UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS NATURAIS E EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA E MODELAGEM QUANTITATIVA

ESTIMAÇÃO DOS PARÂMETROS DO MODELO DE REGRESSÃO LINEAR SIMPLES PELO MÉTODO NÃO-PARAMÉTRICO

MONOGRAFIA DE ESPECIALIZAÇÃO

Alícia Bolfoni Dias

Santa Maria, RS, Brasil

2005

ESTIMAÇÃO DOS PARÂMETROS DO MODELO DE REGRESSÃO LINEAR SIMPLES PELO MÉTODO NÃO-PARAMÉTRICO

por

Alícia Bolfoni Dias

Monografia apresentada ao Curso de Especialização do Programa de Pós-Graduação em Estatística e Modelagem Quantitativa, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de
Especialista em Estatística e Modelagem Quantitativa.

Orientadora: Prof. MSc. Luciane Flores Jacobi

Santa Maria, RS, Brasil

2005

Universidade Federal de Santa Maria
Centro de Ciências Naturais e Exatas
Programa de Pós-Graduação em Estatística e Modelagem Quantitativa

A Comissão Examinadora, abaixo assinada,
aprova a Monografia de Especialização

ESTIMAÇÃO DOS PARÂMETROS DO MODELO DE REGRESSÃO LINEAR SIMPLES PELO MÉTODO NÃO-PARAMÉTRICO

elaborada por
Alícia Bolfoni Dias

como requisito parcial para obtenção do grau de
Especialista em Estatística e Modelagem Quantitativa

COMISSÃO EXAMINADORA:

Luciane Flores Jacobi, MSc.
(Presidente/Orientadora)

Luis Felipe Dias Lopes, Dr. (UFSM)

Ivanor Müller, Dr. (UFSM)

Santa Maria, 02 de dezembro de 2005.

Dedico com muito carinho este trabalho ao meu irmão Silvano pela ajuda na programação, dedicação, compreensão e principalmente pela sua paciência.

AGRADECIMENTOS

O meu primeiro agradecimento é todo especial aos meus pais que me ampararam com muito amor e carinho, pela paciência, e também pela ajuda em todos os momentos durante esta caminhada.

A minha orientadora, Profª. MSc Luciane Flores Jacobi pelo acompanhamento durante este trabalho, pela oportunidade deste aprendizado, por toda a valiosa contribuição nesta pesquisa, e sua amizade.

A todos os professores do Departamento de Estatística pela sabedoria que souberam transmitir.

Aos meus colegas em especial: Gabriela Bilibio, Gilvete Lírio, Mara Rubia Couto, Mauricio Lutz e Rejane Cardoso pelas horas de estudos, por todas as suas contribuições e amizade.

RESUMO

Monografia de Especialização
Programa de Pós-Graduação em Estatística e Modelagem Quantitativa
Universidade Federal de Santa Maria

ESTIMAÇÃO DOS PARÂMETROS DO MODELO DE REGRESSÃO LINEAR SIMPLES PELO MÉTODO NÃO-PARAMÉTRICO
AUTORA: ALÍCIA BOLFONI DIAS
ORIENTADORA: LUCIANE FLORES JACOBI
Data e local da Defesa: Santa Maria, 02 de dezembro de 2005.

O método mais utilizado para estimar os coeficientes angular e linear da equação da reta no modelo de regressão linear simples é o método de mínimos quadrados, apesar de existirem outros procedimentos. O método não-paramétrico é pouco aplicado em função do seu desenvolvimento ser muito trabalhoso, e neste contexto, o objetivo do trabalho consiste em implementar um algoritmo através do *Visual Basic for Applications* (VBA), o qual pode facilitar a utilização da técnica não-paramétrica e fazer a comparação das estimativas e critérios encontrados com a paramétrica. A metodologia adotada para o método não-paramétrico é o estimador proposto por Theil, para o coeficiente angular (β) e os dois estimadores propostos por Dietz para o intercepto (α). Conclui-se que a implementação de um algoritmo no VBA facilita a utilização da técnica não-paramétrica no meio acadêmico, pois as estimativas encontradas pelos dois métodos são muito próximas, não havendo grandes diferenças entre os critérios utilizados para comparação das equações estimadas.

Palavras-chave: Regressão Linear Simples, Método Não-Paramétrico, Implementar algoritmo.

ABSTRACT

Specialization Monograph
Post Gratuated Program in Statistics and Quantitative Modelling
Federal University of Santa Maria

CONSTRUCTION OF SIMPLE LINEAR MODEL REGRESSION BY THE NONPARAMETRIC METHOD
AUTHOR: ALÍCIA BOLFONI DIAS
ADVISOR: LUCIANE FLORES JACOBI
Date and place of defense: Santa Maria, Dezember 2, 2005.

The most used method to estimate the angular and linear coefficients of the line equation in simple linear regression model is the least square method, in spite do the existence of another. The nonparametric method, which is less used because its laborious process. In this context, the goal of this work is implement an algorithm, using the *Visual Basic for Applications* (VBA) which can became easier to use the nonparametric method to compare the estimates and criterias found by parametric method. The methodology used was the estimator of Theil's estimator angular coefficient (β) and the two estimators proposed by Dietz for the intercept (α). At the end of research the algorithm implemented in VBA really makes easier the use of the nonparametric method in academic environment, because the estimates found by the two methods were very close, with small differences among the criterias used to compare the estimated equations.

Key words: Simple Linear Regression, Nonparametric Method, Implement algorithm.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SIGLAS

AIC – *Akaike Information Criteria*

BIC – *Bayesian Information Criteria*

EQM – Erro Quadrado Médio

MMQ – Métodos dos Mínimos Quadrados

SQR – Soma dos Quadrados dos Resíduos

VBA – *Visual Basic for Applications*

SUMÁRIO

1 INTRODUÇÃO

Nos dias de hoje, geralmente, os pesquisadores necessitam tomar decisões baseadas na relação entre duas ou mais variáveis que estão inerentemente relacionadas, sendo necessário explorar a natureza dessa ligação.

A técnica estatística que é mais utilizada para pesquisar e modelar o relacionamento existente entre as diversas variáveis de um processo é a análise de regressão que pode ser usada com vários objetivos como: descrição, predição, controle e estimação.

O estudo da dependência de uma variável em relação a uma única variável explicativa chama-se análise de regressão de duas variáveis ou análise de regressão simples (Gujarati, 2000).

A análise de regressão linear simples significa encontrar a equação da reta que melhor se ajuste aos dados. Este melhor ajuste é a tentativa de achar a equação da reta para a qual a soma dos quadrados das diferenças entre os valores reais (Y_i) e os valores que seriam previstos na reta de regressão ajustada ($\hat{Y}_i$) seja a menor possível. A sua utilização amplia-se a cada dia, devido ao fato da análise de regressão ser baseada na idéia relativamente simples de se empregar uma equação para expressar o relacionamento entre as variáveis de interesse (Levine, Berenson, Stephan, 2000; Werkema, Aguiar, 1996).

Quando um modelo de regressão linear simples é usado em estudos, é necessário que os quatros principais pressupostos existentes sejam válidos. Se alguns desses não se confirmarem, o modelo poderá ser inadequado para fazer as inferências de interesse.

Para solucionar esse problema a maioria dos autores indica a utilização de transformação da variável resposta a qual depende da pressuposição que não é válida.

Nesse estudo é apresentada uma nova maneira de solucionar este problema, utiliza-se, aqui, uma técnica pouco aplicada a não-paramétrica para se estimar os coeficientes do modelo de regressão linear simples.

Os modelos não-paramétricos possuem a vantagem de serem aplicados a uma grande variedade de situações, porque não possuem as exigências rígidas dos métodos paramétricos.

Dessa forma, pretende-se contribuir para que a relação entre duas variáveis seja melhor avaliada, mesmo quando o método dos mínimos quadrados que é o mais usado, pois fornece os parâmetros que melhor se ajustam aos dados, não é o mais indicado.

1.1 Objetivos

1.1.1 Objetivo Geral

Este trabalho tem como objetivo geral estimar os coeficientes do modelo de regressão linear simples pelo método não-paramétrico.

1.1.2 Objetivos Específicos

- Realizar um estudo teórico sobre a regressão linear pelo método não-paramétrico;
- Determinar as estimativas dos coeficientes angular e linear pelo método não-paramétrico;
- Comparar as estimativas dos coeficientes do modelo de regressão linear simples pelos métodos paramétrico e não-paramétrico;
- Implementar um algoritmo por meio do *Visual Basic for Applications* (VBA).

1.2 Justificativa

O método dos mínimos quadrados (MMQ) fornece os parâmetros que melhor ajustam os dados à reta de regressão, ou seja, que minimiza a soma dos quadrados dos erros (Jacobi, 2001).

A estimação dos parâmetros do modelo de regressão pelo método não-paramétrico é útil quando não se pode estimar os coeficientes da equação de regressão pelo método dos mínimos quadrados. Quando o MMQ não pode ser estabelecido, usualmente transforma-se a variável resposta, para que desta forma as pressuposições sejam atendidas, mas muitas vezes essa técnica não resolve o problema, devido a isso a regressão não-paramétrica será uma alternativa a mais nessas situações.

Além disso, como o método não-paramétrico é pouco aplicado, pois, para utilizá-lo, é exigido muito trabalho e a não existência de um software apropriado para se estimar os coeficientes angular e linear da equação da reta, implementou-se um algoritmo por meio do VBA para facilitar a utilização desta técnica.

1.3 Metodologia

Este trabalho propõe encontrar as estimativas pelo método não-paramétrico considerando a proposta feita por Theil (1950 apud Daniel, 1999, p. 717) para calcular o coeficiente angular (β) e ou; a proposta feita por Dietz (1989 apud Daniel, 1999. p.718) para os dois estimadores intercepto (α).

Após o estudo das teorias aplicadas para exemplificar a metodologia, utilizaram-se exemplos hipotéticos, sendo que para cada exemplo se obteve três equações: uma para o método paramétrico e duas para o não-paramétrico, comparando-as através dos critérios *Akaike Information Criteria* (AIC) e *Bayesian Information Criteria* (BIC) e do cálculo do Erro Quadrado Médio (EQM).

Os dados foram rodados utilizando-se a macro implementada no VBA, linguagem de programação usada no programa Microsoft Excel, e após, os resultados foram analisados.

1.4 Estrutura do trabalho

Para atender aos objetivos propostos, a estrutura deste trabalho será da seguinte forma: no Capítulo 1 tem-se os objetivos, geral e específicos, a justificativa e a metodologia; no Capítulo 2, apresentação da revisão de literatura sobre a Análise de regressão e seus pressupostos e os Critérios para seleção de modelos; no Capítulo 3, a fundamentação teórica onde se aborda o método não-paramétrico e a linguagem de programação VBA; no Capítulo 4, exemplos e discussões; no Capítulo 5, conclusão.

2 REVISÃO DE LITERATURA

Este capítulo traz a revisão de literatura dividida em três itens. No item 2.1 realizou-se o estudo da Análise de Regressão. O item 2.2 aborda os pressupostos à análise de regressão linear simples, e o item 2.3 são os critérios para a seleção de modelos. Estes três itens servirão como suporte para o desenvolvimento do trabalho e também onde se pretende mostrar como estas técnicas serão utilizadas.

2.1 Análise de Regressão

A origem histórica do termo regressão é devida a Francis Galton. Em um famoso ensaio, ele verificou que, mesmo havendo uma tendência de pais altos terem filhos altos e de pais baixos terem filhos baixos, a altura média dos filhos de pais de uma dada altura tendia a se deslocar ou "regredir" até a altura média da população como um todo. De outra maneira pode-se dizer que a altura dos filhos de pais extraordinariamente altos ou baixos tende a se mover para a altura média da população. A lei de regressão universal de Galton foi confirmada por seu amigo Karl Pearson que coletou mais de mil registros das alturas dos membros de grupos de família. Ele verificou que a altura média dos filhos de um grupo de pais altos era inferior à altura de seus pais, e a altura média dos filhos de um grupo de pais baixos era superior à altura de seus pais. Assim, tanto os filhos altos como baixos "regrediram" em direção à altura média de todos os homens (Gujarati, 2000).

A moderna interpretação da análise de regressão é, porém, bem diferente. É utilizada principalmente na elaboração de modelos de previsão. A análise de regressão é o desenvolvimento de um modelo estatístico que possa ser utilizado para prever valores de uma variável dependente ou variável resposta, com base nos valores de pelo menos uma variável independente ou explicativa, com o objetivo de estimar e/ou prever a média (da população) ou o valor médio dependente em termos dos valores conhecidos ou fixos (em amostragem repetida) das explicativas (Gujarati, 2000; Levine, Berenson, Stephan, 2000).

Segundo Toledo e Ovalle (1985), a análise de regressão linear simples tem por objetivo descrever, através de um modelo matemático, a relação existente entre duas variáveis: a variável explicativa X e variável explicada Y, a partir de uma amostra de n pares de valores (x_i, y_i), i = 1, 2, 3, ..., n, considerando apenas a variável Y como aleatória e a variável X como, supostamente, sem erro.

De maneira geral, a análise de regressão pode ser utilizada para diferentes objetivos:

a) Descrição: É utilizada uma equação para sumarizar ou descrever um conjunto de dados, onde a análise de regressão pode ser empregada para ajustar uma equação.

b) Predição: O interesse está em predizer os valores da variável resposta.

c) Controle: É muito utilizado na regressão com o objetivo de controlar a variável de interesse em faixas de valores pré-fixados. Destaca-se, quando a equação de regressão for empregada, a relação existente entre a variável de interesse e as variáveis utilizadas, para o seu controle, sejam do tipo causa-e-efeito.

d) Estimação: A análise de regressão é utilizada para estimar parâmetros desconhecidos de equações teóricas que representam o relacionamento entre as variáveis de interesse (Werkema, Aguiar, 1996).

As regressões classificam-se em: lineares e não lineares, sendo que as lineares podem ser simples ou múltiplas.

A regressão linear simples é a técnica estatística mais utilizada para se pesquisar o relacionamento entre as variáveis de um processo, possuindo o seguinte modelo matemático:

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \quad (i = 1,2,3,...,n) \tag{1}$$

onde: Y_i é a variável dependente, X_i é a variável independente, ε_i é o erro, α e β são os parâmetros a serem estimados, n indica o tamanho da amostra e o índice i refere-se à unidade de observações dos valores das variáveis. β é o coeficiente angular da reta, também chamado coeficiente de regressão e α é o coeficiente linear da reta ou conhecido como termo constante da equação de regressão.

Ao estabelecer o modelo de regressão linear simples, pressupõe-se que:

- A média dos ε_i é zero e sua variância σ^2 é desconhecida e constante, para $1 \leq i \leq n$ (homocedasticidade);

- Para $i \neq j$, ε_i e ε_j não são correlacionados, isto é, $COV(\varepsilon_i; \varepsilon_j) = 0$, $i \neq j$, $i \leq 1$ e $j \leq n$ (erros independentes);
- A distribuição dos ε_i é normal, $1 \leq i \leq n$;
- A relação entre as variáveis X e Y deve ser linear.

Combinando as pressuposições anteriores, ou seja, que a média dos erros é nula, $E(\varepsilon_i) = 0$, a variância do ε_i é σ^2, $V(\varepsilon_i) = \sigma^2$ e os erros possuem distribuição normal, tem-se que $\varepsilon_i \sim N(0, \sigma^2)$.

Quando se aplica a análise de regressão ao estado da relação funcional entre duas variáveis, geralmente, encontram-se os seguintes problemas:

Especificação do modelo

A especificação do modelo envolve duas etapas:

- Seleção de variáveis

Na regressão linear, o interesse é prever o valor de uma variável aleatória Y. Uma informação inicial sobre os possíveis valores dessa variável é fornecida por sua média estimada com base em uma amostra de seus valores. O segundo passo é verificar se essa variável aleatória se correlaciona com alguma variável que se pode controlar.

Para que o estudo tenha consistência, será preciso construir uma teoria que informe qual a origem dessa correlação e por que isso acontece (Silva, Silva,1999).

- Especificação da forma funcional

Selecionada a variável, o seguinte passo é construir o diagrama de dispersão de Y contra a variável explicativa X. A observação do diagrama sugerirá a forma funcional, ou as possíveis formas funcionais, lineares ou linearizáveis, que aparentemente ajuste os pontos do diagrama (Silva, Silva,1999).

Estimação dos parâmetros

Após a determinação do modelo, deve-se determinar o valor dos diversos parâmetros.

Se o modelo escolhido for uma forma linear $Y = \alpha + \beta X + \varepsilon_i$, será necessário estimar os parâmetros α e β. Nesse caso, designar-se-á por a e b os estimadores de α e β, respectivamente. A partir da observação de uma amostra de n pares de valores (x_i, y_i); i = 1,2,...,n; obter-se-á as estimativas a e b e, dessa forma será encontrada uma estimativa do modelo adotado, compondo as estimativas a e b através da fórmula: $\hat{Y} = a + bX$ (Fonseca, Martins, Toledo, 1985).

Adaptação e significância do modelo adotado

Nesta fase se verifica se a especificação escolhida na primeira etapa, se adapta conveniente aos dados observados, através do cálculo da diferença entre os valores de Y observados e os valores de $\hat{Y}$ estimados pela equação de regressão, obtendo-se assim os erros de estimação (Jacobi, 2001).

2.1.1 Estimativas dos Coeficientes Linear e Angular pelo Método dos Mínimos Quadrados (MMQ)

A idéia do MMQ é construir uma estimativa para a verdadeira reta de regressão, tornando a soma dos quadrados das diferenças $(y_i - \hat{y}_i)$ a menor possível entre os pontos observados y_i e o correspondente valor sobre a reta estimada $\hat{Y}$ (Silva, Silva,1999).

A reta estimada tem a forma: $\hat{Y} = a + bX$

onde: $\hat{Y}$ é o estimador de $\overline{Y}$

a e b são estimadores de α e β.

A Figura 1, a seguir, ilustra como se calcula as estimativas, apresentando os pontos amostrados e a reta de regressão estimada.

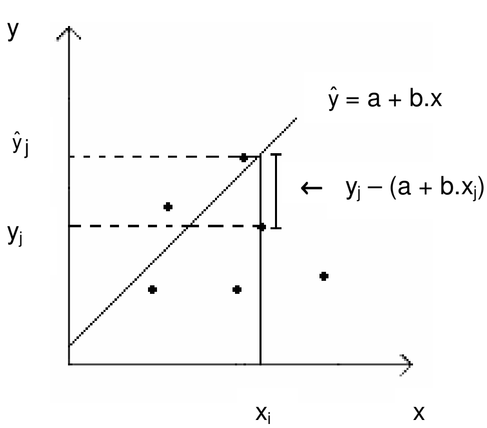


Figura 1 – Pontos amostrados e a reta de regressão estimada.

Assim, a reta ajustada é denominada reta de mínimos quadrados, pois os valores de a e b são obtidos de tal forma que é mínima a soma dos quadrados das diferenças entre os valores observados de Y e os obtidos a partir da reta ajustada para os mesmos valores de X (Toledo e Ovalle, 1985). Temos: $\Sigma\hat{e} = \Sigma(Y - \hat{Y})^2 = \Sigma(Y - a - bX)^2$ seja mínima, onde: $\hat{e} = Y - \hat{Y}$.

Para obter os estimadores a e b aplica-se a condição necessária de mínimo à função $\Sigma(Y - \hat{Y})^2$. Basta derivá-la com relação a esses parâmetros e igualar as derivadas a zero:

$$a)\frac{\partial\Sigma\hat{e}^2}{\partial a} = 0 \qquad e \qquad b)\frac{\partial\Sigma\hat{e}^2}{\partial b} = 0$$

$$a)\frac{\partial}{\partial a}\Sigma\hat{e}^2 = \frac{\partial}{\partial a}\Sigma(Y - \hat{Y})^2 = \frac{\partial}{\partial a}\Sigma(Y - a - bX)^2 = -2\Sigma(Y - a - bX) = 0 \qquad (I)$$

$$b)\frac{\partial}{\partial b}\Sigma\hat{e}^2 = \frac{\partial}{\partial b}\Sigma(Y - \hat{Y})^2 = \frac{\partial}{\partial b}\Sigma(Y - a - bX)^2 = -2\Sigma(Y - a - bX)X = 0 \qquad (II)$$

Retornando para as expressões I e II, e sabendo que n é o número de observações, tem-se:

$$-2\Sigma(Y - a - bX) = 0 \Rightarrow \Sigma(Y - a - bX) = \Sigma Y - na - b\Sigma X = 0 \qquad (III)$$

$$-2\Sigma(Y - a - bX)X = 0 \Rightarrow \Sigma(Y - a - bX)X = \Sigma XY - a\Sigma X - b\Sigma X^2 = 0 \qquad (IV)$$

As equações III e IV podem ser escritas de outra forma:

$$\left.\begin{array}{l} De(III) \Rightarrow \Sigma Y = na + b\Sigma X \\ De(IV) \Rightarrow \Sigma XY = a\Sigma X + b\Sigma X^2 \end{array}\right\} \text{equações normais do ajustamento}$$

Tem-se então um sistema de duas equações com duas incógnitas (a e b), basta resolvê-lo a partir da primeira equação para se obter os valores dos parâmetros.

$$\Sigma Y = na + b\Sigma X \Rightarrow \frac{\Sigma Y}{n} = a + b\frac{\Sigma X}{n} \qquad e \qquad a = \frac{\Sigma Y}{n} - b\frac{\Sigma X}{n} = \overline{Y} - b\overline{X}$$

$$a = \overline{Y} - b\overline{X} \qquad (2)$$

Substituindo *a* pela expressão (2) na segunda equação do sistema tem-se:

$$\Sigma XY = \left(\frac{\Sigma Y}{n} - b\frac{\Sigma X}{n}\right)\Sigma X + b\Sigma X^2 = \frac{(\Sigma Y)(\Sigma X)}{n} - \frac{b(\Sigma X)^2}{n} + b\Sigma X^2 \quad \Rightarrow$$

$$\Sigma XY = \frac{(\Sigma Y)(\Sigma X)}{n} + b\left[\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right] \quad \Rightarrow \quad b\left[\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right] = \Sigma XY - \frac{(\Sigma Y)(\Sigma X)}{n}$$

$$b = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}} \quad \text{ou} \quad b = \frac{S_{XY}}{S_{XX}} \qquad (3)$$

O coeficiente de determinação é um indicador que fornece elementos para a análise do modelo adotado, indicando quanto por cento a variação explicada pela regressão representa da variação total, definido por (Fonseca, Martins, Toledo, 1985):

$$R^2 = \frac{VE}{VT} = \frac{b^2 S_{XX}}{S_{YY}} = \frac{bS_{XY}}{S_{YY}}.100 \qquad (4)$$

variando no intervalo de $0 \leq R^2 \leq 1$.

As Figuras 2, 3 e 4, a seguir, mostram a configuração do diagrama de dispersão, nos casos de $0 < R^2 < 1$, $R^2 = 1$ e $R^2 = 0$, respectivamente:

No caso de $0 < R^2 < 1$, Figura 2, os pontos encontram-se tão próximos à linha de regressão, quanto mais próximos de 1 for o valor do coeficiente de determinação. As variações de Y são mais explicadas pelas variações de X através da função especificada, quanto maior for o valor de R^2.

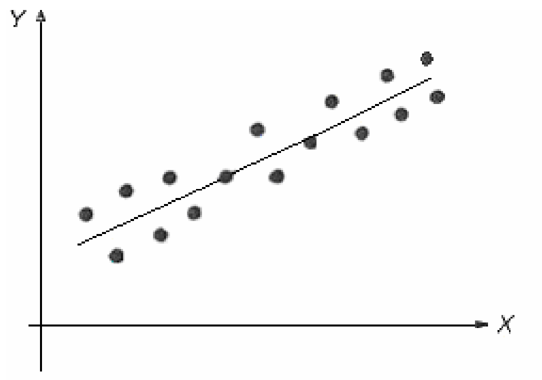


Figura 2 – Diagrama de dispersão e reta estimada para duas variáveis com coeficiente de determinação $0 < R^2 < 1$.

Fonte: Adaptada, Toledo e Ovalle (1985)

No caso em que $R^2 = 1$, Figura 3, todos os pontos observados se situam "exatamente" sobre a reta de regressão, é um ajuste perfeito. As variações de Y são 100% explicadas pelas variações de X através da função especificada, não havendo desvios em torno da função estimada (Fonseca, Martins, Toledo, 1985).

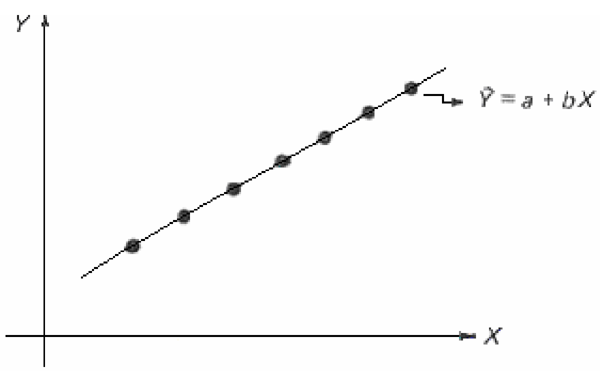


Figura 3 – Diagrama de dispersão e reta estimada para duas variáveis com coeficiente de determinação $R^2 = 1$.

Fonte: Adaptada, Toledo e Ovalle (1985)

Por outro lado, se $R^2 = 0$, Figura 4, conclui-se que as variações de Y são exclusivamente aleatórias, e a introdução da variável X no modelo não incorporará informação alguma sobre as variações de Y (Fonseca, Martins, Toledo,1985).

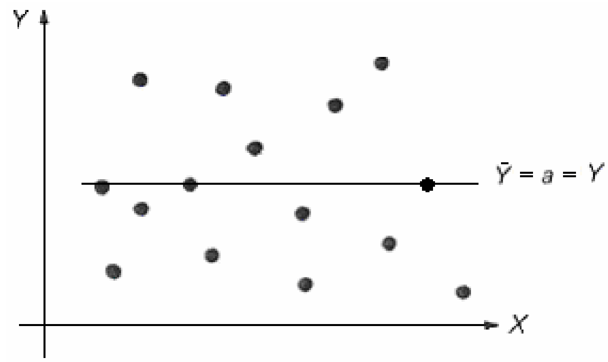


Figura 4 – Diagrama de dispersão e reta estimada para duas variáveis com coeficiente de determinação $R^2 = 0$.

Fonte: Adaptada, Toledo e Ovalle (1985)

2.2 Pressupostos à Análise de Regressão Linear Simples

Para se estabelecer o modelo de regressão linear simples é necessário a observação de alguns pressupostos, sendo quatro principais: normalidade, homocedasticidade, independência de erros e linearidade.

Normalidade: requer que os valores de Y sejam normalmente distribuídos para cada valor de X. Enquanto a distribuição dos valores de y_i em torno de cada nível de X não for exatamente diferente de uma distribuição normal, inferências sobre a linha de regressão e sobre coeficientes de regressão não são seriamente afetadas (Levine, Berenson, Stephan, 2000).

Para se verificar a normalidade de uma amostra de dados, pode-se utilizar os testes de Lilliefors, Shapiro-Wilks entre outros.

Homocedasticidade: requer que as variações em torno da linha de regressão sejam constantes para todos os valores de X. Isto significa que Y varia na mesma proporção, quando X for um valor baixo e quando X for um valor elevado. O pressuposto da homocedasticidade é importante na utilização do método dos mínimos quadrados, para determinar os coeficientes de regressão. Se houver sérios afastamentos desse pressuposto, podem-se aplicar transformações de dados ou método dos mínimos quadrados ponderados (Levine, Berenson, Stephan, 2000).

A presença de heterocedasticidade gera: estimadores dos parâmetros (β) não viesados, mas ineficientes; e as variâncias estimadas dos parâmetros são viesadas, gerando problemas com os intervalos de confiança e os testes de hipóteses (Vasconcellos e Alves, 2000; Hill, Griffiths, Judge, 1999).

A heterocedasticidade pode surgir como resultado de presença de observações aberrantes ("outliers"). Uma observação é aberrante quando é muito diferente (ou muito pequena ou muito grande) das outras observações na amostra. A inclusão ou a exclusão desta observação, principalmente se a amostra for pequena, pode alterar substancialmente os resultados da análise de regressão (Gujarati, 2000).

Para detectar a existência de heterocedasticidade, consiste em estimar o modelo, utilizando mínimos quadrados e fazer o gráfico dos resíduos de mínimos quadrados.

Um dos testes para detectar a presença da heterocedasticidade é o Teste de Goldfeld-Quandt.

Independência de erros ou ausência de autocorrelação: requer que o erro (a diferença residual entre valores observados e previstos de Y) deva ser independente para cada valor de X. Esse pressuposto geralmente se refere a dados que são coletados ao longo de um período de tempo. Quando os dados são coletados dessa maneira, os resíduos para um determinado período de tempo são freqüentemente correlacionados com os do período de tempo anterior (Levine, Berenson, Stephan, 2000).

Este pressuposto pode ser avaliado se desenharmos o gráfico com os resíduos na ordem ou seqüência em que os dados observados foram obtidos. Os dados coletados ao longo de períodos de tempo geralmente exibem um efeito de autocorrelação entre as observações sucessivas, isto é, existe correlação entre uma determinada observação e os valores que a antecedem ou que lhe sucedem. Esse efeito pode ser mensurado pela estatística de Durbin-Watson.

As conseqüências da autocorrelação para o estimador de mínimos quadrados são essencialmente as mesmas da heterocedasticidade:

- O estimador de mínimos quadrados ainda é um estimador linear e não-tendencioso, mas não é mais o *melhor* (Hill, Griffiths, Judge, 1999);

- As fórmulas dos desvios-padrão calculados na forma usual para o estimador de mínimos quadrados já não são corretas e, assim, os intervalos de confiança e os testes de hipóteses baseados nesses desvios-padrão podem ser enganosos (Hill, Griffiths, Judge, 1999).

Linearidade: estabelece que a relação entre as variáveis deve ser linear. Essa relação pode ser verificada por meio do coeficiente de correlação linear simples, que tem por objetivo medir e avaliar o grau de relação existente entre duas variáveis aleatórias.

A correlação linear procura medir a relação entre as variáveis X e Y. Esta relação pode ser observada, por meio da disposição dos pontos de X e Y, plotados em um gráfico chamado de diagrama de dispersão, que possibilita avaliar possíveis relações de causa e efeito entre duas variáveis, conforme mostrado na Figura 5 (Toledo e Ovalle, 1985).

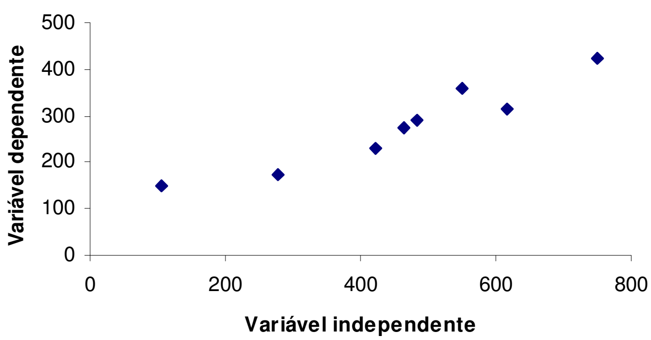


Figura 5 - Representação genérica do gráfico de dispersão.

O instrumento de medida de correlação linear é dado pelo coeficiente de correlação de Pearson:

$$r_{XY} = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{\sqrt{\left[\sum X^2 - \frac{(\sum X)^2}{n}\right]}\sqrt{\left[\sum Y^2 - \frac{(\sum Y)^2}{n}\right]}} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}} \tag{5}$$

onde: n = número de observações;

X e Y são as variáveis em estudo;

campo de variação: $-1 \leq r_{xy} \leq 1$.

A correlação pode ser: Correlação Linear Positiva $0 < r_{xy} < 1$, Correlação Linear Perfeita Positiva $r_{xy} = 1$, Correlação Negativa $-1 < r_{xy} < 0$, Correlação Perfeita Negativa $r_{xy} = -1$, Correlação Nula $r_{xy} = 0$ (Toledo e Ovalle, 1985).

2.3 Critérios para Seleção de Modelos

2.3.1 Critérios "*Akaike Information Criteria*" (AIC) e "*Bayesian Information Criteria*" (BIC)

As formas de comparação dos modelos comumente utilizadas são a análise dos resíduos e a avaliação da ordem do modelo cujos critérios mais usados são os AIC e BIC, que levam em conta a variância do erro, o tamanho da amostra e os valores dos coeficientes estimados. Segundo Farias, Rocha e Lima (2000):

$$AIC = n\ ln(\text{somatório dos quadrados dos resíduos}) + 2T \quad (6)$$

$$BIC = n\ ln(\text{somatório dos quadrados dos resíduos}) + T\ ln(n) \quad (7)$$

Onde: n = número de observações utilizadas;

T = número de parâmetros estimados.

2.3.2 Cálculo do Erro Quadrado Médio (EQM)

Uma outra maneira de se comparar modelos é determinar os desvios em relação aos valores observados da variável Y, ou seja, calcula-se a diferença entre o valor observado para Y e sua respectiva estimativa $\hat{Y}$ determinada pelas equações ajustadas. Eleva-se esses desvios ao quadrado e divide-se a soma dos quadrados pelo número de valores amostrados, encontra-se, desta forma, o erro quadrado médio determinado por:

$$EQM = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{8}$$

2.4 Comentários do Capítulo

Neste capítulo, desenvolveu-se a técnica da análise de regressão, sua validade e suas pressuposições. O MMQ é usado, pois fornece os parâmetros que melhor se ajustam aos dados, mas quando fracassam uma ou mais pressuposições da análise de regressão, este método não pode ser mais usado. Assim no Capítulo seguinte desenvolveu-se outro método, o não-paramétrico para encontrar as estimativas do modelo como alternativa ao MMQ, bem como a macro no VBA para fazer estes cálculos automaticamente.

3 FUNDAMENTAÇÃO TEÓRICA

A metodologia proposta neste capítulo será o método não-paramétrico, descrito no item 3.1, que se utiliza de medianas para encontrar as estimativas dos coeficientes do modelo de regressão linear simples. E no item 3.2 será apresentado a macro do VBA que foi implementada para facilitar a utilização da técnica não-paramétrica.

3.1 Método Não-Paramétrico

Alguns procedimentos não-paramétricos são úteis na análise de regressão linear simples, quando não se é capaz de fazer as suposições necessárias para a aplicação válida das técnicas paramétricas análogas (Daniel, 1978).

A principal vantagem do método não-paramétrico é que não são exigidas as suposições sobre a população da qual se originam os dados, e uma de suas desvantagens segundo Sen (1968) é que pode ser bastante laborioso quando n não for muito pequeno para se estimar os coeficientes da equação de regressão.

Para encontrar as estimativas dos coeficientes através dos cálculos das medianas, no método não-paramétrico foi considerado o estimador proposto por Theil (1950 apud Daniel, 1999, p. 717), para o coeficiente angular (β) e os dois estimadores para o intercepto (α), propostos por Dietz (1989 apud Daniel, 1999. p.718).

3.1.1 O Estimador para o Coeficiente Angular

Theil (1950 apud Daniel, 1999, p. 717) recomenda um método para obter um ponto estimado de β, usando o modelo clássico $Y_i = \alpha + \beta X_i + \varepsilon_i$, $i = 1,...,n$. Onde X_i são constantes conhecidas, α e β são parâmetros desconhecidos, Y_i um valor observado da variável contínua Y em X_i, e ε_i são mutuamente independentes. Os dados consistem em n pares de observações amostrais $(x_{1,}\, y_1),..., (x_n,y_n)$.

Para se obter o estimador β, primeiramente forma-se todas as possíveis inclinações entre dois pontos da amostra, conforme mostra a Figura 6:

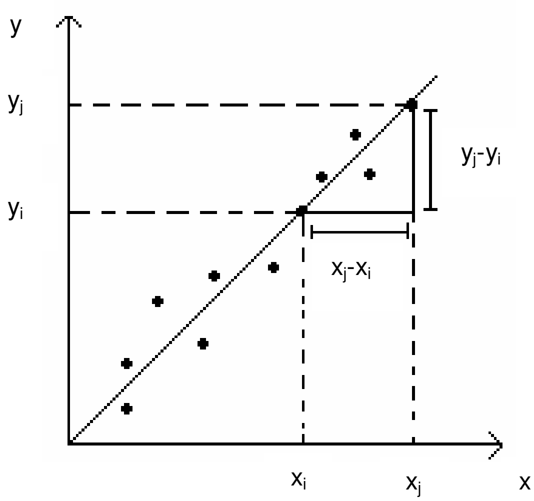


Figura 6 – Interpretação geométrica dos valores de S_{ij}.

Onde:

$$S_{ij} = (y_j - y_i)/(x_j - x_i) \qquad (9)$$

para $i < j$ encontra-se $N = {}_nC_2$ valores para S_{ij} que é o coeficiente angular da reta que passa por dois pontos. Assim, o cálculo de S_{ij} $\forall$ i, j, representa todas as combinações possíveis de coeficientes angulares de retas tomadas, duas a duas, em todos os pontos do diagrama de dispersão. O estimador β que se designa através de $\hat{\beta}$, é a mediana dos valores de S_{ij}. Assim $\hat{\beta}$ = mediana $\{S_{ij}\}$ (Dietz, 1989).

Deste modo para n pares de valores tem-se:

$S_{12} = (y_2 - y_1)/(x_2 - x_1)$

$S_{13} = (y_3 - y_1)/(x_3 - x_1)$

.

.

.

$S_{1n} = (y_n - y_1)/(x_n - x_1)$
$S_{23} = (y_3 - y_2)/(x_3 - x_2)$
$S_{24} = (y_4 - y_2)/(x_4 - x_2)$
.
.
.
$S_{2n} = (y_n - y_2)/(x_n - x_2)$
.
.
.
$S_{n-1,\, n} = (y_n - y_{n-1})/(x_n - x_{n-1})$

Conforme Sen (1968), os S_{ij} são as inclinações das retas ligando cada par de pontos (x_i, y_i) e (x_j, y_j) onde $x_i \neq x_j$; os pares de pontos para os quais x_i é igual a x_j não são considerados.

3.1.2 Os Estimadores para o Coeficiente Linear

Dietz (1989 apud Daniel, 1999. p.718) recomenda um método para obter dois estimadores para α: $\hat{\alpha}_1$ e $\hat{\alpha}_2$.

Para determinar $\hat{\alpha}_1$, encontra-se a partir dos n pares de valores observados, e da equação (10) todas as possíveis estimativas para α, assim tem-se: $Y_1 - \hat{\beta} X_1$, $Y_2 - \hat{\beta} X_2$,...

$$Y_i - \hat{\beta} X_i \tag{10}$$

Após determina-se a mediana dos possíveis valores estimados para α, sendo a mesma a estimativa de $\hat{\alpha}_1$, quando se assume que os erros não são simétricos.

Para determinar $\hat{\alpha}_2$, primeiro determina-se a média para os valores da variável dependente e dos valores da variável independente consideradas em S_{ij}.

Assim: $\overline{Y}_{12} = \dfrac{y_2 + y_1}{2}$ e $\overline{X}_{12} = \dfrac{x_2 + x_1}{2}$

$$\overline{Y}_{13} = \frac{y_2 + y_3}{2} \quad \text{e} \quad \overline{X}_{13} = \frac{x_2 + x_3}{2}$$

$$\vdots$$

$$\overline{Y}_{1n} = \frac{y_n + y_1}{2} \quad \text{e} \quad \overline{X}_{1n} = \frac{x_n + x_1}{2}$$

$$\overline{Y}_{23} = \frac{y_3 + y_2}{2} \quad \text{e} \quad \overline{X}_{23} = \frac{x_3 + x_2}{2}$$

$$\overline{Y}_{24} = \frac{y_4 + y_2}{2} \quad \text{e} \quad \overline{X}_{24} = \frac{x_4 + x_2}{2}$$

$$\vdots$$

$$\overline{Y}_{2n} = \frac{y_n + y_2}{2} \quad \text{e} \quad \overline{X}_{2n} = \frac{x_n + x_2}{2}$$

$$\vdots$$

$$\overline{Y}_{n-1,n} = \frac{y_n + y_{n-1}}{2} \quad \text{e} \quad \overline{X}_{n-1,n} = \frac{x_n + x_{n-1}}{2}$$

Após aplica-se na equação (10) os $\overline{Y}_s$ e $\overline{X}_s$ encontrados. Junta-se a esses, os valores encontrados para $\hat{\alpha}_1$, determinando-se desta forma, $\frac{n(n+1)}{2}$ possíveis estimativas para $\hat{\alpha}_2$ sendo que a mediana desses valores será a estimativa do intercepto, quando se considera que os erros são simétricos.

3.2 *Visual Basic for Applications* (VBA)

O VBA é a linguagem de programação utilizada no programa Microsoft Excel, do tipo planilha eletrônica, onde foi implementado um algoritmo para se estimar os coeficientes angular e linear, calcular os critérios, e a construção do gráfico com o intuito de facilitar a utilização desta técnica não-paramétrica.

A Figura 7, a seguir, mostra a tela inicial onde foi realizada a programação no VBA.

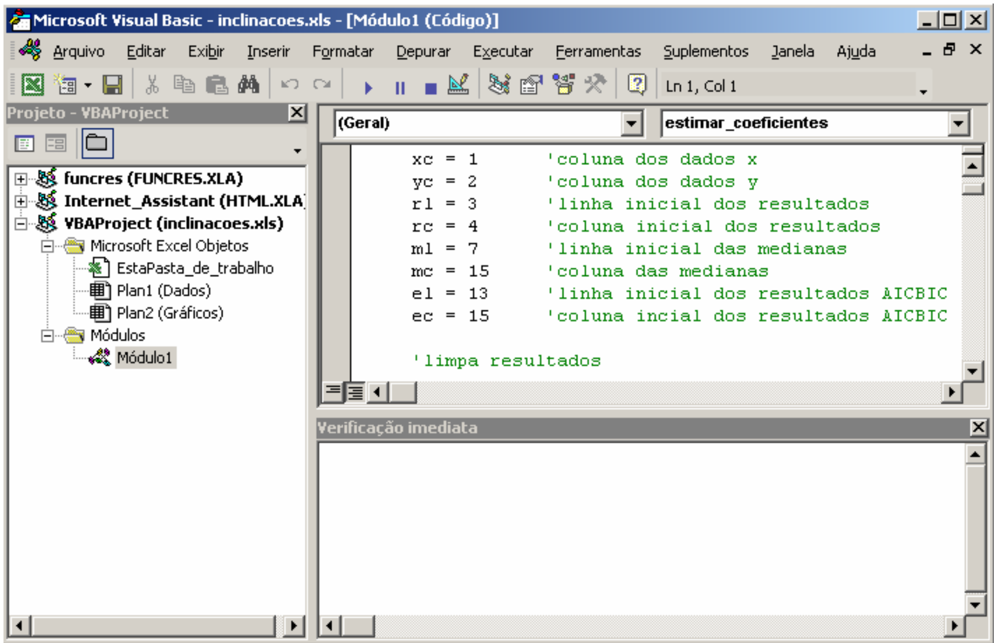


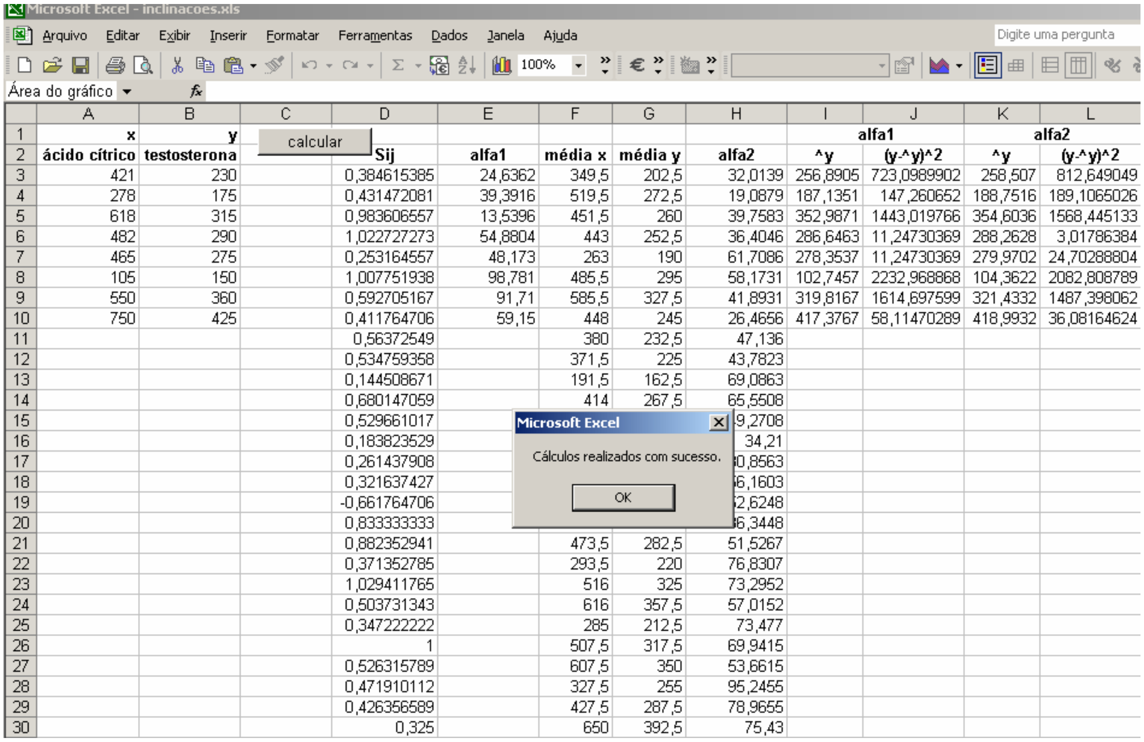
Figura 7 – Tela inicial do VBA.

Com o Microsoft Excel instalado no computador é só utilizar a planilha que contém a macro e digitar os valores para as variáveis independente e dependente nos lugares indicados e clicar no botão calcular, conforme mostra a Figura 8.

O programa realiza todos os cálculos e fornece os resultados na própria planilha, como mostra a Figura 9.

Para mostrar como foram realizados os cálculos por meio da planilha do Microsoft Excel, utilizou-se duas planilhas: uma para saída dos dados, com nome Dados e outra para o gráfico, nome Gráficos.

Na Figura 8 a seguir, visualiza-se a utilização da macro implementada através de dados do exemplo retirado de Daniel (1999; p.717).

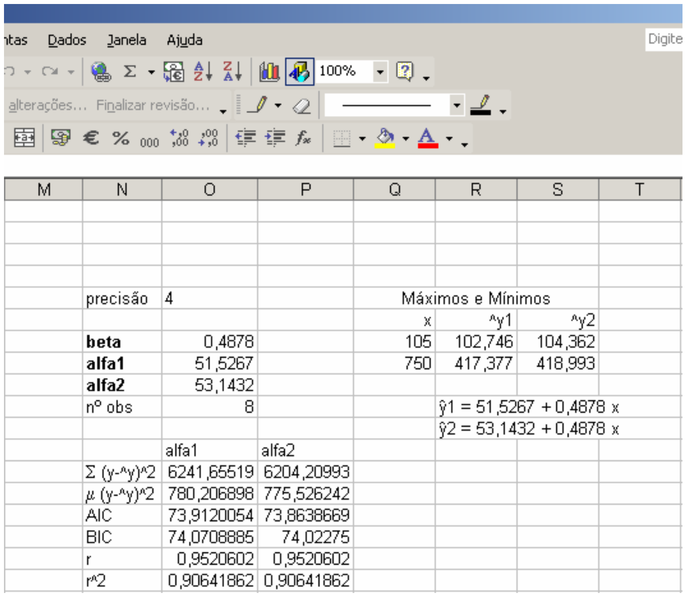


	A	B	C	D	E	F	G	H	I	J	K	L
1	x	y							alfa1		alfa2	
2	ácido cítrico	testosterona		Sij	alfa1	média x	média y	alfa2	^y	(y-^y)^2	^y	(y-^y)^2
3	421	230		0,384615385	24,6362	349,5	202,5	32,0139	256,8905	723,0989902	258,507	812,649049
4	278	175		0,431472081	39,3916	519,5	272,5	19,0879	187,1351	147,260652	188,7516	189,1065026
5	618	315		0,983606557	13,5396	451,5	260	39,7583	352,9871	1443,019766	354,6036	1568,445133
6	482	290		1,022727273	54,8804	443	252,5	36,4046	286,6463	11,24730369	288,2628	3,01786384
7	465	275		0,253164557	48,173	263	190	61,7086	278,3537	11,24730369	279,9702	24,70288804
8	105	150		1,007751938	98,781	485,5	295	58,1731	102,7457	2232,968868	104,3622	2082,808789
9	550	360		0,592705167	91,71	585,5	327,5	41,8931	319,8167	1614,697599	321,4332	1487,398062
10	750	425		0,411764706	59,15	448	245	26,4656	417,3767	58,11470289	418,9932	36,08164624
11				0,56372549		380	232,5	47,136				
12				0,534759358		371,5	225	43,7823				
13				0,144508671		191,5	162,5	69,0863				
14				0,680147059		414	267,5	65,5508				
15				0,529661017		[illegible]	[illegible]	[illegible]				
16				0,183823529		[illegible]	[illegible]	34,21				
17				0,261437908		[illegible]	[illegible]	[illegible]				
18				0,321637427		[illegible]	[illegible]	[illegible]				
19				-0,661764706		[illegible]	[illegible]	[illegible]				
20				0,833333333		[illegible]	[illegible]	[illegible]				
21				0,882352941		473,5	282,5	51,5267				
22				0,371352785		293,5	220	76,8307				
23				1,029411765		516	325	73,2952				
24				0,503731343		616	357,5	57,0152				
25				0,347222222		285	212,5	73,477				
26				1		507,5	317,5	69,9415				
27				0,526315789		607,5	350	53,6615				
28				0,471910112		327,5	255	95,2455				
29				0,426356589		427,5	287,5	78,9655				
30				0,325		650	392,5	75,43				

Figura 8 – Entrada e análise dos dados.

Clicando-se em calcular, a macro realiza todos os cálculos, determina todos os S_{ij} conforme a equação (9), e também os valores possíveis para o coeficiente linear da equação, conforme visto em 3.1.2.

Ao término de todos os cálculos a seguinte frase aparece: “Cálculos realizados com sucesso”, observado na Figura 8.



M	N	O	P	Q	R	S	T
	precisão	4		Máximos e Mínimos			
				x	^y1	^y2	
	beta	0,4878		105	102,746	104,362	
	alfa1	51,5267		750	417,377	418,993	
	alfa2	53,1432					
	nº obs	8			ŷ1 = 51,5267 + 0,4878 x		
					ŷ2 = 53,1432 + 0,4878 x		
		alfa1	alfa2				
	Σ (y-^y)^2	6241,65519	6204,20993				
	µ (y-^y)^2	780,206898	775,526242				
	AIC	73,9120054	73,8638669				
	BIC	74,0708885	74,02275				
	r	0,9520602	0,9520602				
	r^2	0,90641862	0,90641862				

Figura 9 – Saída dos dados.

Na mesma planilha onde são digitados os dados para as variáveis dependente e independente, sairão os resultados, conforme mostra a Figura 9. Na saída, tem-se os valores para: beta (β), alfa1 ($\hat{\alpha}_1$), alfa2 ($\hat{\alpha}_2$), $\Sigma(y-\hat{y})^2$ (Soma dos quadrados dos Resíduos - SQR), $\mu(y-\hat{y})^2$ (EQM), os critérios AIC e BIC, r (coeficiente de correlação linear de Pearson entre y e $\hat{y}$), r^2 (coeficiente de determinação), as equações $\hat{y}_1$ e $\hat{y}_2$, nº obs (tamanho da amostra), os máximos e mínimos são os quatros pontos que geram as duas retas no gráfico.

Além disso, pode-se escolher o número de casas decimais desejadas para as estimativas, bastando digitar o número escolhido na precisão.

Para verificar-se melhor os resultados, foi construído um gráfico de dispersão mostrado na Figura 10, a partir dos dados apresentados na Figura 8. A macro atualiza automaticamente este gráfico com os novos resultados após os cálculos.

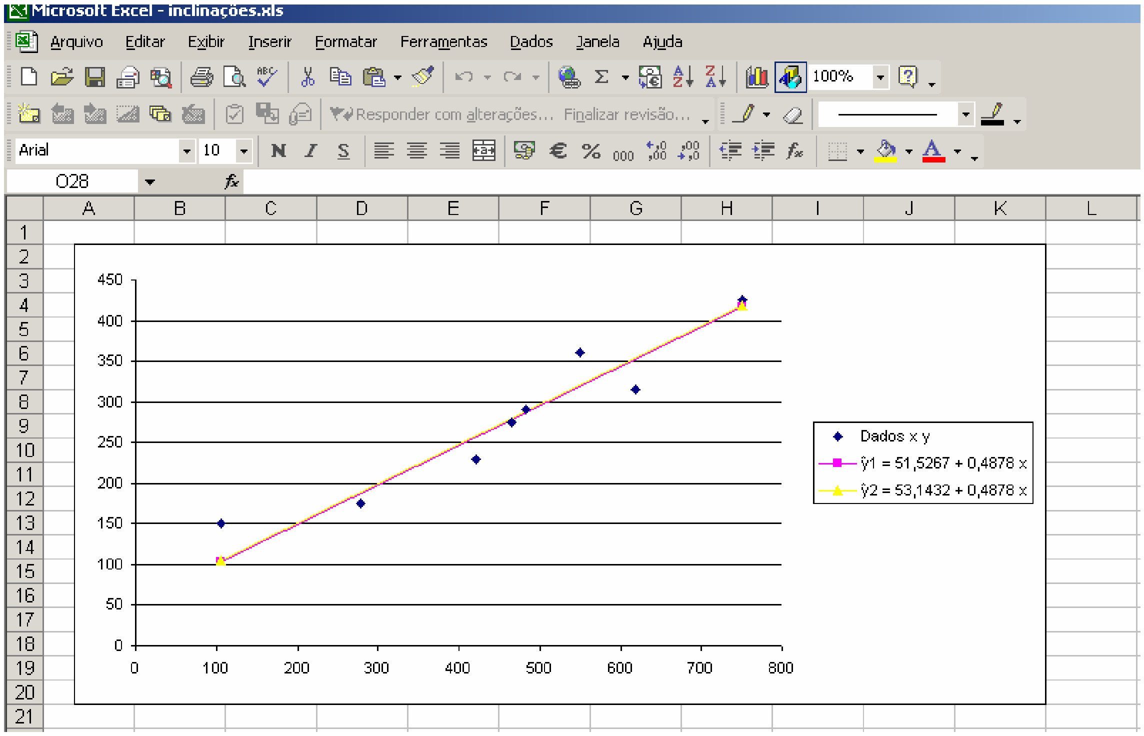


Figura 10 – Representação gráfica dos resultados gerados pela macro.

3.3 Comentários do Capítulo

Neste Capítulo desenvolveu-se como se encontra as estimativas através do método não-paramétrico e apresenta-se a linguagem de programação VBA.

No próximo Capítulo, exemplifica-se esta metodologia através de um exemplo, passo-a-passo, e outros aplicando o VBA.

4 EXEMPLOS E DISCUSSÕES

Neste capítulo são apresentados exemplos hipotéticos para se estimar os coeficientes do modelo de regressão linear simples. Primeiramente, apresenta-se, passo-a-passo, um exemplo com a metodologia aplicada a fim de mostrar que esta é muito trabalhosa e, a seguir, tem-se com a ajuda através do VBA, a resolução de outros exemplos.

Após verifica-se nas tabelas os resultados dos quatros exemplos encontrados através do VBA, onde três foram escolhidos por apresentar alguma violação das pressuposições, e um deles por não apresentar nenhuma falha. Com isso se realiza uma comparação entre os dois métodos não-paramétrico e paramétrico.

O exemplo mostrado, a seguir, foi retirado de Daniel (1999; p.717), com o intuito de mostrar detalhadamente como se utiliza a metodologia do método não-paramétrico, passo-a-passo.

Exemplo 1: (Daniel, 1999; p.717)

Tabela 1 – Testosterona (y) e Ácido Cítrico (x)

testosterona	230	175	315	290	275	150	360	425
ácido cítrico	421	278	618	482	465	105	550	750

Para obter o estimador β pelo método de Theil (1950 apud Daniel, 1999, p. 717), deve-se encontrar todas as possíveis inclinações através da equação (9) para i < j encontra-se $N = {}_8C_2 = 28$ valores para S_{ij}.

$S_{12} = (y_2 - y_1)/(x_2 - x_1)$

$S_{12} = (175 - 230)/(278 - 421)$

$S_{12} = 0,3846$

$S_{13} = (y_3 - y_1)/(x_3 - x_1)$

$S_{13} = (315 - 230)/(618 - 421)$

$S_{13} = 0,4315$

$S_{14} = (y_4 - y_1)/(x_4 - x_1)$
$S_{14} = (290 - 230)/(482 - 421)$
$S_{14} = 0{,}9836$

$S_{15} = (y_5 - y_1)/(x_5 - x_1)$
$S_{15} = (275 - 230)/(465 - 421)$
$S_{15} = 1{,}0227$

$S_{16} = (y_6 - y_1)/(x_6 - x_1)$
$S_{16} = (150 - 230)/(105 - 421)$
$S_{16} = 0{,}2532$

$S_{17} = (y_7 - y_1)/(x_7 - x_1)$
$S_{17} = (360 - 230)/(550 - 421)$
$S_{17} = 1{,}0078$

$S_{18} = (y_8 - y_1)/(x_8 - x_1)$
$S_{18} = (425 - 230)/(750 - 421)$
$S_{18} = 0{,}5927$

$S_{23} = (y_3 – y_2)/(x_3 – x_2)$
$S_{23} = (315 – 175)/(618 – 278)$
$S_{23} = 0{,}4118$

$S_{24} = (y_4 – y_2)/(x_4 – x_2)$
$S_{24} = (290 – 175)/(482 – 278)$
$S_{24} = 0{,}5637$

$S_{25} = (y_5 – y_2)/(x_5 – x_2)$
$S_{25} = (275 – 175)/(465 – 278)$
$S_{25} = 0{,}5348$

$S_{26} = (y_6 – y_2)/(x_6 – x_2)$
$S_{26} = (150 – 175)/(105 – 278)$
$S_{26} = 0{,}1445$

$S_{27} = (y_7 – y_2)/(x_7 – x_2)$
$S_{27} = (360 – 175)/(550 – 278)$
$S_{27} = 0{,}6801$

$S_{28} = (y_8 – y_2)/(x_8 – x_2)$
$S_{28} = (425 – 175)/(750 – 278)$
$S_{28} = 0{,}5297$

$S_{34} = (y_4 – y_3)/(x_4 – x_3)$
$S_{34} = (290 – 315)/(482 – 618)$
$S_{34} = 0{,}1838$

$S_{35} = (y_5 – y_3)/(x_5 – x_3)$
$S_{35} = (275 – 315)/(465 – 618)$
$S_{35} = 0{,}2614$

$S_{36} = (y_6 – y_3)/(x_6 – x_3)$
$S_{36} = (150 – 315)/(105 – 618)$
$S_{36} = 0{,}3216$

$S_{37} = (y_7 – y_3)/(x_7 – x_3)$
$S_{37} = (360 – 315)/(550 – 618)$
$S_{37} = - 0{,}6618$

$S_{38} = (y_8 – y_3)/(x_8 – x_3)$
$S_{38} = (425 – 315)/(750 – 618)$
$S_{38} = 0{,}8333$

$S_{45} = (y_5 - y_4)/(x_5 - x_4)$
$S_{45} = (275 - 290)/(465 - 482)$
$S_{45} = 0,8824$

$S_{46} = (y_6 - y_4)/(x_6 - x_4)$
$S_{46} = (150 - 290)/(105 - 482)$
$S_{46} = 0,3714$

$S_{47} = (y_7 - y_4)/(x_7 - x_4)$
$S_{47} = (360 - 290)/(550 - 482)$
$S_{47} = 1,0294$

$S_{48} = (y_8 - y_4)/(x_8 - x_4)$
$S_{48} = (425 - 290)/(750 - 482)$
$S_{48} = 0,5037$

$S_{56} = (y_6 - y_5)/(x_6 - x_5)$
$S_{56} = (150 - 275)/(105 - 465)$
$S_{56} = 0,3472$

$S_{57} = (y_7 - y_5)/(x_7 - x_5)$
$S_{57} = (360 - 275)/(550 - 465)$
$S_{57} = 1,0000$

$S_{58} = (y_8 - y_5)/(x_8 - x_5)$
$S_{58} = (425 - 275)/(750 - 465)$
$S_{58} = 0,5263$

$S_{67} = (y_7 - y_6)/(x_7 - x_6)$
$S_{67} = (360 - 150)/(550 - 105)$
$S_{67} = 0,4719$

$S_{68} = (y_8 - y_6)/(x_8 - x_6)$
$S_{68} = (425 - 150)/(750 - 105)$
$S_{68} = 0,4264$

$S_{78} = (y_8 - y_7)/(x_8 - x_7)$
$S_{78} = (425 - 360)/(750 - 550)$
$S_{78} = 0,3250$

Para encontrar a mediana, colocam-se os dados de S_{ij} encontrados anteriormente em rol.

-0,6618;	0,1445;	0,1838;	0,2532;	0,2614;	0,3216;	0,3250;	0,3472;
0,3714;	0,3846;	0,4118;	0,4264;	0,4315;	**0,4719;**	**0,5037;**	0,5263;
0,5297;	0,5348;	0,5637;	0,5927;	0,6801;	0,8333;	0,8824;	0,9836;
1,0000;	1,0078;	1,0227;	1,0294				

Após realizar o rol, deve-se determinar a posição dos elementos medianos. Como se tem uma amostra de tamanho par, essa posição é determinada da seguinte maneira, conforme Fonseca e Martins (1996):

$$P_{Md} = \frac{n}{2} = \frac{28}{2} = 14^{\circ} \quad \text{e} \quad P_{Md} = \frac{n}{2} + 1 = \frac{28}{2} + 1 = 15^{\circ}$$

Assim a mediana será a média aritmética dos elementos que ocupam a 14º e a 15º posições, que estão em negrito no rol, então:

$$Md = \frac{0{,}4719 + 0{,}5037}{2} = 0{,}4878$$

Logo o valor estimado para o coeficiente angular da equação de regressão é $\hat{\beta} = 0{,}4878$

Para obter o estimador α pelos dois métodos de Dietz (1989 apud Daniel, 1999. p.718), primeiramente determina-se $\hat{\alpha}_1$ pela equação (10) encontrando as possíveis estimativas para α:

$Y_1 - \hat{\beta}X_1$ = 230 - 0,4878 . 421 = 24,6362

$Y_2 - \hat{\beta}X_2$ = 175 – 0,4878 . 278 = 39,3916

$Y_3 - \hat{\beta}X_3$ = 315 – 0,4878 . 618 = 13,5396

$Y_4 - \hat{\beta}X_4$ = 290 – 0,4878 . 482 = 54,8804

$Y_5 - \hat{\beta}X_5$ = 275 – 0,4878 . 465 = 48,1730

$Y_6 - \hat{\beta}X_6$ = 150 – 0,4878 . 105 = 98,7810

$Y_7 - \hat{\beta}X_7$ = 360 – 0,4878 . 550 = 91,7100

$Y_8 - \hat{\beta}X_8$ = 425 – 0,4878 . 750 = 59,1500

Para encontrar a mediana, utiliza-se o mesmo procedimento visto anteriormente.

Colocam-se os dados em rol:

13,5396; 24,6362; 39,3916; **48,1730; 54,8804;** 59,1500; 91,7100; 98,7810

Determinam-se as posições dos elementos medianos:

$$P_{Md} = \frac{n}{2} = \frac{8}{2} = 4^{\circ} \qquad \text{e} \qquad P_{Md} = \frac{n}{2} + 1 = \frac{8}{2} + 1 = 5^{\circ}$$

A mediana será a média aritmética dos elementos que ocupam a 4º e a 5º posições em destaque no rol, logo:

$$Md = \frac{48,1730 + 54,8804}{2} = 51,5267$$

Assim o valor estimado do coeficiente linear é: $\hat{\alpha}_1 = 51,5267$

Após encontrar $\hat{\alpha}_1$, encontra-se $\hat{\alpha}_2$ através das médias dos valores das variáveis dependente e independente como abaixo:

$$\overline{Y}_1 = \frac{y_2 + y_1}{2} = \frac{175 + 230}{2} = 202,5 \qquad \overline{X}_1 = \frac{x_2 + x_1}{2} = \frac{278 + 421}{2} = 349,5$$

$$\overline{Y}_2 = \frac{y_2 + y_3}{2} = \frac{175 + 315}{2} = 245,0 \qquad \overline{X}_2 = \frac{x_2 + x_3}{2} = \frac{278 + 618}{2} = 448,0$$

$$\overline{Y}_3 = \frac{y_2 + y_4}{2} = \frac{175 + 290}{2} = 232,5 \qquad \overline{X}_3 = \frac{x_2 + y_4}{2} = \frac{278 + 482}{2} = 380,0$$

$$\overline{Y}_4 = \frac{y_2 + y_5}{2} = \frac{175 + 275}{2} = 225,0 \qquad \overline{X}_4 = \frac{x_2 + x_5}{2} = \frac{278 + 465}{2} = 371,5$$

$$\overline{Y}_5 = \frac{y_2 + y_6}{2} = \frac{175 + 150}{2} = 162,5 \qquad \overline{X}_5 = \frac{x_2 + x_6}{2} = \frac{278 + 105}{2} = 191,5$$

$$\overline{Y}_6 = \frac{y_2 + y_7}{2} = \frac{175 + 360}{2} = 267,5 \qquad \overline{X}_6 = \frac{x_2 + x_7}{2} = \frac{278 + 550}{2} = 414,0$$

$$\overline{Y}_7 = \frac{y_2 + y_8}{2} = \frac{175 + 425}{2} = 300,0 \qquad \overline{X}_7 = \frac{x_2 + x_8}{2} = \frac{278 + 750}{2} = 514,0$$

$$\overline{Y}_8 = \frac{y_3 + y_1}{2} = \frac{315 + 230}{2} = 272,5 \qquad \overline{X}_8 = \frac{x_3 + x_1}{2} = \frac{618 + 421}{2} = 519,5$$

$$\overline{Y}_9 = \frac{y_3 + y_4}{2} = \frac{315 + 290}{2} = 302,5 \qquad \overline{X}_9 = \frac{x_3 + x_4}{2} = \frac{618 + 482}{2} = 550,0$$

$$\overline{Y}_{10} = \frac{y_3 + y_5}{2} = \frac{315 + 275}{2} = 295{,}0 \qquad \overline{X}_{10} = \frac{x_3 + x_5}{2} = \frac{618 + 465}{2} = 541{,}5$$

$$\overline{Y}_{11} = \frac{y_3 + y_6}{2} = \frac{315 + 150}{2} = 232{,}5 \qquad \overline{X}_{11} = \frac{x_3 + x_6}{2} = \frac{618 + 105}{2} = 361{,}5$$

$$\overline{Y}_{12} = \frac{y_3 + y_7}{2} = \frac{315 + 360}{2} = 337{,}5 \qquad \overline{X}_{12} = \frac{x_3 + x_7}{2} = \frac{618 + 550}{2} = 584{,}0$$

$$\overline{Y}_{13} = \frac{y_3 + y_8}{2} = \frac{315 + 425}{2} = 370{,}0 \qquad \overline{X}_{13} = \frac{x_3 + x_8}{2} = \frac{618 + 750}{2} = 684{,}0$$

$$\overline{Y}_{14} = \frac{y_4 + y_1}{2} = \frac{290 + 230}{2} = 260{,}0 \qquad \overline{X}_{14} = \frac{x_4 + x_1}{2} = \frac{482 + 421}{2} = 451{,}5$$

$$\overline{Y}_{15} = \frac{y_4 + y_5}{2} = \frac{290 + 275}{2} = 282{,}5 \qquad \overline{X}_{15} = \frac{x_4 + x_5}{2} = \frac{482 + 465}{2} = 473{,}5$$

$$\overline{Y}_{16} = \frac{y_4 + y_6}{2} = \frac{290 + 150}{2} = 220{,}0 \qquad \overline{X}_{16} = \frac{x_4 + x_6}{2} = \frac{482 + 105}{2} = 293{,}5$$

$$\overline{Y}_{17} = \frac{y_4 + y_7}{2} = \frac{290 + 360}{2} = 325{,}0 \qquad \overline{X}_{17} = \frac{x_4 + x_7}{2} = \frac{482 + 550}{2} = 516{,}0$$

$$\overline{Y}_{18} = \frac{y_4 + y_8}{2} = \frac{290 + 425}{2} = 357{,}5 \qquad \overline{X}_{18} = \frac{x_4 + x_8}{2} = \frac{482 + 750}{2} = 616{,}0$$

$$\overline{Y}_{19} = \frac{y_5 + y_1}{2} = \frac{275 + 230}{2} = 252{,}5 \qquad \overline{X}_{19} = \frac{x_5 + x_1}{2} = \frac{465 + 421}{2} = 443{,}0$$

$$\overline{Y}_{20} = \frac{y_5 + y_6}{2} = \frac{275 + 150}{2} = 212{,}5 \qquad \overline{X}_{20} = \frac{x_5 + x_6}{2} = \frac{465 + 105}{2} = 285{,}0$$

$$\overline{Y}_{21} = \frac{y_5 + y_7}{2} = \frac{275 + 360}{2} = 317{,}5 \qquad \overline{X}_{21} = \frac{x_5 + x_7}{2} = \frac{465 + 550}{2} = 507{,}5$$

$$\overline{Y}_{22} = \frac{y_5 + y_8}{2} = \frac{275 + 425}{2} = 350{,}0 \qquad \overline{X}_{22} = \frac{x_5 + x_8}{2} = \frac{465 + 750}{2} = 607{,}5$$

$$\overline{Y}_{23} = \frac{y_6 + y_1}{2} = \frac{150 + 230}{2} = 190{,}0 \qquad \overline{X}_{23} = \frac{x_6 + x_1}{2} = \frac{105 + 421}{2} = 263{,}0$$

$$\overline{Y}_{24} = \frac{y_6 + y_7}{2} = \frac{150 + 360}{2} = 255{,}0 \qquad \overline{X}_{24} = \frac{x_6 + x_7}{2} = \frac{105 + 550}{2} = 327{,}5$$

$$\overline{Y}_{25} = \frac{y_6 + y_8}{2} = \frac{150 + 425}{2} = 287{,}5 \qquad \overline{X}_{25} = \frac{x_6 + x_8}{2} = \frac{105 + 750}{2} = 427{,}5$$

$$\overline{Y}_{26} = \frac{y_7 + y_1}{2} = \frac{360 + 230}{2} = 295{,}0 \qquad \overline{X}_{26} = \frac{x_7 + x_1}{2} = \frac{550 + 421}{2} = 485{,}5$$

$$\overline{Y}_{27} = \frac{y_7 + y_8}{2} = \frac{360 + 425}{2} = 392{,}5 \qquad \overline{X}_{27} = \frac{x_7 + x_8}{2} = \frac{550 + 750}{2} = 650{,}0$$

$$\overline{Y}_{28} = \frac{y_8 + y_1}{2} = \frac{425 + 230}{2} = 327{,}5 \qquad \overline{X}_{28} = \frac{x_8 + x_1}{2} = \frac{750 + 421}{2} = 585{,}5$$

Agora se determinam todos os valores possíveis para a estimativa de α substituindo-se as médias para Y e X encontrados na equação (10). Tem-se:

$\overline{Y}_1 - \hat{\beta}.\overline{X}_1$ = 202,5 – 0,4878 . 349,5 = 32,0139

$\overline{Y}_2 - \hat{\beta}.\overline{X}_2$ = 245 – 0,4878 . 448 = 26,4656

$\overline{Y}_3 - \hat{\beta}.\overline{X}_3$ = 232,5 – 0,4878 . 380 = 47,1360

$\overline{Y}_4 - \hat{\beta}.\overline{X}_4$ = 225 – 0,4878 . 371,5 = 43,7823

$\overline{Y}_5 - \hat{\beta}.\overline{X}_5$ = 162,5 – 0,4878 . 191,5 = 69,0863

$\overline{Y}_6 - \hat{\beta}.\overline{X}_6$ = 267,5 – 0,4878 . 414 = 65,5508

$\overline{Y}_7 - \hat{\beta}.\overline{X}_7$ = 300 – 0,4878 . 514 = 49,2708

$\overline{Y}_8 - \hat{\beta}.\overline{X}_8$ = 272,5 – 0,4878 . 519,5 = 19,0879

$\overline{Y}_9 - \hat{\beta}.\overline{X}_9$ = 302,5 – 0,4878 . 550 = 34,2100

$\overline{Y}_{10} - \hat{\beta}.\overline{X}_{10}$ = 295 – 0,4878 . 541,5 = 30,8563

$\overline{Y}_{11} - \hat{\beta}.\overline{X}_{11}$ = 232,5 – 0,4878 . 361,5 = 56,1603

$\overline{Y}_{12} - \hat{\beta}.\overline{X}_{12}$ = 337,5 – 0,4878 . 584 = 52,6248

$\overline{Y}_{13} - \hat{\beta}.\overline{X}_{13}$ = 370 – 0,4878 . 684 = 36,3448

$\overline{Y}_{14} - \hat{\beta}.\overline{X}_{14}$ = 260 – 0,4878 . 451,5 = 39,7583

$\overline{Y}_{15} - \hat{\beta}.\overline{X}_{15}$ = 282,5 – 0,4878 . 473,5 = 51,5267

$\overline{Y}_{16} - \hat{\beta}.\overline{X}_{16}$ = 220 – 0,4878 . 293,5 = 76,8307

$\overline{Y}_{17} - \hat{\beta}.\overline{X}_{17}$ = 325 – 0,4878 . 516 = 73,2952

$\overline{Y}_{18} - \hat{\beta}.\overline{X}_{18}$ = 357,5 – 0,4878 . 616 = 57,0152

$\overline{Y}_{19} - \hat{\beta}.\overline{X}_{19}$ = 252,5 – 0,4878 . 443 = 36,4046

$\overline{Y}_{20} - \hat{\beta}.\overline{X}_{20}$ = 212,5 – 0,4878 . 285 = 73,4770

$\overline{Y}_{21} - \hat{\beta}.\overline{X}_{21}$ = 317,5 – 0,4878 . 507,5 = 69,9415

$\overline{Y}_{22} - \hat{\beta}.\overline{X}_{22}$ = 350 – 0,4878 . 607,5 = 53,6615

$\overline{Y}_{23} - \hat{\beta}.\overline{X}_{23}$ = 190 – 0,4878 . 263 = 61,7086

$\overline{Y}_{24} - \hat{\beta}.\overline{X}_{24}$ = 255 – 0,4878 . 327,5 = 95,2455

$\overline{Y}_{25} - \hat{\beta}.\overline{X}_{25}$ = 287,5 – 0,4878 . 427,5 = 78,9655

$\overline{Y}_{26} - \hat{\beta}.\overline{X}_{26}$ = 295 – 0,4878 . 485,5 = 58,1731

$\overline{Y}_{27} - \hat{\beta}.\overline{X}_{27}$ = 392,5 – 0,4878 . 650 = 75,4300

$\overline{Y}_{28} - \hat{\beta}.\overline{X}_{28}$ = 327,5 – 0,4878 . 585,5 = 41,8931

Encontrados todos os valores, junta-se com os do $\hat{\alpha}_1$ e determina-se desta forma $\frac{n(n+1)}{2} = \frac{8(8+1)}{2} = 36$ possíveis estimativas para $\hat{\alpha}_2$.

13,5356; 19,0879; 24,6362; 26,4656; 30,8563; 32,0139; 34,2100; 36,3448; 36,4046; 39,3916; 39,7583; 41,8931; 43,7823; 47,1360; 48,1730; 49,2708; 51,5267; **52,6248; 53,6615;** 54,8804; 56,1603; 57,0152; 58,1731; 59,1500; 61,7086; 65,5508; 69,0863; 69,9415; 73,2952; 73,4770; 75,4300; 76,8307; 78,9655; 91,7100; 95,2455; 98,7810

Com os dados em rol, determinam-se as posições dos elementos medianos:

$$P_{Md} = \frac{n}{2} = \frac{36}{2} = 18^\circ \qquad e \qquad P_{Md} = \frac{n}{2} + 1 = \frac{36}{2} + 1 = 19^\circ$$

Encontra-se a mediana que será a média aritmética dos elementos que ocupam a 18º e a 19º posições em negrito no rol, assim:

$$Md = \frac{52,6248 + 53,6615}{2} = 53,1432$$

Portanto o outro valor estimado para o coeficiente linear é: $\hat{\alpha}_2 = 53,1432$

Logo as equações estimadas pelo método não-paramétrico são:

$$\hat{y}_1 = 51{,}5267 + 0{,}4878x$$
$$\hat{y}_2 = 53{,}1432 + 0{,}4878x$$

Conforme Dietz (1989), para se calcular a inclinação no problema de regressão usa-se $\hat{\beta}$. A melhor escolha para um estimador do intercepto é $\hat{\alpha}_2$ pois assume a simetria da distribuição dos erros.

Mostrar-se-á a partir das Tabelas, a seguir, os exemplos realizados através do VBA.

A Tabela 2 mostra os resultados da comparação dos parâmetros e critérios pelos métodos não-paramétrico e paramétrico, onde as pressuposições de homocedasticidade, independência de erros e normalidade falham na paramétrica. A Tabela 3 mostra os resultados da comparação dos parâmetros e critérios para os dois métodos, não ocorrendo nenhuma violação das pressuposições requeridas no MMQ.

As amostras analisadas, para o estabelecimento das Tabelas 2 e 3, foram retiradas de exemplos de livros e dissertações.

Na Tabela 2 quando se coloca homocedasticidade, independência de erros e normalidade é porque somente essas pressuposições foram violadas. Para a verificação das pressuposições foi realizado o teste de Goldfeld-Quandt para homocedasticidade, Darbin-Watson para a independência dos resíduos e Lilliefors para a normalidade.

Tabela 2 - Resultados das estimativas e dos critérios utilizados para comparar os modelos nos métodos paramétrico e não-paramétrico, quando há violações das pressuposições de homocedasticidade, independência de erros e normalidade.

Homocedasticidade	**PARAMÉTRICA**	**NÃO – PARAMÉTRICA**	
		$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\hat{\alpha}$	3,526	3,500	3,500
$\hat{\beta}$	0,512	0,500	0,500
AIC	235,374	235,523	235,523
BIC	239,198	239,347	239,347
EQM	2,045	2,051	2,051
SQR	102,257	102,562	102,562
r	0,540	0,540	0,540
r^2	0,292	0,292	0,292
Equação	$y=3,526+0,512x$	$y_1=3,500+0,500x$	$y_2=3,500+0,500x$
Independência de erros		$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\hat{\alpha}$	199,950	187,130	186,120
$\hat{\beta}$	-0,361	-0,250	-0,250
AIC	238,237	239,708	239,409
BIC	241,039	242,511	242,211
EQM	82,000	86,121	85,265
SQR	2460,002	2583,647	2557,963
r	0,543	0,543	0,543
r^2	0,294	0,294	0,294
Equação	$y=199,950-0,361x$	$y_1=187,130-0,250x$	$y_2=186,120-0,250x$
Normalidade		$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\hat{\alpha}$	-67,810	-71,100	-71,000
$\hat{\beta}$	73,452	75,000	75,000
AIC	565,345	566,037	565,883
BIC	569,754	570,447	570,292
EQM	64,948	65,623	65,472
SQR	4351,545	4396,767	4386,647
r	0,692	0,692	0,692
r^2	0,479	0,479	0,479
Equação	$y=-67,810+73,452x$	$y_1=-71,100+75,000x$	$y_2=-71,000+75,000x$

Pela análise da Tabela 2, percebe-se que as estimativas de α e β encontradas pelos dois métodos foram muito próximas, não havendo grandes diferenças entre os critérios AIC, BIC e EQM.

No método não-paramétrico pode-se observar que o EQM de $\hat{\alpha}_2$ foi menor ou igual a $\hat{\alpha}_1$, o que está de acordo com Dietz (1989) que diz que o estimador $\hat{\alpha}_2$ tem um EQM menor que outros estimadores, com exceção de erros normais, onde o estimador de mínimos quadrados é preferível.

Observa-se que, quando a pressuposição de normalidade não é atendida o EQM da paramétrica é inferior ao EQM da não-paramétrica, contrariando Dietz (1989). O autor coloca que, quando a suposição de normalidade dos erros é satisfeita, o EQM é menor no MMQ, mas quando essa suposição não é aceita o EQM do MMQ é maior que qualquer outro método.

Tabela 3 - Resultados das estimativas e dos critérios utilizados para comparar os modelos nos métodos paramétrico e não-paramétrico, quando não há violação das pressuposições.

	PARAMÉTRICA	**NÃO – PARAMÉTRICA**	
		$\hat{\alpha}_1$	$\hat{\alpha}_2$
$\hat{\alpha}$	124,823	118,816	124,900
$\hat{\beta}$	0,006	0,006	0,006
AIC	92,914	1048,270	1047,773
BIC	97,549	1052,905	1052,408
EQM	14727,780	14855,355	14757,234
SQR	1104583,000	1114151,700	1106792,560
R	0,920	0,920	0,920
r^2	0,848	0,848	0,848
Equação	$y = 124,823+0,006x$	$y_1 = 118,816+0,006x$	$y_2 = 124,900+ 0,006x$

Pela Tabela 3, pode-se verificar que as estimativas $\hat{\alpha}$ e $\hat{\beta}$ foram muito próximas nos dois métodos. Nos critérios AIC e BIC observou-se que na não-paramétrica os valores foram maiores que na paramétrica, pois não ocorreu nos dados, falha das pressuposições.

Pode-se observar que, quando as pressuposições ao MMQ são atendidas, este é a melhor opção para se estimar os coeficientes no modelo de regressão linear simples. Porém, quando alguma hipótese básica ao MMQ não é atendida, o método não-paramétrico é bem próximo ao paramétrico, embora se deva dar preferência ao não-paramétrico.

4.1 Comentários do Capítulo

Neste Capítulo, desenvolveu-se um exemplo passo-a-passo usando a metodologia não-paramétrica proposta no Capítulo 3, para mostrar como esta é muito trabalhosa, quando utilizada. Apresentou-se, também, através de exemplos, resultados realizados com ajuda do VBA que facilitou o uso da técnica não-paramétrica. Fez-se uma comparação entre os dois procedimentos com o intuito de verificar se o método não-paramétrico pode ser considerado uma alternativa ao MMQ.

5 CONCLUSÃO

Ao analisar a relação entre duas variáveis através de um modelo de regressão linear simples não se deve incorrer no erro de não verificar primeiramente se as hipóteses básicas do modelo estão sendo satisfeitas. Por meio da análise dos resíduos é possível encontrar o modelo que melhor se ajusta aos dados e possíveis violações dos pressupostos do MMQ, como: normalidade, homocedasticidade e independência de erros.

Neste presente trabalho o tema é o método não-paramétrico. Para avaliar o mesmo, compara-se as estimativas encontradas por ele com as estimativas obtidas pelo MMQ. Para facilitar os cálculos no método não-paramétrico, implementa-se um algoritmo no VBA. Analisa-se esses resultados e observa-se que, as estimativas encontradas pelos dois métodos são muito próximas e isso também acontece com os critérios. Através de estudos, afirma-se que o EQM de $\hat{\alpha}_2$ é menor que $\hat{\alpha}_1$ em todas as situações no método não-paramétrico. Em outro exemplo não ocorre violação, as estimativas são também muito próximas, mas os critérios AIC e BIC no método não-paramétrico obtêm valores maiores que o paramétrico. Assim como todas as hipóteses básicas são atendidas, pode-se dizer que o MMQ é a melhor opção para se estimar os coeficientes do modelo de regressão linear simples.

Salienta-se que o VBA facilita a utilização da técnica não-paramétrica para encontrar os coeficientes angular e linear da equação, pois este método exige uma grande quantidade de cálculos.

Diante disso, observa-se que a metodologia não-paramétrica não traz muitas divergências em relação à paramétrica, pode ser uma alternativa ao MMQ quando este não é o mais indicado.

Para trabalhos futuros, recomenda-se que se determine os intervalos de confiança e os testes de hipóteses para as estimativas encontradas pelo método não-paramétrico.

6 REFERÊNCIAS BIBLIOGRÁFICAS

DANIEL, W. W. **Applied nonparametric statistics**. Boston: Houghton Mifflin Company, 1978. p. 343 – 371.

____________. **Biostatistics:** A Foundation for Analysis in the Health Sciences 7th ed. New York: Jonh & Sons Wiley. 1999. p.717 – 720.

DIETZ, E. J. Teaching regression in a Nonparametric Statistics Couse. **The American Statistican**, vol.43, n.1, p.35-40, fev.1989.

FARIAS, E.R. de; ROCHA, F.J.S; LIMA, R.C. Critérios de seleção de modelos sazonais de séries temporais: Uma aplicação usando a taxa de desemprego da região Metropolitana de Recife. III Encontro Regional de Estudos do Trabalho – ABET, 22 a 24 de novembro 2000 – Recife. Disponível em: <http://www.race.nuca.ie.ufrj.br/abet/3seg>. Acesso em: 13 abr. 2005. p.47-67

FONSECA, J. S. da.; MARTINS, G. de A. **Curso de estatística.** 6. ed. São Paulo: Atlas, 1996. 320p.

FONSECA, J. S. da.; MARTINS, G. de A.; TOLEDO, G. L. **Estatística aplicada.** São Paulo: Atlas, 1985. 267p.

GUJARATI, D. N. **Econometria básica.** 3. ed. São Paulo: Makron Books, 2000. 846p.

HILL, C.; GRIFFITHS, W.; JUDGE, G. **Econometria.** São Paulo: Saraiva, 1999. 408p.

JACOBI, L. F. **Monitoração do processo de coletas de resíduos em Santa Maria – RS usando gráficos de controle de regressão**. 2001. 92f. Dissertação (Mestrado em Engenharia de Produção) – Universidade Federal de Santa Maria, Santa Maria, 2001.

LEVINE, D. M.; BERENSON, M. L.; STEPHAN, D. **Estatística:** teoria e aplicações. Rio de Janeiro: Livros Técnicos e Científicos, 2000. 811p.

SEN, K. P. Estimates of Regression Coefficient Based on Kendall's Tau. **Journal of the American Statistical Association**, v. 63, p.1379-1389, dez 1968.

SILVA, E. M. da.; SILVA, E. M. da. **Matemática e estatística aplicada.** São Paulo: Atlas, 1999. p. 13-62.

TOLEDO, G. L.; OVALLE, I. I. **Estatística básica.** 2. ed. São Paulo: Atlas, 1985. 459p.

VASCONCELLOS, M. A. S. & ALVES, D. **Manual de econometria.** São Paulo: Atlas, 2000. 308p.

WERKEMA, M. C. C. & AGUIAR, S. **Ferramentas da qualidade:** análise de regressão: como entender o relacionamento entre as variáveis de um processo. Belo Horizonte: QFCO, v. 7, 1996. 311p.