

UNIVERSIDADE FEDERAL DE SÃO CARLOS
Campus Sorocaba

RICARDO BALARIN MENEGUEL

MÉTODO DOS MÍNIMOS QUADRADOS E APLICAÇÕES

Sorocaba

2014

RICARDO BALARIN MENEGUEL

MÉTODO DOS MÍNIMOS QUADRADOS E APLICAÇÕES

Orientador: Prof. Dr. Antonio Luís Venezuela

Monografia apresentada ao Departamento de Física, Matemática e Química da Universidade Federal de São Carlos, Campus de Sorocaba – DFQM, como condição parcial para a obtenção do título de Licenciatura em Matemática.

Sorocaba

2014

Meneguel, Ricardo Balarin

Métodos dos mínimos quadrados e aplicações / Ricardo Balarin Meneguel.

2014

Dissertação (Trabalho de Conclusão de Curso – Orientador: Prof. Antonio Luís Venezuela) – Fundação Universidade Federal de São Carlos – Campus Sorocaba, 2014.

1. Regressão Linear. 2. Ajuste de Curvas .I Meneguel, Ricardo Balarin. II. Fundação Universidade Federal de São Carlos – Campus Sorocaba. III Métodos dos mínimos quadrados e aplicações

MÉTODO DOS MÍNIMOS QUADRADOS E APLICAÇÕES

RICARDO BALARIN MENEGUEL

Monografia apresentada ao Departamento de Física, Matemática e Química da Universidade Federal de São Carlos, Campus de Sorocaba – DFQM, como condição parcial para a obtenção do título de Licenciatura em Matemática, obtendo o conceito de _____ atribuído pelos professores examinadores.

Prof. Dr. Antonio Luís Venezuela
Orientador DFQM/UFSCar

Prof.^a. Dr.^a. Magda da Silva Peixoto
Examinadora DFQM/UFSCar

Prof. Dr. Renato Fernandes Cantão
Examinador DFQM/UFSCar

Folha de aprovação

Ricardo Balarin Meneguel

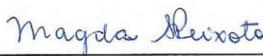
Método dos Mínimos Quadrados e Aplicações

Trabalho de Conclusão de Curso

Universidade Federal de São Carlos – *Campus Sorocaba*

Sorocaba, 12/12/2014.

Orientador 
Prof. Dr. Antonio Luís Venezuela – UFSCar *Campus Sorocaba*

Membro 2 
Prof.^a Dr.^a Magda da Silva Peixoto – UFSCar *Campus Sorocaba*

Membro 3 
Prof. Dr. Renato Fernandes Cantão – UFSCar *Campus Sorocaba*

Agradecimentos

Agradeço primeiramente a Deus por ter me ajudado nas diversas dificuldades. A meus pais pelo apoio e motivação.

Aos diversos professores que me acompanharam desde o Ensino Básico ao Ensino Superior.

Ao Professor Venezuela por sua imensa boa vontade em me orientar e em sua imensa paciência em trabalhar semanalmente no que eu buscava neste trabalho.

A todos os colegas do curso de Matemática Turma 2009 que além do conhecimento compartilhado em diversas áreas, tivemos momentos excelentes.

A todos os colegas da Matemática Turma 2010 pela sua recepção desde o início do curso em relação a mim e pelo companheirismo de todos.

A todos os professores que cursei disciplinas nesta universidade que responderam dúvidas das quais eu não conseguia responder sobre o pensamento humano, científico e matemático, e também sempre instigando dúvidas das quais nem imaginava que eram possíveis.

“A lei de Murphy não quer dizer que algo ruim irá acontecer. Apenas nos diz se houverem certas condições a situação ruim irá ocorrer.”

Interstellar-Warner Bross Pictures

RESUMO

Existem diversos métodos de ajuste linear ou de curvas na matemática. Neste trabalho, estuda-se a técnica denominada método dos mínimos quadrados (MMQ), no caso discreto e no caso contínuo. Essa metodologia parte do princípio que, dados pontos experimentais e plotando-os no plano cartesiano, pretende-se ajustar uma função a estes pontos. A partir de dados provindos da literatura, elaboram-se modelos gerados via método dos mínimos quadrados. Para se determinar a confiabilidade destes modelos, utiliza-se o coeficiente de correlação de Pearson.

Palavras chaves: regressão linear, ajuste de curvas, coeficiente de correlação de Pearson.

LISTA DE FIGURAS

Figura 1. Representação de ponto (X_i, Y_i) e da função de ajuste com o respectivo resíduo.	2
Figura 2. Representação das estimativas relativas às equações (28) á direita e (24) á esquerda.	8
Figura 3. Representação variação total da Equação (25) representada no plano cartesiano.	9
Figura 4. Representação do gráfico de dispersão e da linha de tendência linear, cuja equação é dada por: $f(Q) = T = -0,86Q + 196,92$	14
Figura 5. Gráfico do modelo exponencial que relaciona as notas das teclas <i>versus</i> a frequência (Hz).	16
Figura 6. Gráfico que aproxima a função $f(x) = e^x$ por um polinômio p , de grau 1.	17

LISTA DE TABELAS

Tabela 1. Dados relativos ao índice da quantidade demandada e índice da tarifa real média, sendo $n=10$	12
Tabela 2. Cálculos para ajuste linear, considerando a variável independente, Q , e a variável independente, T , tomando $f_0(Q)=1$ e $f_1(Q)=Q$	13
Tabela 3. Cálculos para o coeficiente de correlação de Pearson.....	13
Tabela 4. Dados relativos à frequência em Hertz de cada tecla da escala temperada. Fonte Casio (Casio, acesso 10/2014).....	15
Tabela 5. Dados provindos da Tabela 5 e linearizados.....	16

SUMÁRIO

1	INTRODUÇÃO.....	1
1	TEORIA MATEMÁTICA	2
1.1.	Método dos Mínimos Quadrados – Caso Discreto.....	2
1.2.	Método dos Mínimos Quadrados – Caso Contínuo	5
1.3.	Coeficiente de Correlação de Pearson	7
1.3.1.	Desvio Padrão de Dados Absolutos.....	7
1.3.2.	Erro Padrão de Estimativa	7
1.3.3.	Coeficiente de Determinação.....	10
2	APLICAÇÕES DO MÉTODO DOS MÍNIMOS QUADRADOS	12
2.	12
2.1.	APLICAÇÃO 1: ENERGIA ELÉTRICA – DEMANDA VERSUS TARIFA MÉDI..	12
2.2.	Aplicação 2: Notas do Teclado de um Piano <i>versus</i> Frequência.....	15
2.3.	Aproximação de uma função qualquer para um polinômio- caso contínuo	17
3	CONCLUSÃO	18
	REFERÊNCIAS	18

1 INTRODUÇÃO

Nos primeiros feitos e resultados importantes da Astronomia obtidas na Grécia Antiga, observava-se que em diferentes momentos e diferentes lugares o mesmo resultado variava consideravelmente. Isto foi observado por Hiparco (c. 180-125 a.C) que mediu o brilho e a posição das estrelas a olho nu, Erastótenes (c. 276-194 a.C) que mediu pela primeira vez o raio da Terra e Aristarco (c. 310-230 a.C) que mediu as distâncias relativas do Sol e da Lua (Nunes, 1997). Como se sabe a preocupação, para esses astrônômos, estava em como se determinar essas grandezas, e não em medidas precisas.

Tycho Brahe (1546-1601) teve um grande desenvolvimento na era pré-telescópica: em seus registros constam de dados obtidos com uma variação esta descrita pela média ou mediana dos dados. Em meados do Século XVIII surge o Método dos Mínimos Quadrados através dos estudo de parâmetros de órbitas de cometas, desenvolvida inicialmente por Legendre (1805) e anos mais tarde por Gauss (1809)

Segundo Nunes (apud Plackett, 1972) com publicações em datas tão próximas, estes dois matemáticos, vieram a envolver-se em uma polêmica sobre a autoria da descoberta. Embora que Légendre tenha divulgado primeiro os seus resultados, sabe-se que Gauss os tinha obtido muito antes, em 1794–1795, pelo que hoje se atribui a este último a prioridade da criação do método.

Após um século Pearson e Fischer desenvolveram e sistematizaram de uma forma elegante as ferramentas da Estatística já conhecidas anteriormente. A grande contribuição de Pearson foi o coeficiente de correlação de Pearson.

Neste trabalho utilizou-se o Método dos Mínimos Quadrados (MMQ). No Capítulo 1 traz-se a teoria matemática para o caso discreto e contínuo. Ainda neste capítulo, obteve-se o coeficiente de correlação de Pearson.

O Capítulo 2 são tratadas as aplicações do MMQ no caso discreto e contínuo, estas encontradas nas referências bibliográficas. Para o caso discreto apresentam-se duas aplicações, uma na econometria e outra na música. Para o caso contínuo, expõe-se um exemplo de uma função exponencial, um caso aplicado de interesse para a área de matemática.

Pode-se utilizar o (MMQ) no Ensino Médio onde dois fenômenos podem ser correlacionados. Para tal utiliza-se tabelas de medidas relativas a estes fenômenos. Deve-se auxiliar o aluno nos cálculos de somatórios e aplicando-as nas devidas fórmulas.

1 - TEORIA MATEMÁTICA

Neste primeiro capítulo trataremos do ajuste de pontos seja ele através de uma função, este estudo chama-se Método dos Mínimos Quadrados (MMQ), será estudado o caso discreto e contínuo. Também será estudado o desvio padrão de dados brutos e em seguida o erro padrão de estimativa e o coeficiente de correlação de Pearson.

1.1. Método dos Mínimos Quadrados – Caso Discreto

Dados os pontos (X_i, Y_i) $i = 0, 1, 2, \dots, n$, pretende-se encontrar uma função da forma:

$$f(X) = \sum_{j=0}^m a_j f_j(X), \quad n \geq m+1 \quad (1)$$

sendo f_j funções contínuas e a_j coeficientes reais, tal que a função f , escolhida *a priori*, minimize o funcional:

$$M(a_0, \dots, a_n) = \sum_{i=0}^n (Y_i - f(X_i))^2 \quad (2)$$

A distância entre Y_i e $f(X_i)$ poderá ser chamado de resíduo ou erro, como se vê na Figura 1, um ajuste padrão para o caso discreto.

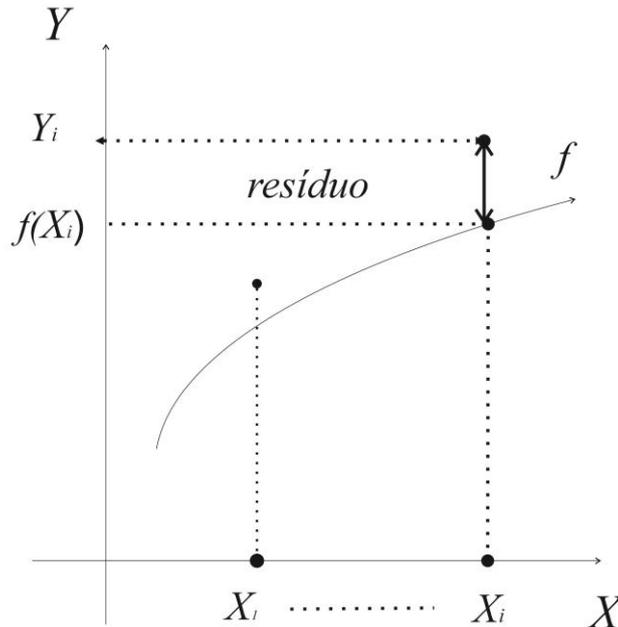


Figura 1. Representação de ponto (X_i, Y_i) e da função de ajuste com o respectivo resíduo.

Assim têm-se pontos conhecidos, (X_i, Y_i) , $i = 0, 1, 2, \dots, n$, e pretende-se aproximá-los por uma combinação de funções quaisquer, por exemplo: lineares, exponenciais, periódicas, etc.

Para que o funcional M seja mínimo, faz-se:

$$\frac{\partial M}{\partial a_j} = 0, \quad j = 0, 1, 2, \dots, m \quad (3)$$

Na Equação (2) aplica-se a derivada parcial com relação a a_j , $j = 0, 1, 2, \dots, m$, e obtém-se:

$$\frac{\partial M}{\partial a_j} = -2 \sum_{i=0}^n [Y_i - f(X_i)]^2$$

Aplicando as regras de derivação, obtém-se:

$$\frac{\partial M}{\partial a_j} = -2 \sum_{i=0}^n [Y_i - f(X_i)] \frac{\partial}{\partial a_j} f(X_i)$$

Substituindo a Equação (1) na expressão acima, tem-se:

$$\frac{\partial M}{\partial a_j} = -2 \sum_{i=0}^n [Y_i - f(X_i)] \cdot \frac{\partial}{\partial a_j} \sum_{k=0}^m a_k f_k(X_i).$$

Assim

$$\frac{\partial M}{\partial a_j} = -2 \sum_{i=0}^n [Y_i - f(X_i)] \sum_{k=0}^m f_k(X_i) \frac{\partial a_k}{\partial a_j} \quad (4)$$

Como $\delta_{kj} = \frac{\partial a_k}{\partial a_j}$, sendo δ_{kj} o delta de Kronecker, o qual é definido por:

$$\delta_{kj} = \begin{cases} 0, & \text{se } k \neq j \\ 1, & \text{se } k = j \end{cases} \quad (5)$$

Com isto, a partir da Equação (4), obtém-se:

$$\frac{\partial M}{\partial a_j} = -2 \sum_{i=0}^n [Y_i - f(X_i)] f_j(X_i)$$

Considerando a Equação (3) e substituindo a Equação (1) na expressão acima, tem-se .

$$0 = -2 \sum_{i=0}^n [Y_i - \sum_{k=0}^m a_k f_k(X_i)] f_j(X_i), \text{ com } j = 0, 1, 2, \dots, m. \quad (6)$$

Aplicando lei do cancelamento do produto, o ponto de mínimo do funcional M , em relação a a_j , é descrito como:

$$0 = - \sum_{i=0}^n Y_i f_j(X_i) + \sum_{i=0}^n \sum_{k=0}^m a_k f_k(X_i) f_j(X_i)$$

Dai, tem-se:

$$\sum_{i=0}^n \sum_{k=0}^m a_k f_k(X_i) f_j(X_i) = \sum_{i=0}^n Y_i f_j(X_i)$$

Para cada $j = 0, 1, 2, \dots, m$, obtém-se o sistema.

$$\left\{ \begin{array}{l} \sum_{i=0}^n \sum_{k=0}^m a_k f_k(X_i) f_0(X_i) = \sum_{i=0}^n Y_i f_0(X_i) \\ \sum_{i=0}^n \sum_{k=0}^m a_k f_k(X_i) f_1(X_i) = \sum_{i=0}^n Y_i f_1(X_i) \\ \vdots \\ \sum_{i=0}^n \sum_{k=0}^m a_k f_k(X_i) f_m(X_i) = \sum_{i=0}^n Y_i f_m(X_i) \end{array} \right. \quad (7)$$

No sistema de equações (7), tomando-se o lado esquerdo da equação com $j = s$, fixo, $s = 0, 1, \dots, m$, tem-se o seguinte desenvolvimento:

$$\sum_{i=0}^n \sum_{k=0}^m a_k f_k(X_i) f_s(X_i) = \sum_{i=0}^n (a_0 f_0(X_i) f_s(X_i) + \dots + a_m f_m(X_i) f_s(X_i))$$

Obtendo assim o seguinte:

$$\sum_{i=0}^n \sum_{k=0}^m a_k f_k(X_i) f_s(X_i) = a_0 \sum_{i=0}^n f_0(X_i) f_s(X_i) + \dots + a_m \sum_{i=0}^n f_m(X_i) f_s(X_i) \quad (8)$$

Substituindo a Equação (8) no sistema de equações (7), tem-se:

$$\left\{ \begin{array}{l} a_0 \sum_{i=0}^n f_0(X_i) f_0(X_i) + \dots + a_m \sum_{i=0}^n f_m(X_i) f_0(X_i) = \sum_{i=0}^n Y_i f_0(X_i) \\ a_0 \sum_{i=0}^n f_0(X_i) f_1(X_i) + \dots + a_m \sum_{i=0}^n f_m(X_i) f_1(X_i) = \sum_{i=0}^n Y_i f_1(X_i) \\ \vdots \\ a_0 \sum_{i=0}^n f_0(X_i) f_m(X_i) + \dots + a_m \sum_{i=0}^n f_m(X_i) f_m(X_i) = \sum_{i=0}^n Y_i f_m(X_i) \end{array} \right.$$

Como são dados os pontos (X_i, Y_i) , $i = 0, 1, 2, \dots, n$, e as funções contínuas, f_j , $j = 0, 1, 2, \dots, m$, motivado pelo problema fenômeno estudado, deve-se determinar os coeficientes reais a_j . O sistema acima pode ser escrito na forma matricial como segue:

$$\begin{bmatrix} \sum_{i=0}^n f_0(X_i)f_0(X_i) & \dots & \sum_{i=0}^n f_m(X_i)f_0(X_i) \\ \sum_{i=0}^n f_0(X_i)f_1(X_i) & \dots & \sum_{i=0}^n f_m(X_i)f_1(X_i) \\ \vdots & & \vdots \\ \sum_{i=0}^n f_0(X_i)f_m(X_i) & \dots & \sum_{i=0}^n f_m(X_i)f_m(X_i) \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^n Y_i f_0(X_i) \\ \sum_{i=0}^n Y_i f_1(X_i) \\ \vdots \\ \sum_{i=0}^n Y_i f_m(X_i) \end{bmatrix} \quad (9)$$

1.2. Método dos Mínimos Quadrados – Caso Contínuo

Sejam os intervalos $X=[a,b]$, $Y=[\alpha,\beta]$ e o conjunto, $\Omega[X,Y]$, das funções contínuas $f:X \rightarrow Y$, munido das operações multiplicação e adição de funções. Logo, $\Omega[X,Y]$ é em espaço vetorial com produto interno, definido por:

$$\langle f(X), g(X) \rangle = \int_a^b f(X)g(X) dX \quad (10)$$

$\forall f, g \in \Omega[X,Y]$. A norma é definida por $\|f\|^2 = \langle f, f \rangle$.

Problema: “Dada, $f \in \Omega[X,Y]$, pretende-se encontrar um polinômio:

$$p(X) = \sum_{i=0}^n a_i X^i, \quad n > 0 \quad (11)$$

que minimize:

$$M(a_0, \dots, a_n) = \int_a^b (f(x) - p(x))^2 dx \quad (12)$$

Desta forma, para que M seja um extremo, tem-se:

$$\frac{\partial M}{\partial a_j} = 0, \quad j = 0, 1, 2, \dots, n. \quad (13)$$

Para $j = 0, 1, 2, \dots, n$, calcula-se:

$$\frac{\partial M}{\partial a_j} = \frac{\partial}{\partial a_j} \int_a^b [f(X) - p(X)]^2 dX = \int_a^b \frac{\partial}{\partial a_j} [f(X) - p(X)]^2 dX$$

Para facilitar a notação, omite-se a variável X das respectivas funções. Como a função f independe dos coeficientes a_i , $i=0, \dots, n$, pela regra da cadeia, obtém-se:

$$\frac{\partial M}{\partial a_j} = -2 \int_a^b (f - p) \frac{\partial p}{\partial a_j} dX, \quad (14)$$

Considerando a Equação (11), faz-se:

$$\frac{\partial p}{\partial a_j} = \frac{\partial}{\partial a_j} \left[\sum_{i=0}^n a_i X^i \right] \Rightarrow \frac{\partial p}{\partial a_j} = \sum_{i=0}^n X^i \frac{\partial a_i}{\partial a_j}$$

Utilizando a Equação (5), delta de Kronecker, a expressão acima fica escrita como:

$$\frac{\partial p}{\partial a_j} = \sum_{k=0}^n X^k \frac{\partial a_k}{\partial a_j} \Rightarrow \frac{\partial p}{\partial a_j} = X^j, j = 0, 1, 2, \dots, n \quad (15)$$

Substituindo a Equação (15) em (14) teremos:

$$\frac{\partial M}{\partial a_j} = -2 \int_a^b (f - p) X^j dX = -2 \int_a^b f X^j dX + 2 \int_a^b p X^j dX \quad (16)$$

Trabalhando apenas com, $\int_a^b p X^j dX$, tem-se :

$$\int_a^b p X^j dX = \int_a^b \left(\sum_{k=0}^n a_k X^k \right) X^j dX$$

Assim:

$$\int_a^b p X^j dX = \sum_{i=0}^n a_k \int_a^b X^k X^j dX, j = 0, 1, 2, \dots, n$$

Substituindo a expressão (15) na Equação (16), tem-se.

$$\frac{\partial M}{\partial a_j} = -2 \int_a^b f X^j dX + 2 \sum_{k=0}^n a_k \int_a^b X^k X^j dX, j = 0, 1, 2, \dots, n \quad (17)$$

Chamando $A_{kj} = \int_a^b X^k X^j dX$, $j = 0, 1, 2, \dots, n$, e substituindo na Equação (17) a Equação

(13), obtém-se:

$$0 = -2 \int_a^b f X^j dX + 2 \sum_{k=0}^n a_k \int_a^b A_{kj},$$

do qual se pode chegar a seguinte expressão:

$$\int_a^b f X^j dX = \sum_{k=0}^n a_k A_{kj}.$$

Como $j = 0, 1, 2, \dots, n$, as equações acima podem ser descritas através do seguinte sistema:

$$\begin{cases} \sum_{k=0}^n a_k A_{k0} = \int_a^b f X^0 dX \\ \vdots \\ \sum_{k=0}^n a_k A_{kn} = \int_a^b f X^n dX \end{cases}$$

ou na forma matricial, como segue:

$$\begin{bmatrix} A_{00} & \cdots & A_{n0} \\ \vdots & \cdots & \vdots \\ A_{0n} & \cdots & A_{nn} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \int_a^b fX^0 dX \\ \vdots \\ \int_a^b fX^n dX \end{bmatrix} \quad (18)$$

sendo $A_{kj} = \int_a^b X^k X^j dX = \int_a^b X^{k+j} dX = \frac{X^{k+j+1}}{k+j+1} \Big|_a^b \quad k, j = 0, 1, 2, \dots, n$

1.3. Coeficiente de Correlação de Pearson

Desenvolve-se abaixo a teoria estatística para obtenção do coeficiente de correlação de Pearson, Este coeficiente mede o grau de correlação linear entre duas variáveis quantitativas.

1.3.1. DESVIO PADRÃO DE DADOS ABSOLUTOS

Seja o conjunto $X = \{x_1, \dots, x_n\}$. Por definição tem-se que a média aritmética destes pontos será dada por:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (19)$$

O desvio padrão, que é a média quadrática dos afastamentos em relação à média aritmética dos elementos de X , é dado por:

$$S_x = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (20)$$

Utilizando as propriedades do somatório, obtém-se a seguinte expressão:

$$\sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (21)$$

Com isso, escreve-se o desvio-padrão de uma conjunto, na forma:

$$S_x = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2} \quad (22)$$

1.3.2. ERRO PADRÃO DE ESTIMATIVA

A simbologia utilizada nesta seção é independente dos tópicos anteriores.

Sejam $\tilde{X} = \{x_1, \dots, x_n\}$, $\tilde{Y} = \{y_1, \dots, y_n\}$ e $\tilde{X} \times \tilde{Y} = \{(x_i, y_i) : x_i \in \tilde{X} \text{ e } y_i \in \tilde{Y}\}$. Para verificar o quanto uma linha reta representa a relação entre duas variáveis, utiliza-se a equação da reta de regressão de mínimos quadrados. Uma estimativa para esta reta é:

$$Y = a_0 + a_1 X, \text{ com } x_1 \leq X \leq x_n, x_1, x_n \in \tilde{X} \quad (23)$$

$$X = b_0 + b_1 Y, \text{ com } y_1 \leq Y \leq y_n, y_1, y_n \in \tilde{Y} \quad (24)$$

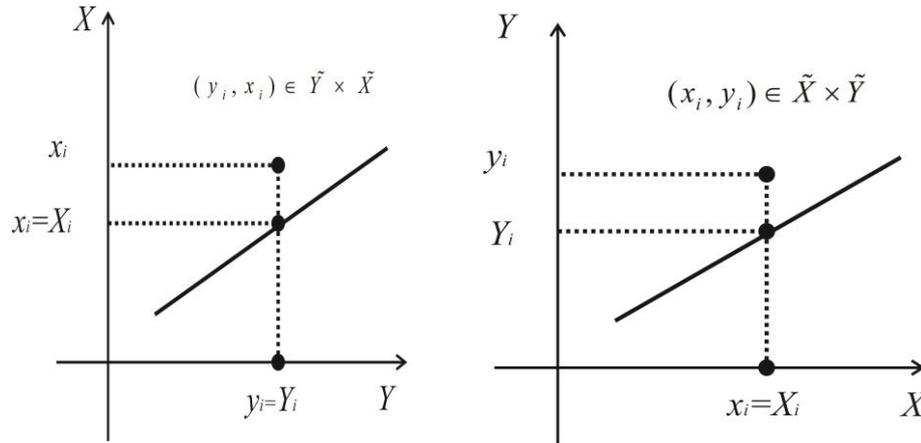


Figura 2. Representação das estimativas relativas às equações (23) á direita e (24) á esquerda.

Analogamente como feito para o desvio padrão, faz-se:

$$S_{xy} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{x})^2} : \text{Erro padrão de estimativa de } X \text{ para } Y$$

$$S_{yx} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \bar{y})^2} : \text{Erro padrão da estimativa } Y \text{ para } X$$

sendo \bar{x} e \bar{y} as médias aritméticas, respectivamente, dos elementos dos conjuntos \tilde{X} e \tilde{Y} .

A *variação total* de Y é dada por: $\sum_{i=1}^n (\bar{y} - y_i)^2$. Considerando $\bar{y} - y_i = (\bar{y} - Y_i) + (Y_i - y_i)$, logo

$$(\bar{y} - y_i)^2 = [(\bar{y} - Y_i)^2 + 2(\bar{y} - Y_i)(Y_i - y_i) + (Y_i - y_i)^2] \quad (25)$$

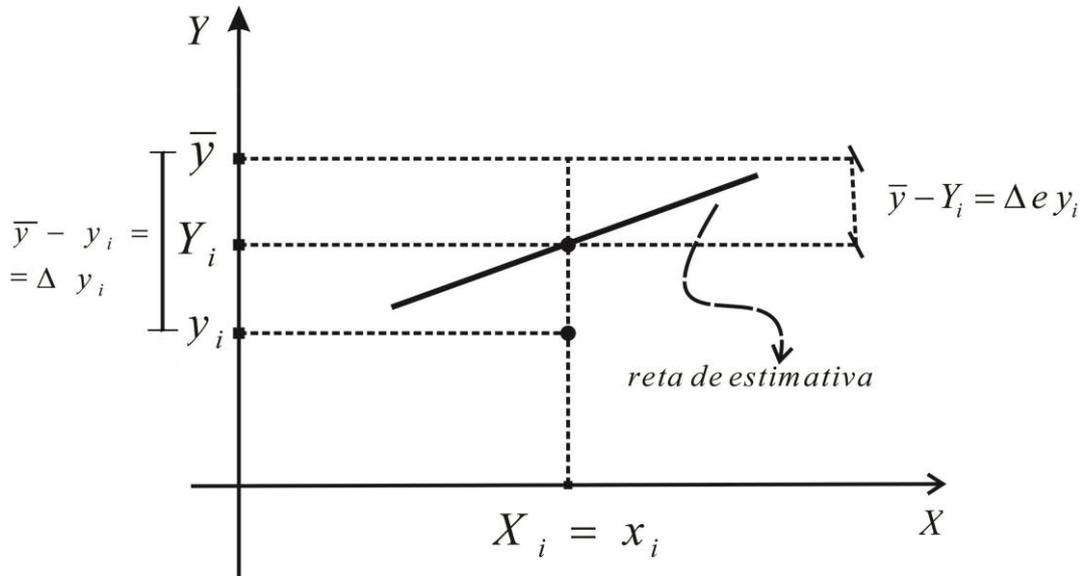


Figura 3. Representação variação total da Equação (25) representada no plano cartesiano.

Daí, a variação total de Y tem a seguinte forma:

$$\sum_{i=1}^n (\bar{y} - y_i)^2 = \sum_{i=1}^n (\bar{y} - Y_i)^2 + \sum_{i=1}^n (Y_i - y_i)^2 + 2 \sum_{i=1}^n (\bar{y} - Y_i)(Y_i - y_i) \quad (26)$$

Analisando a última parcela da equação acima e substituindo $Y_i = a_0 + a_1 X_i$ obtém-se:

$$\sum_{i=1}^n (\bar{y} - Y_i)(Y_i - y_i) = \bar{y} \sum_{i=1}^n (Y_i - y_i) - a_0 \left(\sum_{i=1}^n Y_i - y_i \right) - a_1 \sum_{i=1}^n X_i (Y_i - y_i) \quad (27)$$

Pela regressão linear tem-se a soma dos quadrados:

$$S = \sum_{i=1}^n (Y_i - y_i)^2 \Rightarrow S = \sum_{i=1}^n (a_0 + a_1 X_i - y_i)^2$$

e daí faz-se:

$$\frac{dS}{da_0} = 0 \Rightarrow 2a_0 \sum_{i=1}^n (Y_i - y_i) = 0 \underbrace{\Rightarrow}_{2a_0 \neq 0} \sum_{i=1}^n (Y_i - y_i) = 0$$

Faz-se ainda:

$$\frac{dS}{da_1} = 0 \Rightarrow 2a_1 \sum_{i=1}^n X_i (Y_i - y_i) = 0 \underbrace{\Rightarrow}_{2a_1 \neq 0} \sum_{i=1}^n (Y_i - y_i) = 0$$

Substituindo as Equações acima em (27) obtém-se:

$$\sum_{i=1}^n (\bar{y} - Y_i)(Y_i - y_i) = 0$$

Portando, a variação total é expressa como segue:

$$\sum_{i=1}^n (\bar{y} - y_i)^2 = \sum_{i=1}^n (\bar{y} - Y_i)^2 + \sum_{i=1}^n (Y_i - y_i)^2 \quad (28)$$

sendo $\sum_{i=1}^n (\bar{y} - Y_i)^2$ a variação explicada e $\sum_{i=1}^n (Y_i - y_i)^2$ a variação não explicada.

1.3.3. COEFICIENTE DE DETERMINAÇÃO

O coeficiente de determinação, r^2 , é dado pela seguinte razão

$$r^2 = \frac{\text{variação explicada}}{\text{variação total}} = \frac{\sum_{i=1}^n (\bar{y} - Y_i)^2}{\sum_{i=1}^n (\bar{y} - y_i)^2}, \quad (29)$$

Nomeando cada fator onde cada uma tem o seu significado

$$\Delta e Y_i = \sum_{i=1}^n (\bar{y} - Y_i)^2 \text{ que é o erro de } Y_i \text{ denotado } \Delta e Y_i$$

$$\text{Analogamente ao que se fez na Equação (21)} \quad \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2$$

(21tem-se:

$$\sum_{i=1}^n (\bar{y} - y_i)^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \quad (30)$$

E como

$$S_y = \sqrt{\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2} \Rightarrow n(S_y)^2 = \sum_{i=1}^n (\bar{y} - y_i)^2$$

Pela equação (28)

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

Dividindo nos dois lados da igualdade por $\sum_{i=1}^n (\bar{y} - y_i)^2$ na equação acima, tem-se

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

Substituindo a Equação (29) e (30) na igualdade acima obtém-se

$$\text{Assim } r^2 = 1 - \frac{S_{yx}^2}{S_y^2} \Rightarrow \frac{S_{yx}^2}{S_y^2} = 1 - r^2 \Rightarrow \frac{S_{yx}}{S_y} = \sqrt{1 - r^2}$$

Utilizando o desenvolvimento do Apêndice A, obtém-se o *coeficiente de correlação de Pearson*:

$$r = \frac{\left(\sum_{i=1}^n \Delta x_i \Delta y_i \right)}{\sqrt{\sum_1^n (\Delta x_i)^2 \sum_1^n (\Delta y_i)^2}} = \frac{n \left(\sum_{i=1}^n x_i y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left[n \left(\sum_{i=1}^n x_i \right)^2 - \sum_{i=1}^n x_i^2 \right] \cdot \left[n \left(\sum_{i=1}^n y_i \right)^2 - \sum_{i=1}^n y_i^2 \right]}} \quad (31)$$

onde tem-se o intervalo $-1 \leq r \leq 1$. Portanto pode-se avaliar a correlação entre os dados de duas variáveis qualitativamente, como se descreve a seguir:

$0,0 < r < 0,3$	Correlação Linear Fraca
$0,3 \leq r < 0,6$	Correlação Linear Moderada
$0,6 \leq r < 0,9$	Correlação Linear Forte
$0,9 \leq r < 1,0$	Correção Linear Muito Forte

e

$-1,0 < r < -0,9$	Correlação Linear Muito Forte Negativa
$-0,9 \leq r < -0,6$	Correlação Linear Forte Negativa
$-0,6 \leq r < -0,3$	Correlação Linear Moderada Negativa
$-0,3 \leq r < 0,0$	Correlação Fraca Negativa

2 – APLICAÇÕES DO MÉTODO DOS MÍNIMOS QUADRADOS

Neste capítulo reserva-se a aplicações dos estudos do capítulo 1. Na primeira aplicação tem-se uma regressão com ajuste linear, já na segunda aplicação tem-se o ajuste exponencial. E por fim a aplicação 3, um caso contínuo.

2.1. APLICAÇÃO 1: ENERGIA ELÉTRICA – DEMANDA VERSUS TARIFA MÉDIA

Na Tabela 1 dispõem-se os dados relativos às variáveis: índice da quantidade demandada, Q, (fonte: IBGE (1987) (1981-84) e BACEN (1990)), e índice da tarifa real média, T, (fonte: Eletrobrás-Centrals Elétricas Brasileiras S.A). Para verificar se existe relação entre as referidas variáveis, utiliza-se o método dos mínimos quadrados .

Tabela 1. Dados relativos ao índice da quantidade demandada e índice da tarifa real média, sendo $n=10$.

i	Ano	Q	T
0	1981	69	143
1	1982	76	134
2	1983	81	117
3	1984	90	111
4	1985	94	109
5	1986	100	100
6	1987	103	137
7	1988	108	122
8	1989	113	85
9	1990	115	90

A partir da Equação (9) considera-se $m=1$, tem-se a função estimativa:

$$f(Q) = a_0 f_0(Q) + a_1 f_1(Q)$$

Tomando-se f sendo uma função linear, tem-se $f_0(Q) = 1$ e $f_1(Q) = Q$. Daí:

$$f(Q) = a_0 + a_1 Q$$

Logo, a Equação (9) fica escrita como:

$$\begin{bmatrix} \sum_{i=0}^9 f_0^2(Q_i) & \sum_{i=0}^9 f_1(Q_i) f_0(Q_i) \\ \sum_{i=0}^9 f_0(Q_i) f_1(Q_i) & \sum_{i=0}^9 f_1^2(Q_i) \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^9 T_i f_0(Q_i) \\ \sum_{i=0}^9 T_i f_1(Q_i) \end{bmatrix} \quad (32)$$

Na Tabela 2, tem-se os cálculos necessários para se resolver o sistema acima, nas incógnitas a_0 e a_1 .

Tabela 2. Cálculos para ajuste linear, considerando a variável independente, Q, e a variável independente, T, tomando $f_0(Q)=1$ e $f_1(Q)=Q$.

Ano	Q	T	$f_0(Q_i)$	$f_1(Q_i)$	$f_0(Q_i)f_1(Q_i)$	$f_0^2(Q_i)$	$f_1^2(Q_i)$	$T_i f_0(Q_i)$	$T_i f_1(Q_i)$
0	69	143	1	69	69	1	4761	143	9867
1	76	134	1	76	76	1	5776	134	10184
2	81	117	1	81	81	1	6561	117	9477
3	90	111	1	90	90	1	8100	111	9990
4	94	109	1	94	94	1	8836	109	10246
5	100	100	1	100	100	1	10000	100	10000
6	103	137	1	103	103	1	10609	137	14111
7	108	122	1	108	108	1	11664	122	13176
8	113	85	1	113	113	1	12769	85	9605
9	115	90	1	115	115	1	13225	90	10350
949	1148	10	10	949	949	10	92301	1148	107006

Na Equação (32) substituímos os valores da Tabela 2:

$$\begin{bmatrix} 10,0 & 949,0 \\ 949,0 & 92301,0 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1148,0 \\ 107006,0 \end{bmatrix}$$

Resolvendo o sistema acima, obtém-se: $a_0 = 196,92$ e $a_1 = -0,86$.

O gráfico de dispersão, juntamente com a linha de tendência linear é mostrado na Figura 4.

Tabela 3. Cálculos para o coeficiente de correlação de Pearson.

Ano	Q	T	$f_1^2(Q_i)$	T^2	$T_i f_1(Q_i)$
0	69	143	4761	20449	9867
1	76	134	5776	17956	10184
2	81	117	6561	13689	9477
3	90	111	8100	12321	9990
4	94	109	8836	11881	10246
5	100	100	10000	10000	10000
6	103	137	10609	18769	14111
7	108	122	11664	14884	13176
8	113	85	12769	7225	9605
9	115	90	13225	8100	10350
949	1148	92301	135274	107006	

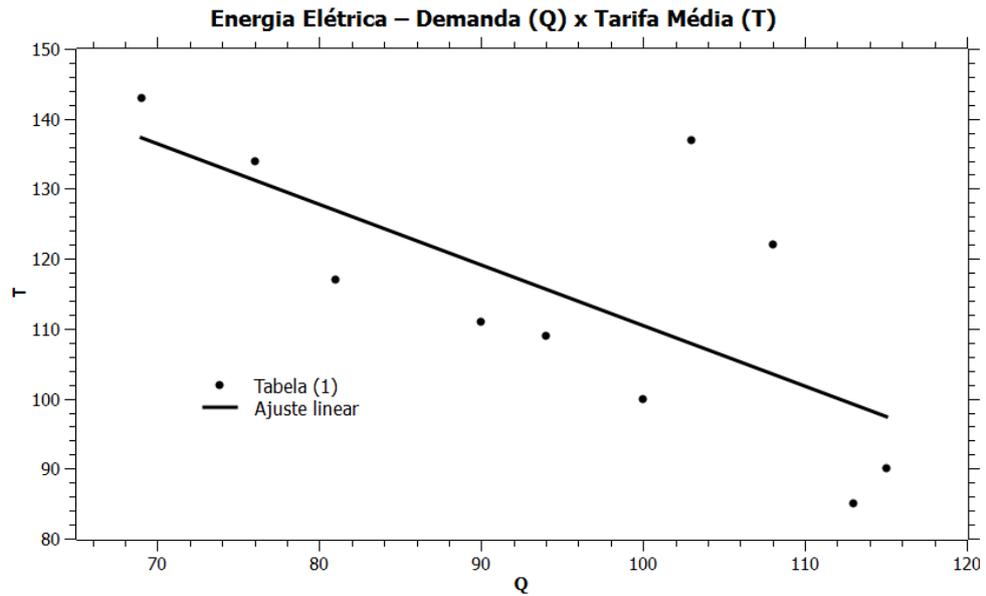


Figura 4. Representação do gráfico de dispersão e da linha de tendência linear, cuja equação é dada por:

$$f(Q) = T = -0,86Q + 196,92$$

Faz-se o cálculo do coeficiente de correlação de Pearson, utilizando a Equação (31) e os dados da Tabela 3, assim:

$$r = \frac{10(107006) - 949(1148)}{\sqrt{[10(92301) - (949)^2] \cdot [10(135274) - (1148)^2]}} = -0,69$$

Como $r = -0,69$ tem-se que os dados das variáveis T e Q tem ajuste negativo forte. Com isto, o modelo linear descreve o comportamento aproximado da tarifa média, T , em função da demanda, Q .

2.2. Aplicação 2: Notas do Teclado de um Piano versus Frequência

A Tabela 4 representa as frequências das teclas de um piano em relação á escala média, referente a teclas 25 a 37 numeradas do piano.

Tabela 4. Dados relativos à frequência em Hertz de cada tecla da escala temperada. Fonte Casio (Casio, acesso 10/2014)

Nota (x)	Freq (y)
1	263,15
2	277,77
3	294,11
4	312,5
5	333,33
6	344,82
7	370,37
8	400,00
9	416,66
10	454,54
11	476,19
12	500,00
13	526,31

Considerando que este fenômeno pode ser modelado pela expressão:

$$y = a e^{bx} \quad (33)$$

sendo $a, b \in R$ os parâmetros a serem estimados.

Para se utilizar o método dos mínimos quadrados, deve-se linearizar a Equação (33) como se segue:

$$\ln(y) = \ln(a e^{bx}) \Rightarrow \ln(y) = \ln(a) + bx$$

Chamando: $Y = \ln(y)$, $A = \ln(a)$, tem-se:

$$Y = A + bx \quad (34)$$

Os parâmetros A e b serão estimados pelo método dos mínimos quadrados.

A partir da Tabela 5 e utilizando o programa SciDAVis, calcula-se os seguintes parâmetros: $A = 5,51$ e $b = 0,06$, cujo coeficiente de determinação é $R^2 = 0,998$. Como

$$A = \ln(a) \Rightarrow a = e^A \Rightarrow a = e^{5,51} \Rightarrow a = 246,81$$

Assim o modelo expresso pela Equação (33) fica determinado:

$$y = 246,81 e^{0,06x} \quad (35)$$

Constrói-se o gráfico de dispersão, Figura 5, onde é apresentada a curva de tendência dada pela Equação (35).

Tabela 5. Dados providos da Tabela 5 e linearizados.

x	$Y = \ln(y)$
1	5,57
2	5,63
3	5,68
4	5,74
5	5,81
6	5,84
7	5,91
8	5,99
9	6,03
10	6,12
11	6,17
12	6,21
13	6,27

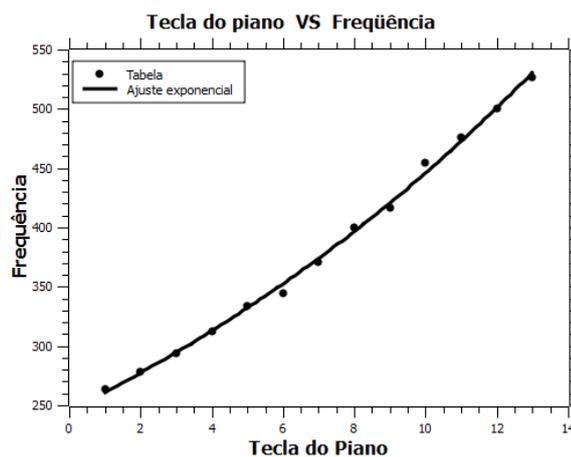


Figura 5. Gráfico do modelo exponencial que relaciona as notas das teclas *versus* a frequência (Hz)..

O coeficiente de correlação de Pearson é $r = 0,999$ e isto garante que o modelo, Equação (35), tem correlação linear muito forte com os dados providos da Tabela 4. Portanto, o modelo exponencial descreve o comportamento aproximado da frequência, y , em função da nota, x .

2.3. Aproximação de uma função qualquer por um polinômio- caso contínuo

Pretende-se aproximar a função definida por $f(x) = e^x$, no intervalo $[0,1]$, por um polinômio p , de grau 1.

Utilizando a equação (18), tem-se:

$$\begin{bmatrix} A_{00} & A_{01} \\ A_{10} & A_{11} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \int_0^1 e^x dx \\ \int_0^1 e^x x dx \end{bmatrix}$$

sendo $A_{kj} = \frac{X^{k+j+1}}{k+j+1} \Big|_0^1 = \frac{1}{k+j+1}$ $k, j = 0, 1$. isto é, $A_{00} = 1$, $A_{01} = A_{10} = \frac{1}{2}$ e $A_{11} = \frac{1}{3}$.

Calculando as integrais do lado direito da equação matricial acima, tem-se:

$$\int_0^1 e^x dx = e - 1$$

$$\int_0^1 e^x x dx = e - e + 1 = 1$$

Substituindo os resultados acima na expressão matricial, tem-se:

$$\begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} e-1 \\ 1 \end{bmatrix}$$

Assim, pelo Método de Gauss, calculam-se as soluções: $a_0 = 4e - 10$ e $a_1 = 18 - 6e$. Logo, obtém-se o polinômio $p(x) = (18 - 6e)x + 4e - 10$.

Na Figura 6 apresenta-se o polinômio de grau 1, $p(x) = (18 - 6e)x + 4e - 10$, que se ajusta à função exponencial $f(x) = e^x$ no intervalo $[0,1]$.

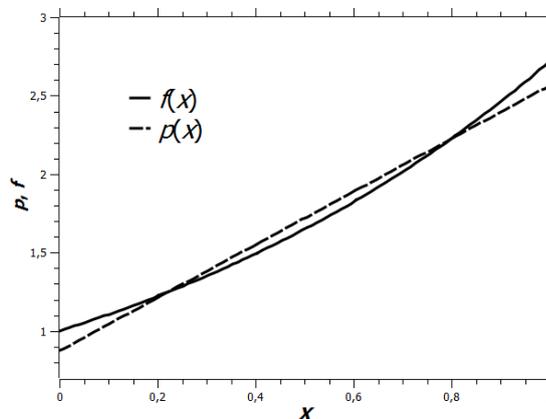


Figura 6. Gráfico que aproxima a função $f(x) = e^x$ por um polinômio p , de grau 1.

3 - CONCLUSÃO

Neste trabalho estudou-se o ajuste de curvas por Métodos dos Mínimos Quadrados (MMQ) e o coeficiente que determina a confiabilidade da aproximação.

Observa-se através deste desenvolvimento que há dois casos de (MMQ), o discreto e o contínuo. Apresentou as suas aplicações. O primeiro caso discreto apresentado na seção 2.1 teve um ajuste linear forte negativo observado através do coeficiente de Pearson, portanto para este ajuste as distâncias dos pontos experimentais são pequenas em relação aos pontos não experimentais da função de tendência. Na seção 2.2 o coeficiente de Pearson foi muito bom, portanto o ajuste dos pontos dados pode ser descrito exponencialmente. Na última aplicação foi um caso contínuo exposto em 2.3, onde a função exponencial apresentada teve um ajuste linear bom, no entanto esta conclusão teve-se como referência apenas em relação ao gráfico.

Para futuros estudos sugere-se a aplicação do Método dos Mínimos Quadrados no ensino básico, de preferência no ensino médio.

REFERÊNCIAS

- ABDI, H. Least Square. Encyclopedia of Research Design. Disponível em: <https://www.utdallas.edu/~herve/abdi-LeastSquares2010-pretty.pdf> . Acesso em: 10 out. 2014.
- BARROS, L. C., BASSANEZI, R. C. Tópicos de Lógica *Fuzzy* e Biomatemática. Coleção IMECC - Textos didáticos v.5, Campinas, 2006.
- BASSANESI, R.C. Modelagem Matemática. *Dynamus*, Blumenau, v. 1, n. 7, p. 55-83, abr./jun. 1994.
- BASSANEZI, R. C. Ensino-Aprendizagem com Modelagem Matemática: uma Nova Estratégia. Editora Contexto, São Paulo, 2006.
- BUSSAB, W.O, MORRETTIN P. A. Estatística Básica. 5.ed. São Paulo: Saraiva, 2003.

- BOLDRINI, J. L.; COSTA, S. L. R.; FIGUEIREDO, V. L., WETZLER, H. G. Álgebra Linear, 3a edição, Editora Harbra, São Paulo, 1986.
- CHARNET, R. et al. Análise de Modelos de Regressão Linear com Aplicações – Campinas, SP: Editora da Unicamp, 1999.
- CUNHA, M. C. C. Métodos Numéricos. Editora Unicamp, Campinas-SP, 2000.
- GUJARATI, D. N. Econometria Básica. 4ª ed. Rio de Janeiro: Campus, 2006
- GUIDORIZZI, H. Um Curso de Cálculo, Vol. 1, LTC Editora, 5a. Edição, 2011.
- GUIDORIZZI, H. Um Curso de Cálculo, Vol. 2, LTC Editora, 5a. Edição, 2001..
- HILL, C., GRIFFITHS, W., JUDGE, G. Econometria, Ed. Saraiva, São Paulo, 1999.
- HUMES, A.F.P.C, MELO, I.S.H., YOSHIDA, L.K., MARTINS, W.T., Noções de Cálculo Numérico. Editora LTC, Rio de Janeiro, 2001.
- MADDALA, G. S. Introdução à Econometria. 3ª ed. Rio de Janeiro: LTC. 2003
- MARGARIDA, A. S. D. Modelos Matemáticos na Música. Revista Casio News nº9, Lisboa-Portugal, pág. 3. 2013.
- MATOS, O. C. Econometria Básica: Teoria e Aplicações. Atlas, São Paulo, 2000.
- MORETTIN, L. G. Estatística Básica - Probabilidade (vol.1). Makron, São Paulo, 2002.
- MORETTIN, L. G. Estatística Básica - Inferência (vol.2). Makron, , São Paulo, 2002.
- SciDAVIs. Disponível em: <http://scidavis.sourceforge.net/>. Acesso em: 10 out. 2014.

SPIEGEL, M. R.; STEPHENS, L. J. NASCIMENTO, J. L. Estatística - Coleção Schaum.
Bookman , São Paulo , 2009.

VIEIRA, S. Introdução à Bioestatística. Ed. Campus, Rio de Janeiro, 1981.

WOOLDRIDGE, J. Introduction to Econometrics: A Modern Approach. Thomson – South
Western, 2003.

Apêndice A

Pelo mesmo raciocínio da Equação (20).

$$\sum_1^n \Delta x_i \Delta y_i = \sum_1^n x_i y_i - \frac{1}{n} \sum_1^n x_i y_i \quad (36)$$

Então pode-se escrever:

$$\sum_{i=1}^n \Delta x_i^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (37)$$

Pelo desenvolvimento de regressão linear simples , e se $Y = a_0 + a_1 X$, logo se

$$a_0 = \frac{\left(\sum_{i=1}^n y_i \right) \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n x_i y_i \right)}{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2} \quad (38)$$

$$a_1 = \frac{n \left(\sum_{i=1}^n x_i y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \left(\sum_{i=1}^n x_i^2 \right) - \left(\sum_{i=1}^n x_i \right)^2} \quad (39)$$

Substituindo as Equações (36)e (37) na Igualdade (38) , tem-se

$$a_1 = \frac{\left(\sum_{i=1}^n \Delta x_i \Delta y_i \right)}{n \left(\sum_{i=1}^n \Delta x_i^2 \right)} \quad (40)$$

Se a média dos elementos de x for zero, logo

$$Y = \bar{y} + a_1 \Delta x_i \text{ se } x_i = X_i$$

Como $\Delta e Y_i = \bar{y} - Y_i = -a_1 X_i$, como

$$r^2 = \frac{\sum_{i=1}^n (\Delta e Y)^2}{\sum_1^n (\Delta y_i)^2} = a_1^2 \frac{\sum_1^n (\Delta x_i)^2}{\sum_1^n (\Delta y_i)^2}$$

Substituindo a Equação (40) na equação acima temos

$$r^2 = \frac{\left(\sum_{i=1}^n \Delta x_i \Delta y_i \right)^2}{\sum_1^n (\Delta x_i)^2 \sum_1^n (\Delta y_i)^2}$$

logo

$$r = \frac{\left(\sum_{i=1}^n \Delta x_i \Delta y_i \right)}{\sqrt{\sum_1^n (\Delta x_i)^2 \sum_1^n (\Delta y_i)^2}} \quad (41)$$

Obtém por um processo análogo á equação (29)

$$n \sum_1^n \Delta x_i \Delta y_i = n \sum_1^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i, \text{ assim temos}$$

De onde se Obtém

$$n \sum_1^n (\Delta x_i)^2 = n \sum_1^n x_i^2 - \sum_{i=1}^n x_i^2 \quad (42)$$

$$n \sum_1^n (\Delta y_i)^2 = n \sum_1^n y_i^2 - \sum_{i=1}^n y_i^2 \quad (43)$$

Substituindo as Equações (36) e (40), na Equação (41) acima tem-se

$$r = \frac{\left(\sum_{i=1}^n \Delta x_i \Delta y_i \right)}{\sqrt{\sum_1^n (\Delta x_i)^2 \sum_1^n (\Delta y_i)^2}} = \frac{n \left(\sum_{i=1}^n x_i y_i \right) - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left[n \left(\sum_{i=1}^n x_i \right) - \sum_{i=1}^n x_i^2 \right] \cdot \left[n \left(\sum_{i=1}^n y_i \right) - \sum_{i=1}^n y_i^2 \right]}}$$