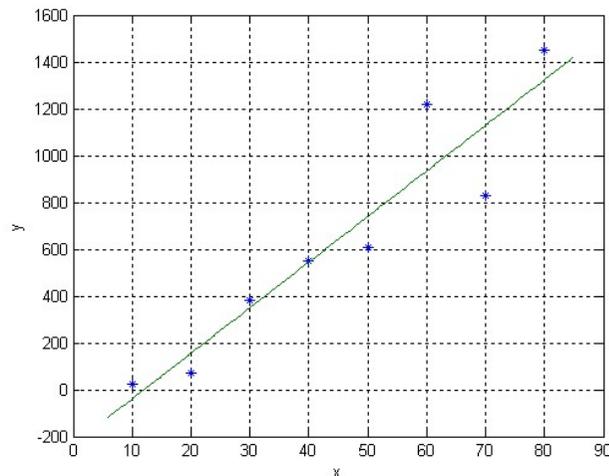
Exemplo 2

- Um objeto é suspenso em um túnel de vento e a força é medida em diversos níveis de velocidade:

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

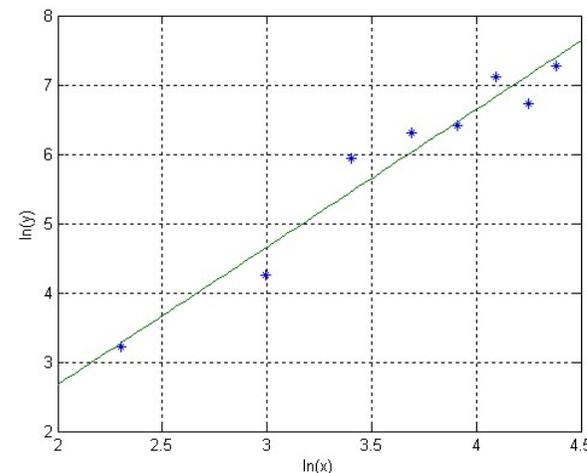
- Através de regressão por mínimos quadrados, verifica-se qual é o melhor ajuste: com uma reta ou com uma equação $y^* = ax^b$

$$y^* = -234,2857 + 19,4702x$$



$$r = 0,9384$$

$$\ln y^* = -1,2941 + 1,9842 \cdot \ln x$$



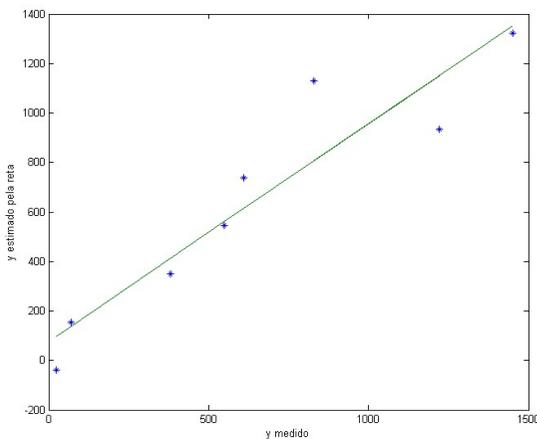
$$r = 0,9737$$

melhor ajuste

Exemplo 2 (continuação)

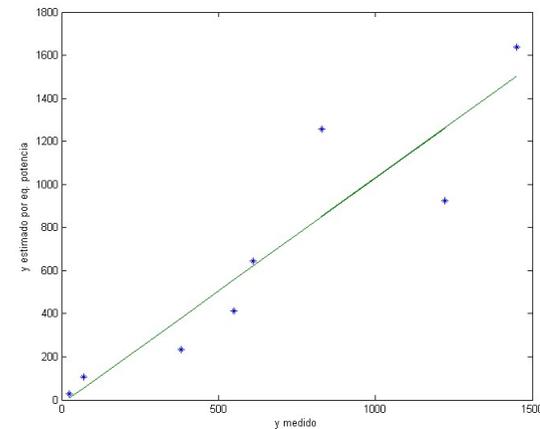
- Outro modo de verificar a qualidade da escolha é comparar os resultados do ajuste nos m pontos originais
- Em outras palavras, faz-se regressão linear nos pontos $(y_1, f^*(x_1)), \dots, (y_m, f^*(x_m))$
- No caso de um ajuste perfeito, será encontrada uma reta com coeficientes $a_0 = 0$ e $a_1 = 1$, que coincide nos pontos tabelados
- Resultados obtidos no caso anterior:

$$y^* = 76,7135 + 0,8805y$$



$$r = 0,9384$$

$$y^* = -18,6452 + 1,0497y$$



$$r = 0,9737$$

mesmo valor

MatLab



- `polyfit(x,y,k)`
 - Vetor com os $k+1$ coeficientes do polinômio de grau k que ajusta a curva descrita pelos dados dos vetores x e y , utilizando o método dos mínimos quadrados