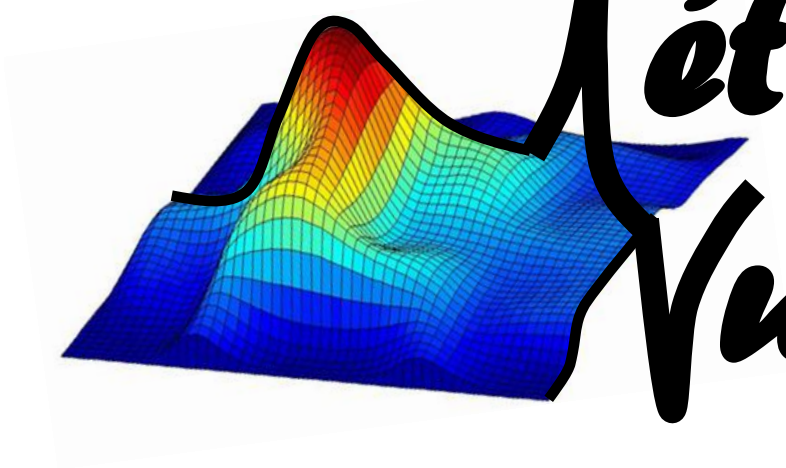


Métodos Numéricos

AJUSTE DE FUNÇÕES

Prof. Erivelton Geraldo Nepomuceno

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

UNIVERSIDADE DE JOÃO DEL-REI
PRÓ-REITORIA DE PESQUISA

CENTRO FEDERAL DE EDUCAÇÃO
TECNOLÓGICA DE MINAS GERAIS
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO





2016

Métodos Numéricos

Conteúdo

1. Diferença Entre Regressão e Interpolação
2. Regressão Linear Simples.
3. Qualidade do Ajuste
4. Regressão Linear Múltipla

Introdução

- É importante relacionar, por meio de um modelo matemático, a variável resposta (ou dependente) com o conjunto de variáveis explicativas (ou independentes).
- Para ter controle, determinar algum parâmetro ou mesmo fazer previsão acerca do comportamento da variável resposta.

Diferença entre regressão e interpolação

- Polinômio interpolador de grau n-1 é construído de modo a passar por n pontos dados:

$$P_{n-1}(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}.$$

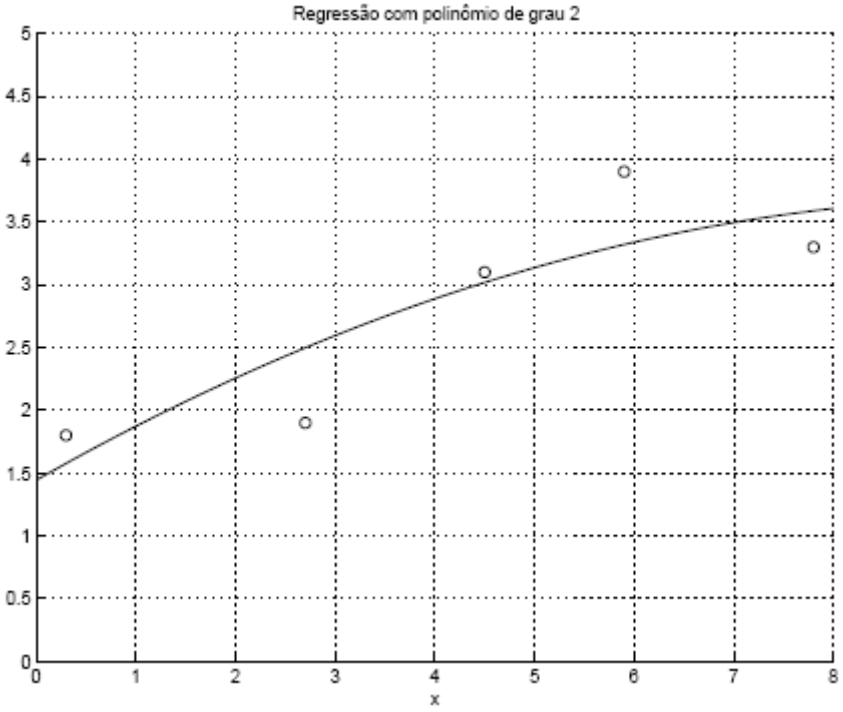
- Possui n coeficientes a_i, i = 0, 1, ... , n - 1.
- O número de pontos utilizados para gerar o polinômio interpolador é igual ao número de coeficientes do polinômio.
- O Polinômio de regressão de grau g, usando n pontos:

$$U_g(x) = b_0 + b_1 x + b_2 x^2 + \ldots + b_g x^g,$$

- sendo g < n - 1.
- Quando g = n - 1 o polinômio de regressão é idêntico ao polinômio interpolador.

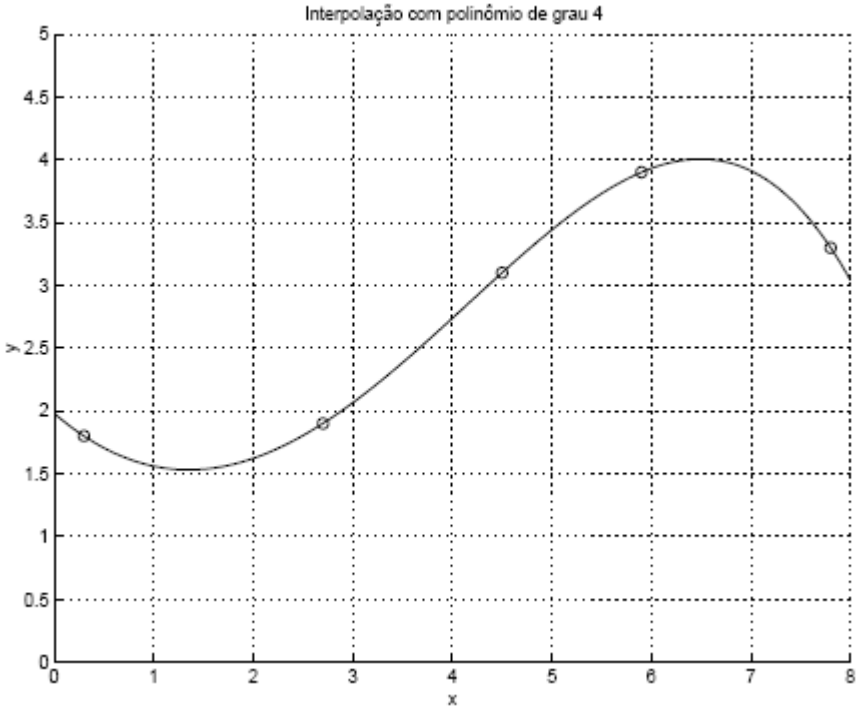
Diferença entre regressão e interpolação

- Polinômio de regressão de grau g = 2 com n = 5 pontos.



- Quando o polinômio de regressão possuir grau g = n - 1 = 4 ele se torna idêntico a um polinômio interpolador de mesmo grau.
- O PI passa por todos os pontos do diagrama de dispersão:

Diferença entre regressão e interpolação



Diferença entre regressão e interpolação

- Em termos de complexidade computacional, a interpolação é um processo mais simples que a regressão polinomial (solução do sistema linear).
- A interpolação deve ser utilizada quando se necessita de um valor intermediário não constante de uma tabela.
- A regressão tem que ser utilizada quando se deseja estimar um parâmetro de um modelo semideterminístico e/ou prever um valor dado por esse modelo.

Relação entre Variáveis

As relações entre as variáveis envolvidas em um experimento podem ser classificadas em três tipos: determinísticas, semideterinísticas e empíricas.

Relações determinísticas:

- Variáveis relacionadas entre si por uma lei expressa por formula matemática precisa.
- Variação nas observações é atribuída a erros experimentais.
- Por exemplo, se r reais forem investidos durante m meses a uma taxa de juros j, ao final do prazo ter-se-a v reais.
- As variáveis r, m, j e v estão relacionadas pela expressão exata fornecida pela Matemática Financeira $v = r(1+j)^m$, que e a lei dos juros compostos.
- Qualquer analise adicional é desnecessária para relacionar estas variáveis.

Relação entre Variáveis

Relações semideterminísticas:

- Alguma teoria prescreve uma forma para a relação entre as variáveis, mas não os valores particulares dos parâmetros que aparecem na relação.
- É necessário realizar experimentos para obter informações acerca desses parâmetros.
- Precisão limitada dos instrumentos de medida, as perturbações incontroláveis dos experimentos e outros fatores introduzem erros nos dados, causam perturbação na verdadeira relação.
- Por exemplo, a concentração c de uma substância após um tempo t em uma reação química de primeira ordem é $c = c_0 e^{-kt}$, c_0: concentração inicial e k: constante de velocidade de uma reação específica. A constante k é obtida experimentalmente.

Relação entre Variáveis

Relações empíricas:

- Relação entre as variáveis envolvidas não é conhecida.
- Determinar uma fórmula matemática que relacione essas variáveis.
- O Gráfico feito com valores observados dessas variáveis fornece uma idéia da relação entre elas com algumas variações aleatórias.
- Pode ser que a relação obtida não siga uma fórmula matemática precisa, dada a complexidade do problema.
- Deve-se ter suficiente conhecimento sobre uma relação empírica, para desenvolver a teoria que conduza a uma fórmula matemática, caso semideterminstico.

Regressão Linear Simples

- As relações mais simples entre duas variáveis são as relações lineares.
- A variável independente ou explicativa x é relacionada com a variável dependente ou resposta y por meio de um modelo linear:

$$y = b_0 + b_1 x.$$

Diagrama de dispersão:

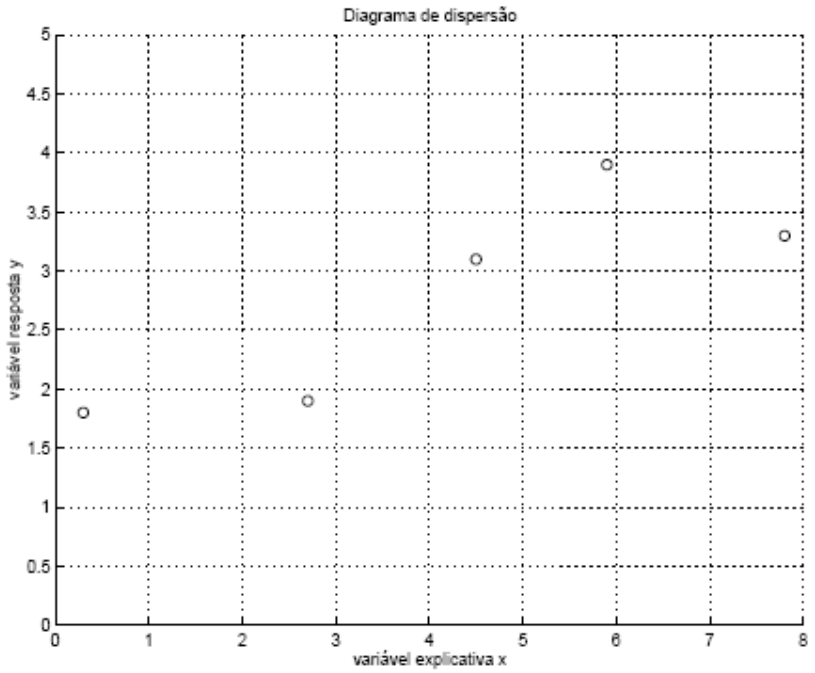
- Uma etapa importante ao analisar a relação entre duas variáveis é esboçar os dados em um gráfico de coordenadas cartesianas denominado diagrama de dispersão.
- Diagrama mostra a natureza da relação intrínseca entre as duas variáveis estudadas.

Regressão Linear Simples

- Sejam os dados da tabela relacionando as variáveis x e y:

x	0,3	2,7	4,5	5,9	7,8
y	1,8	1,9	3,1	3,9	3,3

- Diagrama de dispersão de dados:



Métodos Numéricos

Regressão Linear Simples

Retas de Regressão:

- Modelo simples que relaciona as variáveis x e y é:

$$y = \beta_0 + \beta_1 x + \epsilon,$$

- β_0 e β_1 são os parâmetros a serem estimados.
- ε contém os componentes desconhecidos e aleatórios de erro que se sobrepõem à verdadeira relação linear.
- Como estimar os parâmetros β_0 e β_1 ?

Métodos Numéricos

Regressão Linear Simples

Modelo 1:

- Usar o polinômio interpolador linear.
- Através do diagrama de dispersão apresentado é possível perceber que não se pode traçar uma única reta que passe por todos os pontos simultaneamente.
- Assim a reta é esboçada a partir de dois pontos quaisquer, por exemplo, o primeiro e o último:

x	0,3	7,8
y	1,8	3,3

- Equação da reta u(x) que passa por estes dois pontos:

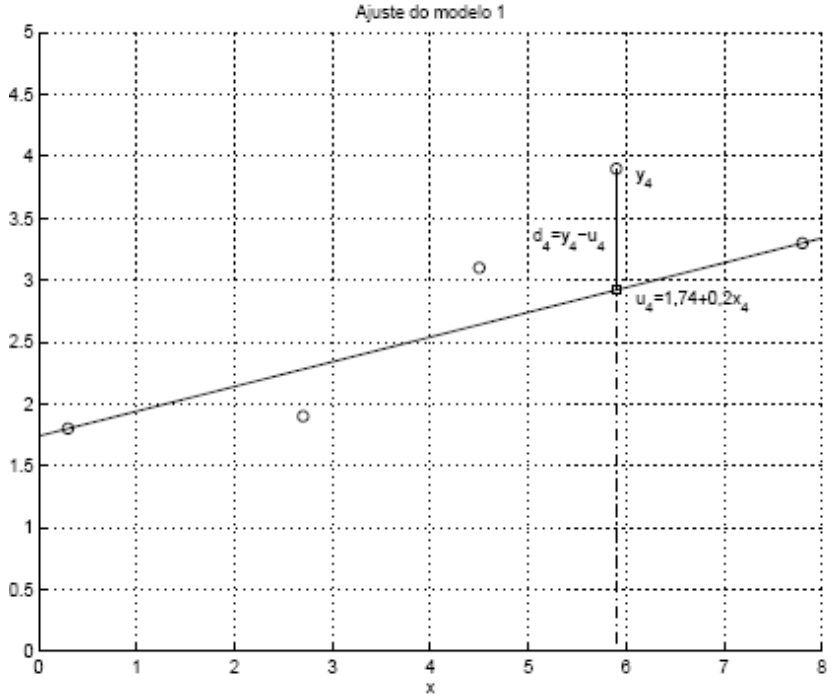
$$u(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = 1{,}8 + \frac{3{,}3 - 1{,}8}{7{,}8 - 0{,}3}(x - 0{,}3),$$

$$u(x) = 1{,}8 + 0{,}2(x - 0{,}3) \rightsquigarrow u(x) = 1{,}74 + 0{,}2x.$$

Regressão Linear Simples

- A Figura a seguir mostra a reta u=1.74+0.2x traçada entre os pontos do diagrama de dispersão.
- A distância vertical d_i entre o i-ésimo ponto dado y_i e o ponto u_i = 1.74+0.2x_i de mesma abscissa x_i é:

$$d_i = y_i - u_i.$$



Regressão Linear Simples

- Uma forma de calcular a qualidade do ajuste é calculando a soma de todas as n distâncias verticais de y_i aos pontos da reta $u_i = 1.74+0.2x_i$ considerando os valores positivos de d_i:

$$D(b_0, b_1) = \sum_{i=1}^{n} (y_i - u_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i)^2 = \sum_{i=1}^{n} d_i^2,$$

$$D(1{,}74; 0{,}2) = \sum_{i=1}^{5} (y_i - (1{,}74 + 0{,}2x_i))^2.$$

i	x_i	y_i	u_i	d_i
1	0,3	1,8	1,80	0,00
2	2,7	1,9	2,28	−0,38
3	4,5	3,1	2,64	0,46
4	5,9	3,9	2,92	0,98
5	7,8	3,3	3,30	0,00
$D(1{,}74; 0{,}2) = 1{,}3164$				

Resultados do ajuste pelo modelo 1.

Regressão Linear Simples

Modelo 2:

- Usa o polinômio interpolador linear.
- Reta traçada por dois pontos quaisquer.
- Pontos escolhidos não pertencentes ao diagrama de dispersão.
- Por exemplo, escolhendo os pontos

x	2	6
y	2	3

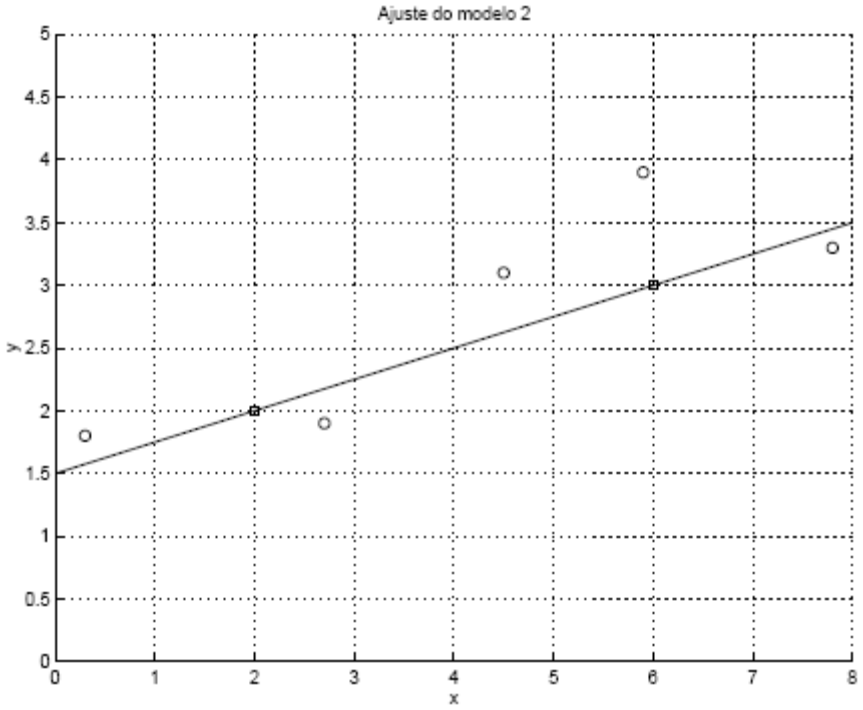
- A reta u(x) será:

$$u(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = 2 + \frac{3-2}{6-2}(x-2)$$

$$2 + 0{,}25(x-2) \rightsquigarrow u(x) = 1{,}5 + 0{,}25x.$$

Regressão Linear Simples

- A Figura a seguir mostra a reta u=1.5+0.25x traçada entre os pontos do diagrama de dispersão.
- A tabela mostra os resultados do ajuste pelo modelo 2.



i	x_i	y_i	u_i	d_i
1	0,3	1,8	1,575	0,225
2	2,7	1,9	2,175	−0,275
3	4,5	3,1	2,625	0,475
4	5,9	3,9	2,975	0,925
5	7,8	3,3	3,450	−0,150
			$D(1{,}5; 0{,}25) = 1{,}2300$	

O modelo 2 é mais adequado:

$$D(1{,}5; 0{,}25) = 1{,}2300 < D(1{,}74; 0{,}2) = 1{,}3164.$$

Regressão Linear Simples

Método dos mínimos quadrados:

- A qualidade do ajuste depende da equação da reta escolhida.
- Reta que não passa por dois pontos dentre aqueles do diagrama de dispersão produziu resultado melhor.
- Por onde se deve traçar a reta de modo a obter o menor valor do desvio D?
- O método dos mínimos quadrados consiste em encontrar uma estimativa da reta u = $\beta_0 + \beta_1 x$ para produzir o menor valor possível do desvio:

$$D(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - u_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

Regressão Linear Simples

- Cujas derivadas parciais são:

$$D(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - u_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

$$\frac{\partial D(\beta_0, \beta_1)}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i),$$

$$\frac{\partial D(\beta_0, \beta_1)}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i.$$

- Os valores para os quais a função $D(\beta_0, \beta_1)$ possui um mínimo são aqueles onde as derivadas parciais se anulam.
- Se $D(b_0, b_1)$ for o ponto de mínimo de $D(\beta_0, \beta_1)$, então:

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0 \rightarrow \sum_{i=1}^{n} b_0 + \sum_{i=1}^{n} b_1 x_i = \sum_{i=1}^{n} y_i,$$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) x_i = 0 \rightarrow \sum_{i=1}^{n} b_0 x_i + \sum_{i=1}^{n} b_1 x_i^2 = \sum_{i=1}^{n} x_i y_i.$$

Métodos Numéricos

Regressão Linear Simples

- Na forma matricial:

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

- Os valores em que D(β_0, β_1) apresentam um mínimo são obtidos pela solução do sistema linear denominado equações normais.
- Utilizando as operações l-elementares, obtém-se:

$$\begin{bmatrix} n & \sum x_i \\ 0 & -\frac{1}{n}(\sum x_i)^2 + \sum x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ -\frac{1}{n}\sum x_i \sum y_i + \sum x_i y_i \end{bmatrix}$$

- Cuja solução é:

$$u(x) = b_0 + b_1 x$$

$$b_1 = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2}$$

$$b_0 = \frac{\sum y_i - b_1 \sum x_i}{n}$$

Métodos Numéricos

Regressão Linear Simples

Exemplo: Calcular a reta de mínimos quadrados usando:

x	0,3	2,7	4,5	5,9	7,8
y	1,8	1,9	3,1	3,9	3,3

- Valores dos somatórios necessários para resolver o sistema:

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	0,3	1,8	0,09	0,54	3,24
2	2,7	1,9	7,29	5,13	3,61
3	4,5	3,1	20,25	13,95	9,61
4	5,9	3,9	34,81	23,01	15,21
5	7,8	3,3	60,84	25,74	10,89
$\sum$	21,2	14,0	123,28	68,37	42,56

Métodos Numéricos

Regressão Linear Simples

- Solução de quadrados mínimos:

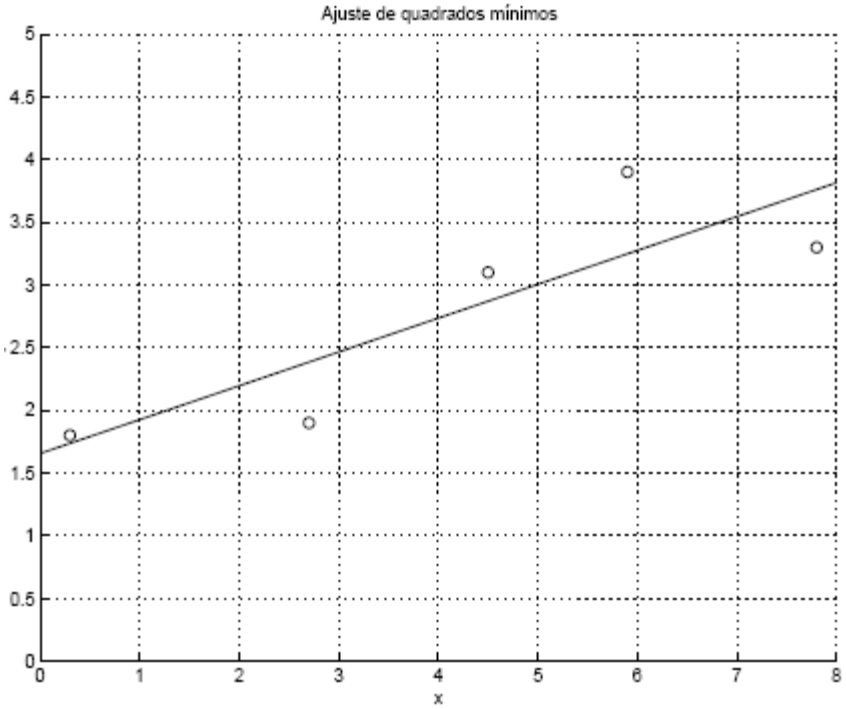
$$b_1 = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2} = \frac{21{,}2 \cdot 14{,}0 - 5 \cdot 68{,}37}{(21{,}2)^2 - 5 \cdot 123{,}28} \rightsquigarrow b_1 = 0{,}2698;$$

$$b_0 = \frac{\sum y_i - b_1 \sum x_i}{n} = \frac{14{,}0 - 0{,}2698 \cdot 21{,}2}{5} \rightsquigarrow b_0 = 1{,}6560$$

$$u = 1{,}6560 + 0{,}2698x$$

Regressão Linear Simples

- Reta u e ajuste de quadrados mínimos:



i	x_i	y_i	u_i	d_i
1	0,3	1,8	1,7369	0,0631
2	2,7	1,9	2,3845	−0,4845
3	4,5	3,1	2,8701	0,2299
4	5,9	3,9	3,2478	0,6522
5	7,8	3,3	3,7604	−0,4604
$D(1{,}6560; 0{,}2698) = 0{,}9289$				

- Melhor dos três modelos propostos:

$$D(1{,}6560; 0{,}2698) = 0{,}9289 < D(1{,}5; 0{,}25) = 1{,}2300 <$$

$$D(1{,}74; 0{,}2) = 1{,}3164.$$

Métodos Numéricos

Referencias Bibliográficas

1. Frederico Ferreira Campos Filho, **Algoritmos Numéricos**.