INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE SANTA CATARINA
CAMPUS FLORIANÓPOLIS
DEPARTAMENTO ACADÊMICO DE ELETRÔNICA

COMPUTAÇÃO CIENTÍFICA

Prof. Marco Valério Miorim Villaça

Capítulo IV

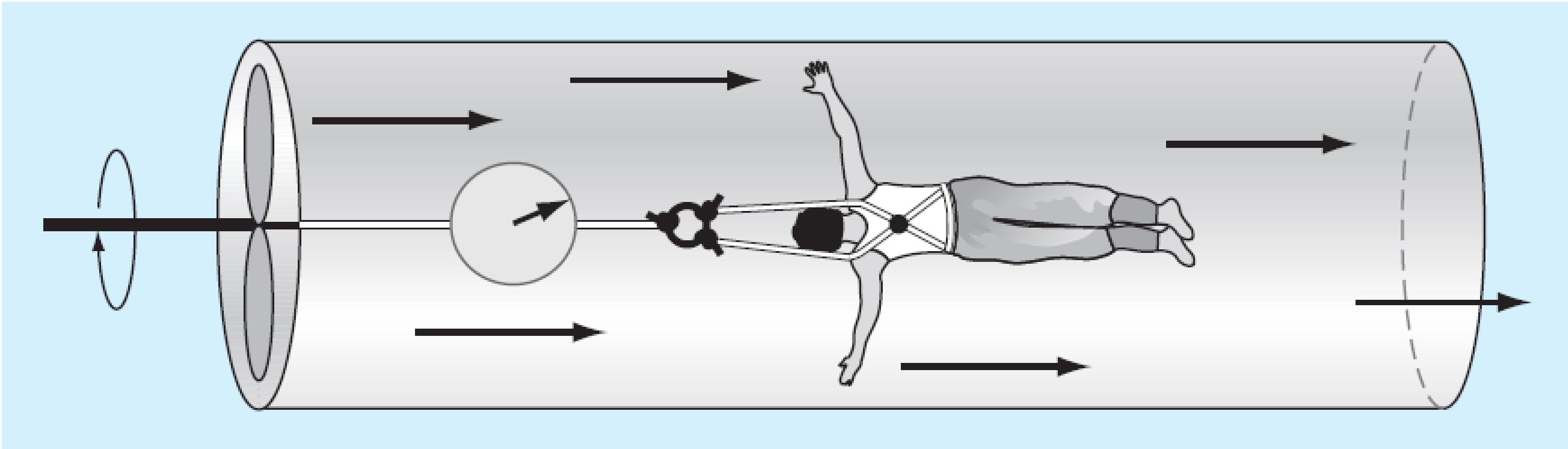
Ajuste de curvas: método dos mínimos quadrados, método polinomial e linearização

Parte I – Regressão Linear

Regressão Linear

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

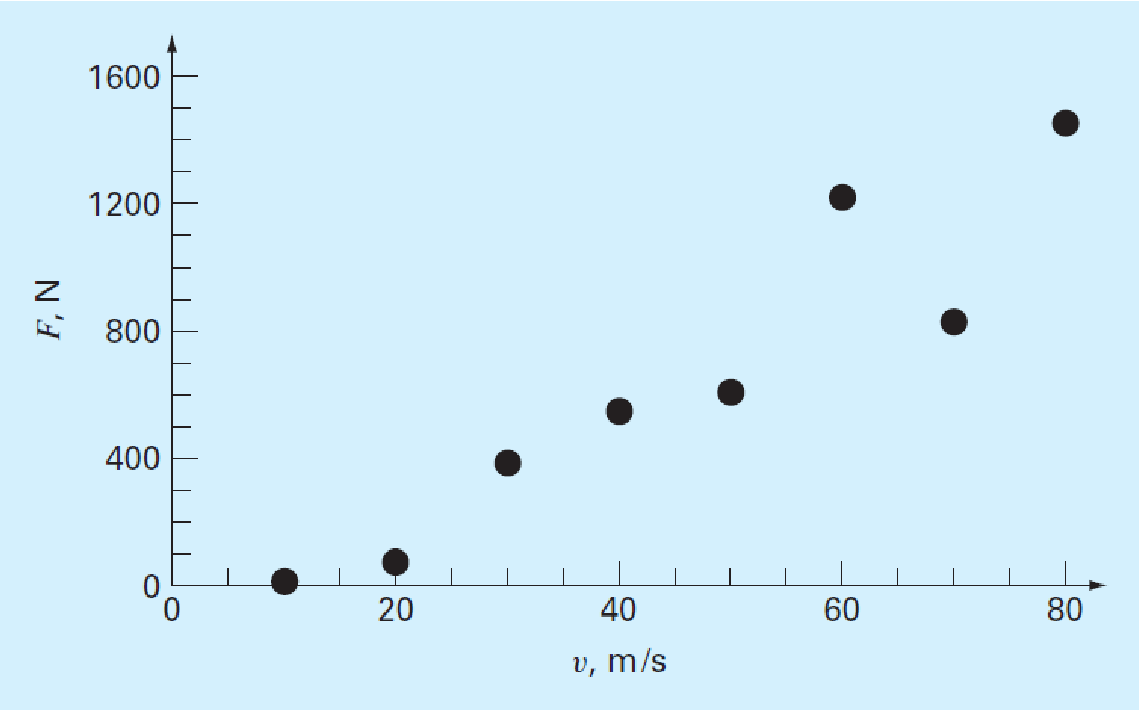
  Exemplo: Experimento do túnel de vento para medir como a força da resistência do ar depende da velocidade.



Regressão Linear

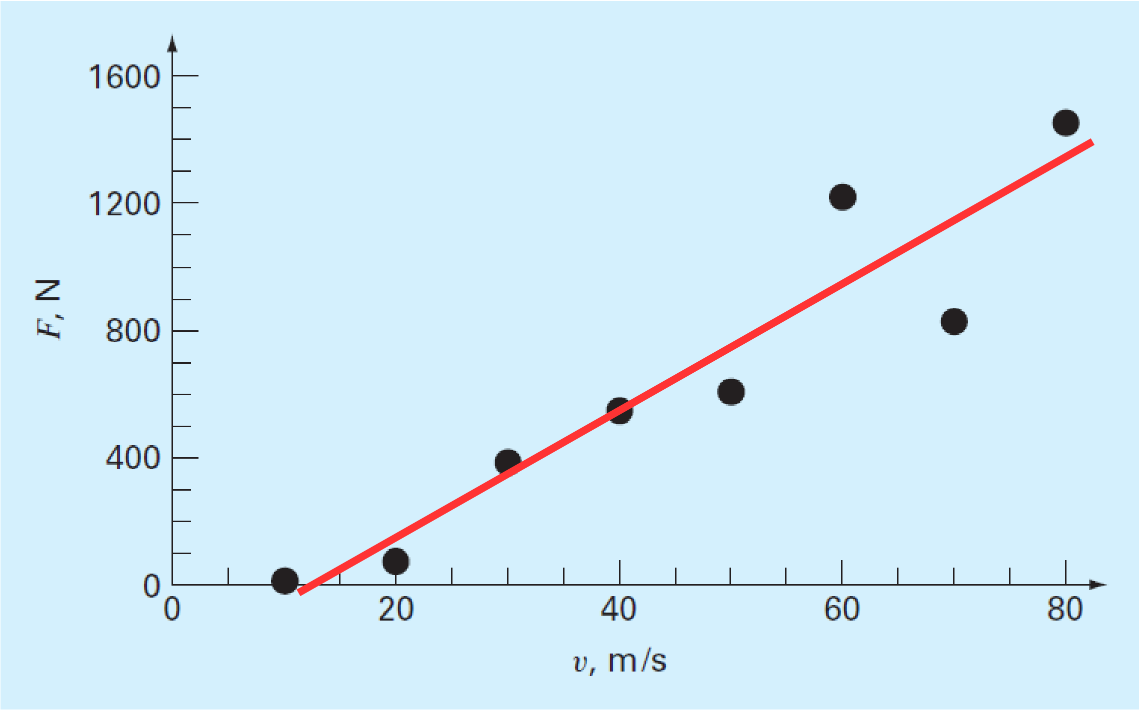
Dados experimentais (Tabela I) para a força e a velocidade em um experimento de túnel de vento e gráfico correspondente

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



Regressão Linear

Como obter a reta que melhor se ajusta aos dados?



Método dos mínimos quadrados

- Quando um erro substancial estiver associado aos dados, a melhor estratégia de ajuste de curva é determinar uma função aproximada que ajuste a forma ou 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

$$\left(x_{1,}\, y_1\right), \left(x_{2,}\, y_2\right), \cdots, \left(x_n\,, y_n\right)$$

- A expressão matemática do ajuste por uma reta é

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

onde: a_0 – intersecção com o eixo y;

a_1 – inclinação da reta e

e = erro ou resíduo entre o modelo e a observação

Método dos mínimos quadrados

- O erro ou resíduo é, portanto, descrito por

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

  ou seja, a discrepância entre o valor verdadeiro de y e o valor aproximado $a_0 + a_1 x$

- O método dos mínimos quadrados (MMQ) é o mais utilizado em muitas ciências experimentais para o ajuste de parâmetros a dados experimentais.
- 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:

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

Método dos mínimos quadrados
Ajuste de uma reta

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

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

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

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

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

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

Método dos mínimos quadrados
Ajuste de uma reta

- Como $\sum_{i=1}^{n} (a_0) = n\, a_0$
- Resulta:

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

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

- Resolvendo o sistema para $a_{1:}$

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

Método dos mínimos quadrados
Ajuste de uma reta

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

$$a_0 = \bar{y} - a_1 \bar{x}$$

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

- Exemplo 1

Ajuste uma reta aos valores da Tabela I

Método dos mínimos quadrados
Ajuste de uma reta

- Nesse caso F = f (v). Reorganizando os dados da Tabela I e totalizando os somatórios necessário, constrói-se a tabela abaixo

i	x_i	y_i	x_i^2	$x_i y_i$
1	10	25	100	250
2	20	70	400	1 400
3	30	380	900	11 400
4	40	550	1 600	22 000
5	50	610	2 500	30 500
6	60	1 220	3 600	73 200
7	70	830	4 900	58 100
8	80	1 450	6 400	116 000
$\sum$	360	5 135	20 400	312 850

- Dessa tabela resulta que $\bar{y} = \dfrac{5135}{8} = 641{,}875 \;\; e \;\; \bar{x} = \dfrac{360}{8} = 45$

Método dos mínimos quadrados
Ajuste de uma reta

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

$$a_1 = \frac{8 \cdot 312 - 360 \cdot 5135}{8 \cdot 20400 - 360^2} = 19{,}47024$$

$$a_0 = 641{,}875 - 19{,}47024 \cdot 45 = -234{,}2875$$

- Utilizando F no lugar de y e v no lugar de x, o ajuste por mínimos quadrados dos dados é

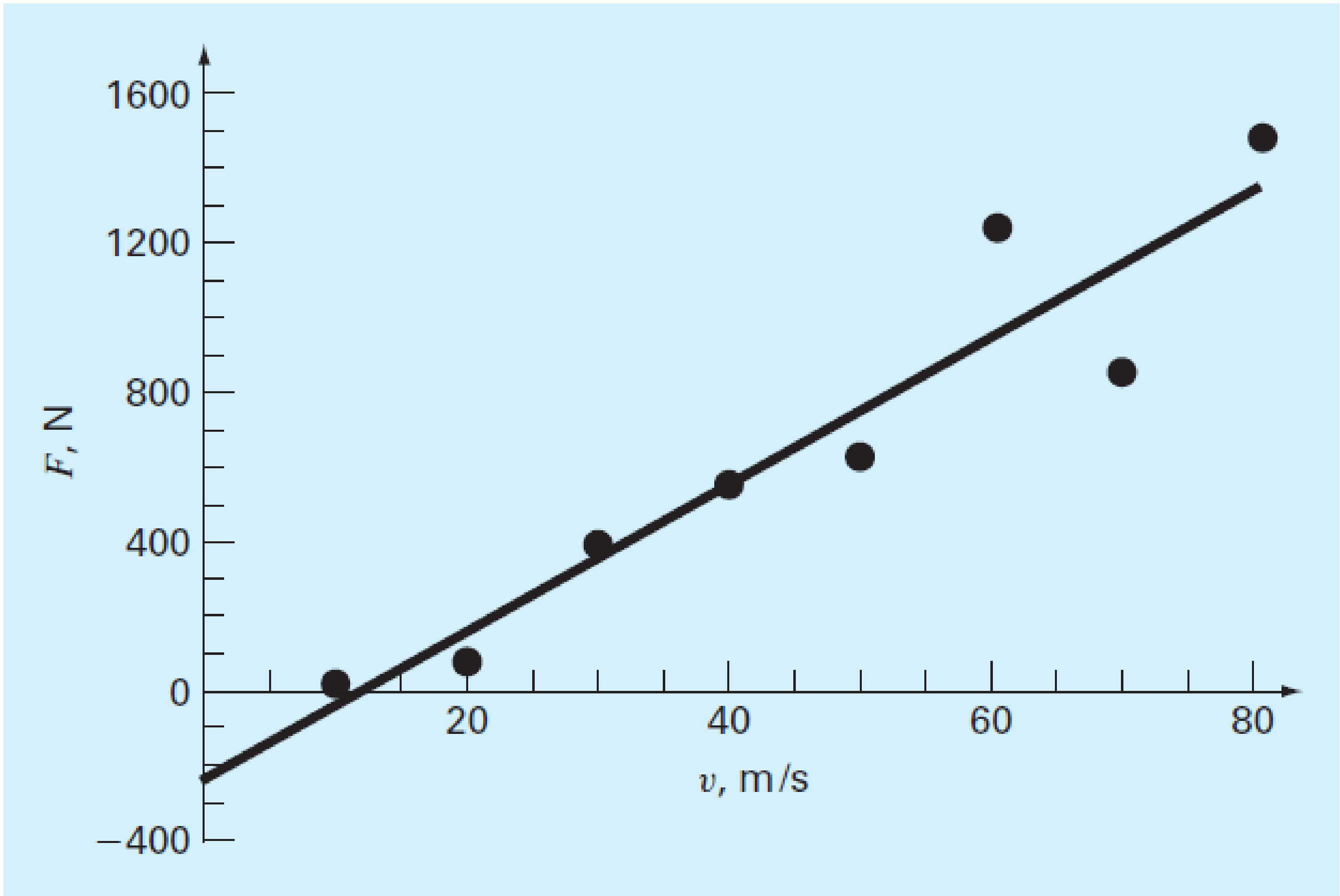
$$F = -234{,}2857 + 19{,}47024\,v$$

- Reta válida para

$$v > \frac{234{,}2857}{19{,}47024} > 12{,}03 \; m/s$$

Método dos mínimos quadrados
Ajuste de uma reta

- A Figura abaixo representa o ajuste por mínimos quadrados de uma reta aos dados da Tabela I



Quantificação do erro
Regressão linear

- Vimos que a soma dos quadrados dos resíduos é definida por:

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

- Equação 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 no cálculo do desvio padrão

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

- Assim, 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 residual que permanece após a regressão

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

normalizada, é chamado 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.

Quantificação do erro
Regressão linear

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

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

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

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

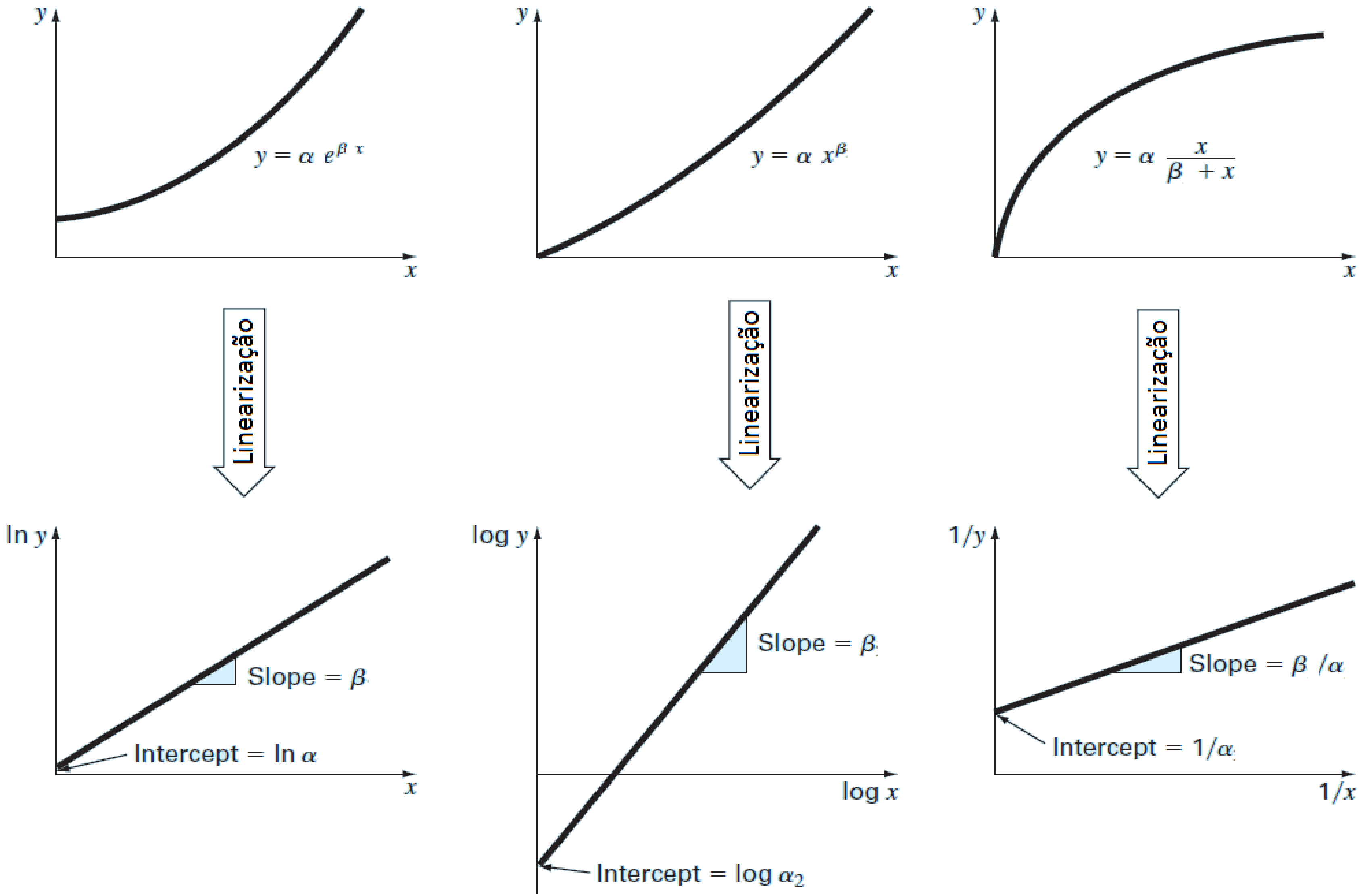
- O que indica que 88,05% da incerteza original foi explicada pelo ajuste linear.

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 de taxa de crescimento de saturação: $y = \alpha \dfrac{x^m}{\beta^m + x^m}$

onde m é a ordem do ajuste.

Linearização de relações não lineares



Linearização de relações não lineares

- Equação exponencial
  - $y = \alpha e^{\beta x} \quad \rightarrow \quad \ln y = \ln \alpha + \beta x$
- Equação de potência simples:
  - $y = \alpha x^{\beta} \quad \rightarrow \quad \log y = \log \alpha \; + \; \beta \log x$
- Equação da taxa de crescimento da saturação:
  - $y = \alpha \dfrac{x^m}{\beta^m + x^m} \longrightarrow \quad \dfrac{1}{y} = \dfrac{1}{\alpha} \; + \; \dfrac{\beta^m}{\alpha} \; \cdot \dfrac{1}{x^m}$

Ajustando os dados com a eq. de potência simples
Exemplo

- Exemplo 2

  Ajuste a equação $y = \alpha x^{\beta}$ aos dados da Tabela I usando uma transformação logarítmica

Ajustando os dados com a eq. de potência simples Exemplo

- Reorganizando os dados da Tabela I e totalizando os somatórios necessário. constrói-se a tabela abaixo

i	x_i	y_i	$\log x_i$	$\log y_i$	$(\log x_i)^2$	$\log x_i \log y_i$
1	10	25	1.000	1.398	1.000	1.398
2	20	70	1.301	1.845	1.693	2.401
3	30	380	1.477	2.580	2.182	3.811
4	40	550	1.602	2.740	2.567	4.390
5	50	610	1.699	2.785	2.886	4.732
6	60	1220	1.778	3.086	3.162	5.488
7	70	830	1.845	2.919	3.404	5.386
8	80	1450	1.903	3.161	3.622	6.016
$\sum$			12.606	20.515	20.516	33.622

- Dessa tabela resulta que

$$\bar{y} = \frac{20,515}{8} = 2,5644 \quad e \quad \bar{x} = \frac{12,606}{8} = 1,5757$$

Ajustando os dados com a eq. de potência simples
Exemplo

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

$$a_1 = \frac{8 \cdot 33{,}622 - 12{,}606 \cdot 20{,}515}{8 \cdot 20{,}516 - 12{,}606^2} = 1{,}9842$$

$$a_0 = 2{,}5644 - 1{,}9842 \cdot 1{,}5757 = -0{,}5620$$

- Utilizando F no lugar de y e v no lugar de x, o ajuste por mínimos quadrados dos dados é

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

Ajustando os dados com a eq. de potência simples
Exemplo

- Para exibir o ajuste utilizando coordenadas não transformadas, utiliza-se

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

$$a_1 = \beta$$

- Substituindo os valores:

$$\alpha = 10^{-0,5620} = 0,2741$$

$$\beta = 1,9842$$

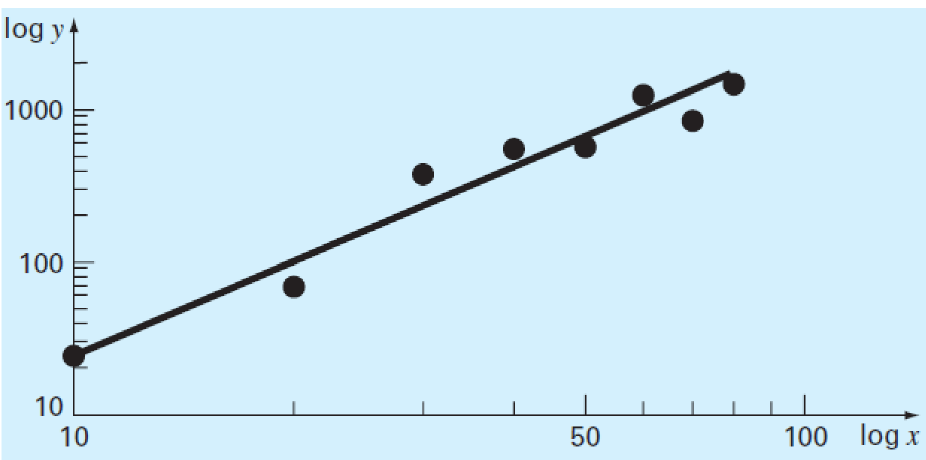
- O ajuste por mínimos quadrados, então, é

$$F = 0,2741\, v^{1,9842}$$

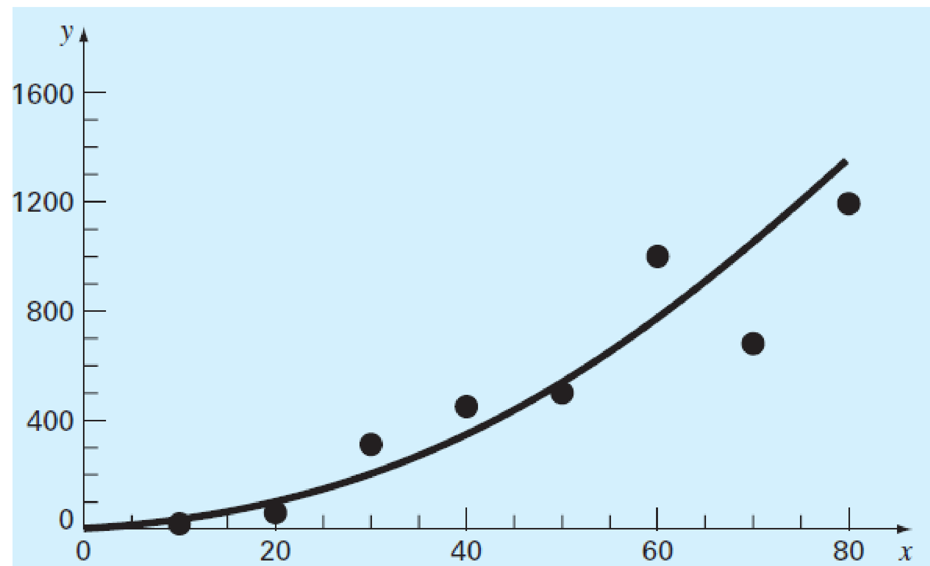
- O coeficiente de determinação é (calcule): $r^2 = 0,9481$

Ajustando os dados com a eq. de potência simples Exemplo

A Figura abaixo representa o ajuste por mínimos quadrados de um modelo de potência simples aos dados da Tabela I, sendo (a) o ajuste dos dados transformados e (b) o ajuste da equação de potência juntamente com os dados



(a)



(b)

Exercícios

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 m*l)	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

Função Scilab

- Faça uma função Scilab

```
function [a,r2] = reglinear(x,y)
```

para realizar a regressão linear ajustando os dados a um dos modelos estudados. A função após ser chamada deve apresentar o seguinte menu:

Digite:

<1> para Ajuste por uma reta

<2> para Ajuste por uma exponencial

<3> para Ajuste por uma potência simples

<4> por uma eq. de taxa de crescimento saturado de ordem m

Função Scilab

- *A função deve:*
  - *Retornar a_0 ,a_1 e o coeficiente de determinação;*
  - *Traçar a curva ajustada e os pontos experimentais;*
  - *Calcular α e β quando for o caso (2, 3 e 4).*
- *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.