REGRESSÃO NÃO-LINEAR :
DESENVOLVIMENTO DE UM SISTEMA COMPUTACIONAL E APLICAÇÕES

EDSON JOSÉ CHIACCHIO
Licenciado em Matemática

ORIENTADOR : Prof. Dr. Cássio Roberto de Melo Godoi

Dissertação apresentada a Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo, para obtenção do título de Mestre em Agronomia. Área de concentração: **Estatística e Experimentação Agronômica.**

Piracicaba
Estado de São Paulo - Brasil
Março/93

Ficha catalográfica preparada pela Seção de Livros da
Divisão de Biblioteca e Documentação - FCLQ/USP

C532r Chiacchio, Edson José
Regressão não-linear; desenvolvimento de um sistema computacional e aplicações. Piracicaba, 1993.
149p.

Diss.(Mestre) - ESALQ
Bibliografia.

1. Análise de regressão 2. Computador - Sistema
3. Estatística matemática I. Escola Superior de Agricultura Luiz de Queiroz, Piracicaba

CDD 519.536

REGRESSÃO NÃO-LINEAR :
DESENVOLVIMENTO DE UM SISTEMA COMPUTACIONAL E APLICAÇÕES

EDSON JOSÉ CHIACCHIO
Licenciado em Matemática

Aprovada em 03/05/1993

Comissão Julgadora :

Prof. Dr. Cássio Roberto de M. Godoi ESALQ/USP

Profª. Drª. Gabriela Stangenhauß IMECC/UNICAMP

Prof. Dr. Rodolfo Hoffmann ESALQ/USP

Prof. Dr. Cassio Roberto de Melo Godoi

Orientador

"A Ciência é construída juntando-se lógica e, ordenadamente, fragmentos do conhecimento, assim como uma casa é construída juntando-se lógica e, ordenadamente, tijolos.

Os fragmentos do conhecimento não formam uma Ciência, assim como um punhado de tijolos não constituem uma casa."

Henri Poincaré

(1854 - 1912)

À meus pais

José (in memorian)

Coraly

pela oportunidade

À meus irmãos

Coraly (in memorian)

Ary

Renato

Nancy

pela convivência

À minha esposa

Liberaci

pelo amor e compreensão

À meus filhos

Karyna

Davy

pela esperança

DEDICO

AGRADECIMENTOS

À Deus pela vida.

Ao Prof. Dr. Cássio Roberto de Melo Godoi, do departamento de Matemática e Estatística da ESALQ, pelo apoio, estímulo, orientação, e sobretudo por acreditar em minhas idéias.

Aos Professores do Departamento de Matemática e Estatística da ESALQ pela amizade, pelos ensinamentos e pelas valiosas sugestões.

Aos colegas da pós-graduação, pelos incentivos, e em especial ao amigo Paulo de Oliveira.

Ao pessoal do CIAGRI pelo apoio computacional, e em especial aos Srs. Andrés Enrique Lai Reyes e Marcelo Zacarias da Silva.

Aos funcionários do Departamento de Matemática e Estatística da ESALQ pela assistência.

Ao pessoal da Biblioteca Central, pelo auxílio na recuperação de muitas referências bibliográficas.

À CAPES pelo auxílio financeiro.

ÍNDICE

	Página
Lista de Figuras	xii
Lista de Quadros	xiii
Simbologia	xiv
Resumo	xvi
Summary	xviii
1. Introdução	1
2. Revisão de Literatura	5
2.1. Histórico	5
2.1.1. Histórico da Regressão de Mínimos Qua- drados	6
2.1.2. Histórico da Regressão de Desvios Ab- solutos	9
2.2. Modelos e Métodos Aplicados na Agricultura e Agropecuária	11
2.3. Classificação dos Modelos	13
2.4. Definição Matemática do Modelo	15
2.5. Suposições a Respeito do Vetor de Erros ε	16
2.6. Caracterização da Regressão em Função das Su- posições do Vetor de Erros ε	17
2.6.1. Modelos Ordinários	17
2.6.2. Modelos Ponderados	18
2.6.3. Modelos Generalizados	18
2.7. Classificação do Tipo da Regressão em Função da Amplitude dos Erros	18
2.8. Método dos Mínimos Quadrados Ordinários Apli- cado aos Modelos Não-Lineares	20
2.9. Linearização do Modelo Não-Linear pela Série de TAYLOR	21
2.10. Método dos Mínimos Quadrados para a Solução do Modelo Linearizado pela Série de TAYLOR ...	23
2.11. Cálculo das Novas Estimativas do Vetor θ	24

2.12. Critérios de Convergência	25
2.12.1. Critério do Erro Relativo	25
2.12.2. Critério de MARQUARDT	25
2.12.3. Critério de DENNIS & SCHNABEL	26
2.13. Método de GAUSS-NEWTON para Modelos Não- Lineares	26
2.13.1. Método Original de GAUSS-NEWTON	26
2.13.1.1. Algoritmo de GAUSS-NEWTON ..	28
2.13.1.2. Método de GAUSS-NEWTON para Modelos Não-Lineares Apli- cado a um Modelo Linear	30
2.13.2. Método de HARTLEY	31
2.14. Método de MARQUARDT para Modelos Não- Lineares	32
2.14.1. Método Original de MARQUARDT	32
2.14.2. Método de MARQUARDT - Estratégia de BARD	35
2.14.3. Método de MARQUARDT - Estratégia de TABATA & ITO	35
2.15. Método de NEWTON para Modelos Não-Lineares ...	36
2.15.1. Obtenção do Termo Não-Linear "\$" do Hessiano	38
2.15.2. A Proposta de DENNIS, GAY & WELSCH ...	41
2.16. Métodos QUASI-NEWTON	41
2.16.1. Regressão Não-Linear Usando os Mé- todos QUASI-NEWTON	42
2.16.2. Algumas Considerações Teóricas e Práticas	43
2.16.3. Algoritmo QUASI-NEWTON para Regres- são Não-Linear	44
2.16.4. Expressões QUASI-NEWTON Utilizadas Nesse Trabalho	45
2.16.4.1. DAVIDON, FLETCHER & POWELL (DFP)	45
2.16.4.2. BROYDEN, FLETCHER, GOLDFARB & SHANNO (BFGS)	46

2.17. Métodos Híbridos para Regressão Não-Linear ...	46
2.17.1. Introdução	46
2.17.1.1. Expressões QUASI-NEWTON no	
Contexto de GILL & MURRAY ..	47
2.17.1.1.1. DAVIDON, FLETCHER	
& POWELL (DFP)	47
2.17.1.1.2. BROYDEN, FLETCHER,	
GOLDFARB & SHANNO	
(BFGS)	47
2.17.2. Método de AL-BAALI	47
2.17.2.1. Algoritmo do Método Híbrido	
HBY2	48
2.17.3. Método de XU	48
2.17.3.1. Introdução	48
2.17.3.2. Algoritmo do Método HY2 ...	49
2.17.3.2.1. Expressão de BROYDEN	
(1965).....	50
2.18. Escolha do Método Dentre os Estudados	50
2.18.1. Considerações Práticas	51
2.19. Método dos Mínimos Quadrados Generalizados ...	52
2.19.1. Método Generalizado de GAUSS-NEWTON ..	52
2.19.2. Método Generalizado de MARQUARDT	52
2.19.3. Método QUASI-NEWTON Generalizado	52
2.20. Método dos Mínimos Quadrados Ponderados	52
2.21. O Problema da Estimativa Inicial	53
2.21.1. Método do Modelo Aproximado	53
2.21.2. Método de HARTLEY & BROOKER	54
2.21.3. Método de GALLANT	54
2.21.4. Método da "MALHA"	55
2.22. Introdução às Medidas de Não-Linearidade	55
2.23. Análise de Variâncias e Teste de Hipótese	
Sobre o Modelo	56
2.24. Matriz de Variâncias e Covariâncias	
Assintóticas	58
2.25. Intervalo de Confiança para os Parâmetros	59

2.26. Matriz de Correlação Assintótica dos	
Parâmetros	60
2.27. Análise dos Resíduos	60
2.27.1. Análise Gráfica	62
2.27.2. Análise Estatística	64
2.27.2.1. Teste de CHORRILHOS	64
2.27.2.2. Teste de LILLIEFORS	66
2.27.2.3. Outras Estatísticas da	
Normalidade	68
2.27.2.4. Teste de ANSCOMBE para	
Homocedasticidade	70
2.27.2.5. Teste de COOK & WEISBERG	
para Homocedasticidade.....	71
2.28. Matriz de Projeção Ortogonal	72
2.29. Simulação	73
2.29.1. Geração de Números Aleatórios com	
Distribuição Normal	74
2.30. Método da Máxima Verossimilhança para Modelos	
Não-Lineares	75
2.31. Regressão Não-Linear pelo Critério dos Desvios	
Absolutos ou Regressão Não-Linear " L_1 "	76
2.31.1. Método de SCHLOSSMACHER	76
2.31.2. Algoritmo de SCHLOSSMACHER	77
2.31.3. Algumas Considerações	77
3. Descrição do Sistema Computacional	79
3.1. Escolha do Modelo	79
3.1.1. Modelos da Biblioteca	80
3.1.1.1. Relação dos Modelos da Biblio-	
teca	80
3.1.1.2. Um Exemplo do Arquivo de Mode-	
los da Biblioteca	84
3.1.2. Modelos do Usuário	85
3.2. Hardware	87
3.2.1. Necessário	87
3.2.2. Opcional	87
3.3. Software Necessário	87

3.4. Descrição do Software	88
3.4.1. Descrição Gráfica das Opções do Software para Regressão Não-Linear.	88
3.4.2. Descrição dos Cálculos	90
3.4.2.1. Método de GAUSS-NEWTON ...	90
3.4.2.2. Método de MARQUARDT	92
3.4.2.3. Métodos QUASI-NEWTON	92
3.4.2.4. Método de SCHLOSSMACHER Original	93
3.4.2.5. Método de SCHLOSSMACHER Modificado	94
3.4.3. Descrição das Estatísticas	94
3.4.3.1. Análise de Variâncias	94
3.4.3.2. Intervalo de Confiança dos Parâmetros	95
3.4.3.3. Matriz de Variâncias/Cova- riâncias dos Parâmetros ..	95
3.4.3.4. Matriz de Correlação dos Parâmetros	96
3.4.3.5. Análise dos Resíduos	96
3.4.3.6. Verificação das Pressupo- sições do Modelo	97
3.4.4. Gráficos	97
3.4.4.1. Gráfico dos Resíduos Normalizados	97
3.4.4.2. Gráficos da Função	98
4. Discussão e Resultados	99
4.1. Introdução	99
4.2. Conjuntos de Dados e Modelos Adotados para Análise	101
4.2.1. Dados e Modelos de SILVA (1978)	101
4.2.2. Dados e Modelo de LANDRIANI et alli (1983)	102
4.3. Análise dos Dados de SILVA (1978)	103
4.3.1. Modelo: 1ª Lei de MITSCHERLICH	103
4.3.2. Modelo: Parábola	109

4.3.3. Modelo: 2ª Lei de MITSCHERLICH	113
4.4. Análise dos Dados de LANDRIANI et alli (1983) .	122
5. Conclusões e Sugestões para Outras Pesquisas	127
6. Referências Bibliográficas	128
ANEXO 1. Limite Superior para o Teste de LILLIEFORS ...	139
ANEXO 2. Limites da Distribuição de χ^2 ($\alpha = 0,05$)	140
ANEXO 3. Limites da Distribuição Bilateral t de Student ($\alpha = 0,05$)	141
ANEXO 4. Limites Críticos de "r" para o Teste de CHORRILHOS	142
ANEXO 5. Resultados Obtidos pelo Método de MARQUARDT e SCHLOSSMACHER modificado para os Dados de SILVA (1978) Usando a 1ª Lei de MITSCHERLICH .	143

LISTA DE FIGURAS

Figura	Página
1. Fluxograma do Método de MARQUARDT	34
2. Critério para Escolha da Constante de MARQUARDT-TABATA & ITO	36
3. Gráfico dos Resíduos	63
4. Relação dos Modelos da Biblioteca	80
5. Gráfico das Regressões L_2 para os Dados de SILVA (1978) Usando a 1ª Lei de MITSCHERLICH	107
6. Gráfico das Regressões L_1 para os Dados de SILVA (1978) Usando a 1ª Lei de MITSCHERLICH	108
7. Gráfico das Regressões L_2 para os Dados de SILVA (1978) Usando Parábola	111
8. Gráfico das Regressões L_1 para os Dados de SILVA (1978) Usando Parábola	112
9. Gráfico das Regressões L_2 para os Dados de SILVA (1978) Usando a 2ª Lei de MITSCHERLICH com 25 Iterações	117
10. Gráfico das Regressões L_1 para os Dados de SILVA (1978) Usando a 2ª Lei de MITSCHERLICH com 100 Iterações	118
11. Gráfico das Regressões L_2 para os Dados de SILVA (1978) Usando a 2ª Lei de MITSCHERLICH com 100 Iterações	120
12. Gráfico das Regressões L_1 para os Dados de SILVA (1978) Usando a 2ª Lei de MITSCHERLICH com 100 Iterações	121
13. Gráfico das Regressões L_2 para os Dados de LANDRIANI et alli (1983)	125
14. Gráfico das Regressões L_1 para os Dados de LANDRIANI et alli (1983)	126

LISTA DE QUADROS

Quadro	Página
1. Resumo das Análises dos Dados de SILVA (1978) Usando a 1ª Lei de MITSCHERLICH	106
2. Resumo das Análises dos Dados de SILVA (1978) Usando Parábola	110
3. Resumo das Análises dos Dados de SILVA (1978) Usando a 2ª Lei de MITSCHERLICH com 25 Iterações ...	116
4. Resumo das Análises dos Dados de SILVA (1978) Usando a 2ª Lei de MITSCHERLICH com 100 Iterações ..	119
3. Resumo das Análises dos Dados de LANDRIANI et alli (1983)	124

SIMBOLOGIA

Apresenta-se nessa seção os símbolos de uso corrente no texto, e seus respectivos significados :

$\underline{Y}$: representa um vetor coluna
 $\underline{X}$: representa uma matriz
 f : representa uma relação funcional
 n, p, α : representam escalares
 i, j, k, m : representam contadores
 $E[\cdot]$: Operador esperança matemática.
 $E[\cdot \cdot \cdot]$: Operador covariância.
 $\| \cdot \|$: Operador norma euclidiana.
 $\nabla R(\underline{X}; \underline{\theta})$: vetor gradiente, definido como:

$$\left[\begin{array}{ccc} \frac{\partial R}{\partial \theta_1} & \dots & \frac{\partial R}{\partial \theta_p} \end{array} \right]'$$

com p sendo o número de elementos em $\underline{\theta}$.

$\nabla^2 R(\underline{X}; \underline{\theta})$: matriz hessiana, tendo como elemento (i, j) :

$$\frac{\partial^2 R}{\partial \theta_i \partial \theta_j}$$

$\underline{\xi}_t$: vetor de correção pela fórmula de TAYLOR

$\hat{a}$: estimativa da variável (ou parâmetro) a

$\underline{\theta}$: vetor de parâmetros, com elementos :

$$[a \quad b \quad c \quad \dots \quad t]'$$

$$\underline{X}^2 = \begin{bmatrix} X_1^2 \\ \vdots \\ X_n^2 \end{bmatrix} \quad ; \quad \ln \underline{\varepsilon} = \begin{bmatrix} \ln \varepsilon_1 \\ \vdots \\ \ln \varepsilon_n \end{bmatrix}$$

Convencionase-se que :

$$Y_i = a + bX_i + cX_i^2 + \epsilon_i$$

será denotado por $\tilde{Y} = a\tilde{1} + b\tilde{X} + c\tilde{X}^2 + \tilde{\epsilon}$, onde $\tilde{1}' = [1 \ 1 \ \dots \ 1]$

e de forma análoga para os demais modelos, exceto no capítulo 3, onde os modelos estão escritos como uma função matemática :

$$Y = a + bX + cX^2 + \epsilon$$

$Z_u = [Z_{ij}^u]$ deve ser interpretado como o elemento Z_{ij} da matriz Z na iteração u .

P_{ii} é algumas vezes usado para representar P_{ii} , ou seja, o elemento da diagonal da matriz P .

REGRESSÃO NÃO-LINEAR :

DESENVOLVIMENTO DE UM SISTEMA COMPUTACIONAL E APLICAÇÕES

AUTOR : Edson José Chiacchio

ORIENTADOR : Prof. Dr. Cassio Roberto de Melo Godoi

RESUMO

Frequentemente deseja-se expressar informações provenientes de dados experimentais sob a forma de uma relação funcional:

$$\underline{Y} = f(\underline{X} ; \underline{\theta}) + \underline{\varepsilon}$$

onde :

$\underline{Y}$:vetor das observações.

$\underline{X}$:matriz do delineamento.

$\underline{\theta}$:vetor dos parâmetros do modelo matemático.

f :relação funcional que descreve o tipo de dependência existente entre as variáveis ($\underline{X}$) e a resposta ($\underline{Y}$) do experimento.

$\underline{\varepsilon}$:vetor dos erros do modelo suposto aleatório e normalmente distribuído, do experimento.

Nesse trabalho a relação funcional é suposta não-linear (podendo eventualmente ser linear), porém conhecida do pesquisador.

O problema então resume-se em encontrar estimativas do vetor dos parâmetros $\underline{\theta}$, que minimize o erro $\underline{\varepsilon}$, através da relação funcional f .

Diversos métodos têm sido propostos para estimar-se o vetor $\underline{\theta}$, dentre eles :

- Método de NEWTON
- Métodos de GAUSS-NEWTON e suas modificações
- Método de MARQUARDT
- Métodos QUASI-NEWTON
- Métodos Híbridos
- Método de SCHLOSSMACHER

Apresenta-se um estudo teórico de cada método e suas principais características e limitações.

Os diversos métodos podem ser melhor justificados quando classificam-se os problemas de acordo com a ordem de grandeza do erro (zero, pequeno e grande).

Projetou-se um "Software" destinado a pesquisadores da área agrônômica, mas que pode ser útil a pesquisadores de outras áreas. Algumas das opções são listadas a seguir :

- Seis métodos de estimação dos parâmetros
- Biblioteca com os modelos usuais (mais de 60 modelos)
- Análise da regressão
- Análise de resíduos
- Intervalo de confiança para os parâmetros
- Testes para verificação das pressuposições do modelo.

Incluso no programa encontra-se um Interpretador de Fórmulas, capaz de avaliar qualquer relação funcional (modelo) fornecida pelo usuário. Assim o sistema pode fazer a regressão de um modelo que depende única e exclusivamente do conhecimento do usuário.

Deve ficar claro que o usuário não necessita conhecer nenhuma linguagem de programação para operar o sistema.

Limitações do sistema proposto :

- Máximo número de observações: 1000
- Máximo número de parâmetros do modelo: 20
- Máximo nº de variáveis independentes: 20
- Se o módulo de algum valor numérico calculado exceder 10^{308} ou for menor que 10^{-308} o sistema aborta.

NONLINEAR REGRESSION :

SOFTWARE DEVELOPMENT AND APPLICATIONS

Author : Edson José Chiacchio

Adviser : Prof. Dr. Cassio Roberto de Melo Godoi

SUMMARY

Often is desired to express information proceeding from experimental data as of a functional relation :

$$\underline{Y} = f(\underline{X} ; \underline{\theta}) + \underline{\varepsilon}$$

Where :

$\underline{Y}$: observation vector

$\underline{X}$: matrix of the experimental design

$\underline{\theta}$: vector of the model parameters

f : functional operator describing the types of dependency existent among the fixed variables ($\underline{X}$) and the response ($\underline{Y}$)

$\underline{\varepsilon}$: vector of errors from the model, assumed to be random and normally distributed, of the experimental design

In this work the functional relation is supposed to be nonlinear (allowed eventually to be linear), but know by the researcher.

The problem then is to find estimates of the parameter vector $\underline{\theta}$, which minimizes the error term $\underline{\varepsilon}$ through the functional relationship vector proposed by the literature, among them

- NEWTON Method
- GAUSS-NEWTON's Method and their modifications
- MARQUARDT's Method
- QUASI-NEWTON's Methods
- Hybrid Methods
- SCHLOSSMACHER's Method

This work proposes a theoretical study of the each method including theirs characteristics and limitations.

The several methods are specially well justified when the problems are classified by the size of the errors (zero, small and large).

A software was written specially for agronomic researchers, but we felt it to be useful also for any kind of researcher. Some options available in the system are listed below :

- Six methods of parameter estimates.
- A library of the most common models (more the 60 models)
- Analysis of regression
- Confidence interval for the model parameters
- Residual analysis
- Test of model validity

Included in the computer program there is an Equation interpreter, which analyses a general model (functional relation) provided by the user. So, this software is able to do a regression whose model is dependent only upon the user knowledge.

It's not necessary to know computer's language to operate this system.

Limitations of the system :

- Maximum number of observations : 1000
- Maximum number of parameters : 20
- Maximum number of independent variables : 20
- Numerical precision : $10^{-308} \leq |K| \leq 10^{308}$

1. INTRODUÇÃO

Regressão é uma técnica estatístico-matemática com o objetivo de estimar o vetor de parâmetros, θ que, aplicado à relação funcional

$$Y = f(X; \theta) + \varepsilon$$

reproduz os dados originais com o menor erro, ε possível.

Os modelos aqui considerados resumem-se em três tipos básicos :

- **Modelos lineares** em relação aos parâmetros
- **Modelos linearizáveis**, que podem ser transformados em lineares através de alguma transformação (anamorfose)
- **Modelos não-lineares**, aqueles que não são lineares em relação aos parâmetros.

Nesse trabalho serão considerados apenas os modelos não-lineares, por tratarem-se de uma classe muito usada na modelagem de fenômenos da natureza, sejam eles físicos, químicos, biológicos, agrônômicos, etc.

Os modelos lineares e os linearizáveis serão tratados como casos particulares dos não-lineares.

Admite-se também, que a relação funcional seja conhecida pelo experimentador. Esse conhecimento pode

- Estar ligado ao desenvolvimento teórico

- Ser baseado em modelo apropriado para situações semelhantes àquela em estudo
- Ser de caráter exploratório

Assim, o problema resume-se em encontrar uma estimativa para o vetor de parâmetros θ .

Dois métodos, dentre outros, tem sido usados para a obtenção dessa estimativa:

- Método dos Mínimos Quadrados ;
- Método da Máxima Verossimilhança.

Técnicas especiais devem ser utilizadas quando se faz uso do Método dos Mínimos Quadrados (minimização da soma dos quadrados dos erros) em regressão não-linear.

As mais comuns são conhecidas como :

- Método de NEWTON
- Método de GAUSS-NEWTON e suas modificações
- Método do GRADIENTE (Steepest Descent)
- Método de MARQUARDT
- Métodos QUASI-NEWTON
- Métodos HÍBRIDOS

que serão objetos de estudo deste trabalho.

Tais métodos são conhecidos como "Regressão de Mínimos Quadrados" ou "Regressão de norma L_2 " ou simplesmente "Regressão L_2 ".

Sob a suposição de normalidade dos erros, ver-se-á que as estimativas de Máxima Verossimilhança são idênticas às de Mínimos Quadrados, pois maximizar a função de verossimilhança corresponde a minimizar a soma dos quadrados dos resíduos.

Um método de estimação mais robusto relativamente a esses dois métodos é conhecido como Método dos Desvios Absolutos ou Regressão de Desvios Absolutos ou Regressão de Norma L_1 ou simplesmente Regressão L_1 , que consiste em minimizar a soma dos módulos dos erros.

Na parte prática apresenta-se um programa de computador, arquitetado e desenvolvido pelo autor, implementando os métodos

- GAUSS-NEWTON
- MARQUARDT
- DAVIDON-FLETCHER & POWELL (DFP)
- BROYDEN-FLETCHER-GOLDFARB-SHANNO (BFGS)
- SCHLOSSMACHER
- SCHLOSSMACHER modificado.

Os modelos mais comuns estão disponíveis em uma "Biblioteca" de fácil acesso ao usuário: para modelos específicos foi desenvolvido um Interpretador de Fórmulas que permite ao usuário escrever sua própria relação funcional (modelo) a partir do teclado. Não necessita :

- A inclusão do modelo no programa fonte
- Um arquivo separado contendo o modelo
- "Linkagem" do modelo como subrotina

Após o ajuste da Regressão L_2 , pode-se obter algumas estatísticas, tais como :

- Intervalo de confiança para parâmetros
- Teste da hipótese de validade do modelo
- Matriz de variâncias e covariâncias das estimativas dos parâmetros
- Análise dos resíduos
 - Teste de aleatoriedade
 - Teste de normalidade
 - Teste de homocedasticidade

Se o usuário optar pela Regressão L_1 , então o relatório do programa só permite comparar os ajustes de modo descritivo (estimativas pontuais e gráficos), pois não se encontrou testes de ajuste para a Regressão L_1 não-linear.

Ressalta-se que o usuário não necessita conhecer qualquer linguagem de programação, pois o programa é baseado em menus auto-explicativos.

Limitações do sistema :

- Nº máximo de observações : 1000
- Nº máximo de parâmetros : 20
- Nº máximo de var. indep. : 20
- Precisão numérica :
 $[-10^{+308}, -10^{-308}] \cup [+10^{-308}, +10^{+308}]$

2. REVISÃO DE LITERATURA

As Técnicas de Regressão surgiram da necessidade de se representar dados provenientes de tabelas, sob a forma de uma expressão matemática. Exemplos de duas utilidades importantes:

- Modelagem de um fenômeno (físico, químico, biológico, agrônômico, etc.). para efeito de estudo, comparação ou previsão.
- Armazenamento e recuperação de dados, sob a forma de uma ou mais expressões matemáticas, como por exemplo a tabela de F (JOHNSON (1973)).

O termo regressão é devido ao "experimento" de GALTON cuja conclusão era que a cada geração que passasse a altura do ser humano tenderia à média (MIRSHAWKA (1983)).

2.1. HISTÓRICO

DE MOIVRES (1738) em seu trabalho *Doctrine of Chances* foi o primeiro a formular um modelo em termos de erros aleatórios. Sua contribuição é conhecida como

"Princípio da Média Aritmética".⁽¹⁾

EULER & MAYER (1748) em trabalhos independentes desenvolveram o Método das Médias, que consiste em subdividir as observações em grupos e trabalhar com as médias dos grupos.⁽¹⁾

DAVIS (1942) apresenta um resumo dessa teoria e alguns exemplos.

2.1.1. Histórico da Regressão de Mínimos Quadrados

GAUSS (1809) sugeriu a minimização da soma dos quadrados dos erros como critério para a Regressão de Mínimos Quadrados - **Regressão L_2** .

SEAL (1967) cita que LEGENDRE (1805) já havia sugerido o Método dos Mínimos Quadrados.

A literatura consultada não faz referência a Métodos de Regressão Não-Linear no período 1900-1930.

CURRY (1944) utiliza o Método do GRADIENTE (Steepest Descent) para resolver problemas não-lineares. Uma breve descrição do método proposto pode ser encontrada em FIGUEIRÔA (1986).

LEVENBERG (1944) sugere que se minimize a soma de quadrado de resíduos localmente, o que levaria o método do GRADIENTE a convergir. Esse método foi melhor explorado por MARQUARDT (1963), embora FLETCHER (1971) tenha afirmado *que as idéias de Levenberg nunca foram bem compreendidas de tal forma que ainda hoje não se consegue escrever um programa de computador capaz de utilizar seu método.*

HARTLEY (1948) apresenta o **Método dos Mínimos Quadrados Internos** que consiste em ter uma equação de regressão em que a variável dependente Y está relacionada

⁽¹⁾ citado em "A Course of Regression" Class notes of Iowa University. Biblioteca do IMECC/UNICAMP.

com sua própria soma repetida, sendo $\underline{Y}$ uma função de $\underline{X}$. O método é ilustrado através de exemplos práticos.

BOOTH, BOX, MÜLLER & PETERSON (1959), citados por MARQUARDT (1963), desenvolveram o primeiro programa (comercial) de computador que resolvia problemas de Regressão Não-Linear.

DAVIDON (1959), citado por FLETCHER & POWELL (1963), descreve um procedimento para minimizar uma função genérica. Atualmente o Método de DAVIDON, simplificado por FLETCHER & POWELL, conhecido por DFP, é dado como o primeiro "Método QUASI-NEWTON".

HARTLEY (1961) apresenta uma modificação no Método de GAUSS-NEWTON. A modificação consiste em encontrar um escalar no intervalo $[0;1]$ de tal forma que haja uma redução na soma dos quadrados do resíduo entre duas iterações consecutivas. Em sua proposta original busca-se o mínimo de uma parábola que passa pela soma dos quadrados dos resíduos, estimada em 3 pontos da atual iteração. Na teoria o método é mostrado ser convergente, mas na prática ocorrem problemas numéricos (aritmética em ponto flutuante).

O próprio HARTLEY (1964) sugere que, ao invés da parábola, seja considerada a sequência $(1/i)^2$, com $i = 0,1,2,\dots$ e toma-se o primeiro valor da sequência para a qual haja uma redução da soma de quadrado de resíduos.

GALLANT (1987) sugere que a sequência a ser considerada seja $[1; 0,9; 0,8; 0,7; 0,6; 1/2; 1/4; 1/8; 1/16; \dots]$ e que o valor a ser escolhido seja o primeiro para o qual haja uma redução da soma de quadrados de resíduos.

MARQUARDT (1963) desenvolve um método que é uma combinação de outros dois já existentes, a saber o Método de GAUSS-NEWTON e o Método do GRADIENTE. Cabe aqui uma explicação:

- O Método de GAUSS-NEWTON (GN) citado por DRAPER & SMITH (1981) transforma o modelo não-linear

original em um modelo linear, usando o primeiro termo da Série de TAYLOR ao redor de uma estimativa inicial θ_0 .

- No Método do GRADIENTE (Steepest Descent) parte-se de um valor inicial θ_0 , toma como correção para a próxima iteração, o valor negativo do gradiente da soma de quadrados de resíduos.

A demonstração da validade do Método de MARQUARDT é complicada, porém a sua aplicação é simples (CHIAICCHIO (1991))

Diversos autores, dentre os quais FLETCHER (1971), SMITH & SHANNO (1971), BARD (1974), TABATA & ITO (1975), têm sugerido modificações no método de MARQUARDT.

FLETCHER (1971) propõe a aplicação de um fator multiplicativo na constante de MARQUARDT, λ , para acelerar o processo de convergência.

SMITH & SHANNO (1971) sugerem o cálculo de autovalores para determinar o valor de λ .

BARD (1974) considera λD ao invés de λI , onde D_{ii} depende dos valores da diagonal do "Hessiano" aproximado ($Z'Z$).

TABATA & ITO (1975) determinam o valor que incrementa ou decrementa λ através do "histórico" das iterações.

FLETCHER & POWELL (1963) descrevem um método de minimização de uma função genérica qualquer. Esse método é conhecido atualmente como DFP. Para ser utilizado em regressão o usuário deve minimizar a Função Soma de Quadrados de Resíduo. O método tem a vantagem de não necessitar do Sistema de Equações Normais (SEN).

VITALE & TAYLOR (1968) mostram através de exemplos como utilizar o Método DFP em Regressão Não-Linear.

BROYDEN, FLETCHER, GOLDFARB, SHANNO (1970) propuseram, independentemente, o que se conhece atualmente como o melhor Método QUASI-NEWTON, conhecido como BFGS. Tem

diversas vantagens sobre seus precedentes, e é utilizado da mesma forma que o DFP.

POWELL (1970), citado por SEBER & WILD (1989), propõe o primeiro Método HÍBRIDO, isto é, uma combinação entre os Métodos de GAUSS-NEWTON e do GRADIENTE.

DENNIS (1973) apresenta diversos métodos e compara suas vantagens e desvantagens. Dentre os métodos estudados estão os de NEWTON (original), GAUSS-NEWTON, LEVENBERG-MARQUARDT, BROWN-DENNIS, BROYDEN-DENNIS.

DENNIS, GAY & WELSCH (1981) aperfeiçoam o algoritmo de BROYDEN-DENNIS, reduzindo o Método de NEWTON (original) a uma combinação entre os Métodos de GAUSS-NEWTON e QUASI-NEWTON.

VESPUCCI (1987) faz uma revisão teórica no trabalho de DENNIS, GAY & WELSCH e apresenta um programa escrito em FORTRAN que implementa o método estudado.

AL-BAALI (1984) propõe vários Métodos Híbridos baseados na estrutura GAUSS-NEWTON/BFGS. Desenvolve um critério para decidir se o algoritmo a ser usado em cada iteração é GN ou BFGS. Conclui que o melhor Método Híbrido é o HBY2.

XU (1986) mostra que o algoritmo proposto por AL-BAALI não preserva algumas das características desejáveis do Método BFGS. Propõe um novo critério para decidir quando a iteração será GN e quando será BFGS. Ele enfatiza que a estrutura GN/BFGS não é alterada, diferindo apenas quanto ao critério de decisão.

2.1.2. Histórico da Regressão de Desvios Absolutos

BOSCOVICH (1757) sugeriu que a regressão deveria satisfazer aos critérios :

- 1- Soma dos resíduos ser igual a zero

2- Soma dos módulos dos erros seja minimizada.⁽²⁾

EDGEWORTH (1887) sugeriu que a primeira condição de BOSCOVICH fosse abandonada e que se usasse apenas a segunda condição. Esse procedimento é conhecido atualmente como Regressão L_1 ou Regressão de Desvios Absolutos. (Notou que na presença de pontos discrepantes, a Regressão de Desvios Absolutos (Regressão L_1) é superior à Regressão de Mínimos Quadrados (Regressão L_2).⁽²⁾)

KARST (1958) propõe um algoritmo para determinar os parâmetros de uma reta (com extensão para um hiperplano), usando a Regressão L_1 . Seu método é justificado geometricamente.

WAGNER (1959) mostra que o problema da Regressão (linear) L_1 pode ser resolvido usando Programação Linear.

FISHER (1961) afirma que o trabalho de KARST (1958) é uma repetição do método proposto por RHODES (1930). Resolve o problema da Regressão (linear múltipla) L_1 usando Programação Linear.

SCHLOSSMACHER (1973) apresenta um método iterativo para o cálculo da Regressão L_1 , baseado na Regressão L_2 . Embora não haja prova matemática que garanta sua convergência, na prática seus resultados tem sido comprovados por outros métodos. O algoritmo resolve tanto problemas de Regressão Linear quanto de Regressão Não-Linear.

ARMSTRONG & FROME (1976) criticam o trabalho de SCHLOSSMACHER, comparando seu método com a Programação Linear. Concluem que o Método da Programação Linear é mais eficiente.

⁽²⁾ citado em "A Course of Regression" Class notes of Iowa University. Biblioteca do IMECC/UNICAMP.

BAZARAA & SHETTY (1979) discutem diversos Métodos de Programação Não-Linear mostrando que o problema da Regressão Não-Linear pode ser resolvido usando a Programação Não Linear.

TISHLER & ZANG (1982) sugerem um algoritmo que apropriado para modelos lineares e não-lineares. Propõe uma aproximação (suavização) para o modelo, de tal forma que esse novo modelo possa ser resolvido usando algum Método QUASI-NEWTON.

2.2. Modelos e Métodos Aplicados na Agricultura e Agropecuária

Normalmente na Agricultura e na Agropecuária são usados modelos oriundos de conhecimentos teóricos do fenômeno estudado. Dessa forma surgiram os modelos de MITSCHERLICH, VON BERTALANFFY, BRODY, GOMPERTZ, SPILLIMAN, COBB-DOUGLAS, dentre outros.

Alguns fenômenos cíclicos tem sido estudados sob a forma de séries temporais, tais como fenômenos meteorológicos e físicos.

Devido à dificuldade de se estimar esses modelos, pois são não-lineares, alguns modelos lineares simples (retas, parábolas, raízes, polinômios, exponenciais, etc., e/ou suas combinações) são escolhidos.

Alguns trabalhos destacam-se na literatura, e são listados a seguir :

CORTARELLI (1973), citado por SILVEIRA JR (1976), ajustou o Modelo BRODY a dados de crescimento de bovinos, utilizando o Método Original de HARTLEY (1961).

SILVEIRA JR (1976) comparou os modelos de VON BERTALANFFY, BRODY, GOMPERTZ e logística quando aplicados aos dados de crescimento de bovinos. Utilizou o Método de HARTLEY (1961).

MELLO (1976) ajustou a 1ª Lei de Mitscherlich para 1, 2 e 3 variáveis independentes (3 modelos distintos), usando o Método de HARTLEY (1961).

SILVA (1978) ajustou as 1ª e 2ª Leis de Mitscherlich (1 variável independente) a dados de adubação de trigo, utilizando o Método de HARTLEY (1961). Comparou esses modelos com o Modelo Quadrático.

SILVA (1980) ajustou a 2ª Lei de Mitscherlich a dados de maturação de cana-de-açúcar pelos Métodos de NEWTON, HARTLEY (1961) e MARQUARDT (1963). Conclui que o Método de MARQUARDT é o mais eficiente.

UDO (1983) ajusta a 2ª Lei de Mitscherlich a dados de maturação de cana usando o Método de MARQUARDT modificado por BARD (1974). Compara os resultados obtidos com os Modelos Quadrático e Raiz Quadrada.

MAGNANI (1985) estuda a 2ª Lei de Mitscherlich (e a Regressão Quadrática) aplicados à maturação de cana-de-açúcar, usando os Métodos de NEWTON, GAUSS-NEWTON e MARQUARDT (pela estratégia de BARD (1974)). Conclui que o Método de MARQUARDT é o melhor.

FIGUEIRÔA (1986), usando o Método Original de GAUSS-NEWTON, estudou a função de COBB-DOUGLAS com erro aditivo. É um trabalho de cunho didático. Conclui que tendo-se boas estimativas iniciais (seção 2.21.) o Método de GAUSS-NEWTON é satisfatório.

MAGNANI (1991) aplicou o Método de MARQUARDT (modificado por BARD) à 2ª Lei de Mitscherlich. Sugere uma reparametização nessa lei baseado em medidas de não-linearidade do modelo. O autor mostra que as estimativas dos parâmetros e as propriedades dos estimadores são severamente afetadas se essas medidas de não-linearidade são consideradas grandes.

FELTRIN (1980) apresentou um estudo sobre Regressão Harmônica. Seu trabalho considerou que os períodos são conhecidos e o modelo tornou-se linear. Citou diversos outros autores que trabalharam com Regressão Harmônica dessa

forma. Nesse trabalho os períodos serão considerados desconhecidos, dessa forma o modelo é dito não-linear.

2.3. Classificação dos Modelos

Seja o modelo:

$$Y_i = f(X_i; \theta) + \varepsilon_i \quad [1]$$

com X e θ considerados "fixos" (conhecidos).

DRAPER & SMITH (1981) citam que os modelos podem ser classificados como :

- **Modelos Lineares**: aqueles que são lineares em relação aos parâmetros, isto é :

$$\frac{\partial f_i(X; \theta)}{\partial \theta_j} = g(X) \quad [2]$$

para $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, p$,
onde n é o número total de observações e p é o número de parâmetros do modelo.

- **Modelos Linearizáveis**: aqueles que podem ser transformados em lineares através de transformação, conhecida por anamorfose.
- **Modelos Não-Lineares**: aqueles que não são lineares em relação aos parâmetros. Nesse caso tem-se,

$$\frac{\partial f_i(X; \theta)}{\partial \theta_j} = g(X, \theta) \quad [3]$$

para $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, p$,
onde n é o número total de observações e p é o número de parâmetros do modelo.

Considere o modelo $Y = f(X; \theta) + \varepsilon$ dado por

$$Y = a1 + bX + cX^2 + \varepsilon$$

Assim $f(X; \theta) = a1 + bX + cX^2$

e $\theta = [a \ b \ c]'$

Então:

$$\frac{\partial f}{\partial a} = 1 \quad \frac{\partial f}{\partial b} = X \quad \frac{\partial f}{\partial c} = X^2$$

donde conclui-se que,

$$\frac{\partial f_i}{\partial \theta_j} = g(X)$$

e o modelo é dito linear.

Seja o modelo $Y = f(X; \theta) + \varepsilon$, caracterizado por :

$$Y = a1 + bc^X + \varepsilon,$$

assim $f(X; \theta) = a1 + bc^X$ e $\theta = [a \ b \ c]'$.

Então:

$$\frac{\partial f}{\partial a} = 1 \quad \frac{\partial f}{\partial b} = c^X \quad \frac{\partial f}{\partial c} = b \ X \ c^{X-1}$$

Conclui-se então que :

$$\frac{\partial f_i(X; \theta)}{\partial \theta_j} = g(X; \theta)$$

e o modelo é classificado como não-linear.

Considere o modelo caracterizado por :

$$Y = a^X \cdot \varepsilon$$

onde o erro ε é dito multiplicativo. Aplicando logaritmo a ambos lados da igualdade :

$$\begin{aligned}\ln \underline{Y} &= \ln (a^{\underline{X}} \cdot \underline{\varepsilon}) \\ \ln \underline{Y} &= \ln a^{\underline{X}} + \ln \underline{\varepsilon} \\ \ln \underline{Y} &= \underline{X} \cdot \ln a + \ln \underline{\varepsilon}\end{aligned}$$

fazendo :

$$\underline{Z} = \ln \underline{Y} ; \quad b = \ln a ; \quad \underline{e} = \ln \underline{\varepsilon}$$

o modelo se torna :

$$\underline{Z} = b\underline{X} + \underline{e}$$

que é linear pois :

$$\frac{\partial f}{\partial b} = \underline{X} = g(\underline{X})$$

Então o modelo $\underline{Y} = a^{\underline{X}} \cdot \underline{\varepsilon}$ é dito linearizável.

Se o modelo for $\underline{Y} = a^{\underline{X}} + \underline{\varepsilon}$, onde o erro $\underline{\varepsilon}$ é dito aditivo, não existe transformação capaz de tornar o modelo linear. É fácil verificar que

$$\frac{\partial f}{\partial a} = \underline{X} a^{\underline{X}-1} = g(\underline{X} ; a)$$

e o modelo considerado é não-linear.

2.4. Definição Matemática do Modelo

Considere o modelo genérico :

$$\underline{Y} = f(\underline{X} ; \underline{\theta}) + \underline{\varepsilon} \quad [1]$$

que de forma mais específica pode ser descrito como :

$$\underline{Y} = f(\underline{X}_1, \underline{X}_2, \dots, \underline{X}_v ; \underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_p) + \underline{\varepsilon} \quad [4]$$

onde :

$\underline{Y}_{n \times 1}$ é o vetor observações.

$\underline{X}_{n \times v}$ é a matriz das variáveis independentes, contendo n linhas de observações e v colunas de variáveis independentes.

$\underline{\theta}_{p \times 1}$ é o vetor dos parâmetros, desconhecido, que se quer estimar de modo a minimizar o erro $\underline{\varepsilon}$; contém p parâmetros.

ε_1 é o vetor de erros (desconhecido) e não observável, tendo cada elemento atribuído a uma observação; contém n valores.

2.5. Suposições a Respeito do Vetor de Erros ε

HOFFMANN & VIEIRA (1987), dentre outros, citam que a validade dos estimadores de mínimos quadrados (do vetor de parâmetros $\hat{\theta}$), da estatística F na análise de variância, da distribuição t para construção de intervalos de confiança, dentre outros, está vinculada a certas suposições sobre o vetor de erros ε , a saber :

I) A média dos erros é nula, isto é,

$$E [\varepsilon] = \phi$$

II) A variância do erro ε_i , $i = 1, 2, \dots, n$, é constante e igual a σ^2 , isto é :

$$E (\varepsilon \varepsilon') = \mathbb{I} \sigma^2$$

Nesse caso diz-se que os erros são homocedásticos.

III) O erro de uma observação é não correlacionado com o erro de outra observação, isto é :

$$E (\varepsilon_i \cdot \varepsilon_j) = 0, \quad i \neq j$$

IV) Os erros tem distribuição normal.

Combinando I, II e IV, pode-se escrever:

$$\varepsilon_i \sim N (0 ; \sigma^2)$$

ou

$$\varepsilon \sim N (\phi ; \mathbb{I} \sigma^2)$$

[5]

"A pressuposição I exclui, por exemplo, a existência de erros sistemáticos de medida da variável Y " (HOFFMANN et alli).

Se a pressuposição II (erros são Homocedáticos) não é atendida, o usuário poderá usar o Método dos Mínimos Quadrados Ponderados (seção 2.20) ou tentar algum tipo de transformação de dados, como por exemplo a sugerida por BOX e COX (1964).

Se a proposição III não é obedecida, diz-se que há autocorrelação nos resíduos. Nesse caso o usuário pode utilizar modelos generalizados (seção 2.19) ou mesmo tentar outros modelos, pois pode estar havendo um erro na especificação do modelo. Se $E[\varepsilon_{t+1} \cdot \varepsilon_t] > 0$ diz-se ter autocorrelação positiva. Se $E[\varepsilon_{t+1} \cdot \varepsilon_t] < 0$ a autocorrelação é negativa.

A pressuposição IV é necessária para que, dentre outras estatísticas, se possa usar o teste F da análise de variâncias e/ou construir intervalos de confiança para os parâmetros.

2.6. Caracterização da Regressão em Função das Suposições do Vetor de Erros $\underline{\varepsilon}$

2.6.1. Modelos Ordinários

São aqueles cuja a estrutura dos erros não viola nenhuma das pressuposições. Assim, [5] pode ser escrito de forma mais eficiente como :

$$\underline{\varepsilon} \sim \text{iid } N(\underline{0}; \mathbb{I} \sigma^2) \quad [6]$$

que se lê : "os erros são independentes e identicamente distribuídos, tendo distribuição normal de média zero e variância constante σ^2 ". $\mathbb{I}$ é a matriz identidade.

Dessa forma os modelos ordinários são resolvidos pelo Método dos Mínimos Quadrados Ordinários.

2.6.2. Modelos Ponderados

São aqueles cuja estrutura dos erros viola a pressuposição II. Nesse caso diz-se que os erros são heterocedásticos. Assim, escreve-se :

$$\varepsilon \sim N(\phi ; \mathbb{D} \sigma^2) \quad [7]$$

onde $\mathbb{D}$ é uma matriz diagonal, positiva definida, que pondera a variância σ^2 .

Modelos Ponderados são resolvidos utilizando o Método dos Mínimos Quadrados Ponderados.

2.6.3. Modelos Generalizados

São aqueles cuja estrutura dos erros viola a pressuposição III e possivelmente a II também. Assim diz-se que os erros são correlacionados (e possivelmente heterocedásticos). Nesse caso escreve-se:

$$\varepsilon \sim N(\phi ; \mathbb{W} \sigma^2) \quad [8]$$

onde $\mathbb{W}$ é uma matriz simétrica, positiva definida, que representa as variâncias e covariâncias dos erros.

Os modelos generalizados são resolvidos usando o Método dos Mínimos Quadrados Generalizados.

2.7. Classificação do Tipo da Regressão em Função da Amplitude dos Erros

A classificação é precedida da revisão (apresentação) de alguns conceitos fundamentais.

O vetor de erros é definido por:

$$\varepsilon = Y - f(X ; \theta) \quad [9]$$

Seja a matriz das derivadas de primeira ordem do vetor ε , conhecida por Jacobiano :

$$Z = Z_{kj} = \frac{\partial \varepsilon_k}{\partial \theta_j}, \quad k = 1, \dots, n, \quad j = 1, \dots, p \quad [10]$$

onde n é o número de observações (dados) e p é o número de parâmetros do modelo.

Considere a função :

$$R(X; \theta) = \varepsilon' \varepsilon \quad [11]$$

conhecida como Soma de Quadrado dos Erros.

DENNIS & SCHNABEL (1983), dentre outros, tem mostrado que o vetor das derivadas de primeira ordem de $R(X; \theta)$, conhecido como "Vetor Gradiente", é dado por :

$$\begin{aligned} {}_p g_1 &= \nabla R(X; \theta) \\ &= 2 \sum_{i=1}^n \varepsilon_i \nabla \varepsilon_i \\ &= 2 Z' \varepsilon \end{aligned} \quad [12]$$

De forma análoga, a matriz das derivadas de segunda ordem de $R(X; \theta)$, conhecida como "Matriz Hessiana", é dada por :

$$\begin{aligned} {}_p H_p &= \nabla^2 R(X; \theta) \\ &= 2 \sum_{i=1}^n (\nabla \varepsilon_i \cdot \nabla \varepsilon_i' + \varepsilon_i \cdot \nabla^2 \varepsilon_i^2), \quad i = 1, \dots, n \\ &= 2 (Z' Z + S) \end{aligned} \quad [13]$$

onde :

$$S = 2 \sum_{i=1}^n \varepsilon_i \cdot \nabla^2 \varepsilon_i \quad [14]$$

é o termo não-linear da Matriz Hessiana.

Seja $\hat{\theta}$ o estimador de θ , assim a estimativa de ε é dada por :

$$\hat{\varepsilon} = Y - f(X; \hat{\theta}) \quad [15]$$

de forma análoga os estimadores de $\mathbf{Z}$ e $\mathbf{S}$ são respectivamente $\hat{\mathbf{Z}}$ e $\hat{\mathbf{S}}$.

DENNIS & SCHNABEL (1983), AL-BAALI (1984), SEBER & WILD (1989), dentre outros, tem classificado o tipo de regressão em função da amplitude dos resíduos, como:

- Modelos de Resíduo Zero

Aqueles para o qual $\hat{\varepsilon} = 0$ ou também $\hat{\mathbf{S}} = 0$

- Modelos de Resíduos Pequenos

Aqueles em que $\hat{\mathbf{S}}$ é considerado pequeno em relação a $\hat{\mathbf{Z}}'\hat{\mathbf{Z}}$ (ou que $\hat{\varepsilon}$ pode ser considerado pequeno).

- Modelos de Resíduos Grandes

Aqueles em que $\hat{\mathbf{S}}$ é considerado grande em relação a $\hat{\mathbf{Z}}'\hat{\mathbf{Z}}$ (ou que $\hat{\varepsilon}$ pode ser considerado grande).

2.8. Método dos Mínimos Quadrados Ordinários Aplicado aos Modelos Não Lineares

Seja o conjunto de variáveis $(X_1, X_2, \dots, X_v, Y)$ ou de forma abreviada $(\mathbf{X}; \mathbf{Y})$ e o modelo $\mathbf{Y} = f(\mathbf{X}; \boldsymbol{\theta}) + \varepsilon$.

O objetivo do método é encontrar um vetor $\hat{\boldsymbol{\theta}}$, para o qual $\|\hat{\varepsilon}\|^2$ seja mínima.

Matematicamente :

$$\begin{aligned} \min_{\boldsymbol{\theta}} T &= \min_{\boldsymbol{\theta}} \|\hat{\varepsilon}\|^2 \\ &= \min_{\boldsymbol{\theta}} \hat{\varepsilon}'\hat{\varepsilon} \end{aligned}$$

considerando que $f(\mathbf{X}; \boldsymbol{\theta})$ é diferenciável com respeito a $\boldsymbol{\theta}$, pode-se escrever :

$$\mathbf{Z} = [Z_{ij}] = \frac{\partial f(X_i; \boldsymbol{\theta})}{\partial \theta_j} \quad ; \quad i = 1, \dots, n, j = 1, \dots, p \quad [16]$$

Do Cálculo sabe-se que uma condição para a existência do mínimo é :

$$\left. \frac{\partial T}{\partial \theta_j} \right|_{\hat{\theta}} = 0 \quad ; \quad j = 1, \dots, p$$

isto é :

$$Z' [Y - f(X; \theta)] = 0 \quad [17]$$

que, de acordo com SEBER & WILD (1989), é conhecido como Sistema de Equações Normais ou SEN.

Exemplo :

Forneça o sistema de equações normais para o modelo $Y_i = e^{-\theta X_i}$, $i = 1, \dots, n$

Solução :

$${}_n Z_1 = \frac{\partial f(X_i; \theta)}{\partial \theta_1} = \left[-X_i e^{-\hat{\theta} X_i} \right]$$

$$\text{SEN :} \quad {}_1 \left[-X_i e^{-\hat{\theta} X_i} \right]_n {}_n \left[Y_i - e^{-\hat{\theta} X_i} \right]_1 = 0$$

Efetuando o produto e organizando vem :

$$\sum_{i=1}^n X_i e^{-2 \hat{\theta} X_i} = \sum_{i=1}^n X_i Y_i e^{-\hat{\theta} X_i}$$

que é um sistema não-linear, e cuja solução $\hat{\theta}$, pode ser obtida por tentativas, ou por algum processo iterativo.

2.9. Linearização do Modelo Não-Linear pela Série de TAYLOR

Uma das técnicas utilizadas para resolver o problema dos modelos não-lineares é linearizar o modelo, expandindo $f(X; \theta)$ pela série de TAYLOR em torno do ponto $\theta_0 = [\theta_{10} \theta_{20} \dots \theta_{p0}]'$. Assim tem-se :

$$\begin{aligned}
f(\mathbf{X} ; \boldsymbol{\theta}) &= f(\mathbf{X} ; \boldsymbol{\theta}) + \left[\frac{\partial f(\mathbf{X} ; \boldsymbol{\theta})}{\partial \theta_1} \right]_{\theta_1 = \theta_{10}} (\theta_1 - \theta_{10}) + \\
&+ \left[\frac{\partial f(\mathbf{X} ; \boldsymbol{\theta})}{\partial \theta_2} \right]_{\theta_2 = \theta_{20}} (\theta_2 - \theta_{20}) + \\
&\dots \\
&+ \left[\frac{\partial f(\mathbf{X} ; \boldsymbol{\theta})}{\partial \theta_p} \right]_{\theta_p = \theta_{p0}} (\theta_p - \theta_{p0}) + \\
&+ (\text{termos de ordem superior a 2}) \quad [18]
\end{aligned}$$

que pode ser escrito como :

$$f(\mathbf{X} ; \boldsymbol{\theta}) \cong f_i^0 + B_1^0 Z_{1i}^0 + B_2^0 Z_{2i}^0 + \dots + B_p^0 Z_{pi}^0 \quad [19]$$

para $i = 1, \dots, n$

$$f(\mathbf{X} ; \boldsymbol{\theta}) \cong f_i^0 + \sum_{k=1}^p B_k^0 Z_{ki}^0$$

onde :

$$f_i^0 = f(\mathbf{X} ; \boldsymbol{\theta})$$

$$B_k^0 = \theta_k - \theta_{k0}$$

$$Z_{ki}^0 = \left. \frac{\partial f(\mathbf{X} ; \boldsymbol{\theta})}{\partial \theta_k} \right|_{\theta_k = \theta_{k0}}$$

Seja o modelo : $Y_i = f(\mathbf{X} ; \boldsymbol{\theta}) + \varepsilon_i$. Usando [19] pode-se escrever :

$$Y_i = f_i^0 + \sum_{k=1}^p B_k^0 Z_{ki}^0 + \varepsilon_i$$

Assim :

$$Y_i - f_i^0 = \sum_{k=1}^p B_k^0 Z_{ki}^0 + \varepsilon_i \quad [20]$$

é um modelo linear em B e pode ser facilmente resolvido.

De forma matricial o modelo é :

$$\mathbf{Y} - \mathbf{f}^0 = \mathbf{Z}^0 \mathbf{B}^0 + \boldsymbol{\varepsilon} \quad [21]$$

onde :

$$\begin{aligned} \underline{Y} - \underline{f}^0 &= \begin{bmatrix} Y_1 - f_1^0 \\ \vdots \\ Y_n - f_n^0 \end{bmatrix} \\ \underline{Z}^0 &= \begin{bmatrix} Z_{11}^0 & Z_{21}^0 & \dots & Z_{p1}^0 \\ Z_{12}^0 & Z_{22}^0 & \dots & Z_{p2}^0 \\ \vdots & \vdots & & \vdots \\ Z_{1n}^0 & Z_{2n}^0 & \dots & Z_{pn}^0 \end{bmatrix} \\ \underline{B}^0 &= \begin{bmatrix} B_1^0 \\ \vdots \\ B_p^0 \end{bmatrix} = \begin{bmatrix} \theta_1 - \theta_{10} \\ \vdots \\ \theta_p - \theta_{p0} \end{bmatrix} \\ \underline{\varepsilon} &= \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \end{aligned}$$

2.10. Método dos Mínimos Quadrados para a Solução do Modelo Linearizado pela Série de TAYLOR

Seja o modelo linearizado :

$$\underline{Y} - \underline{f}^0 = \underline{Z}^0 \underline{B}^0 + \underline{\varepsilon}$$

para o qual se deseja encontrar $\hat{\underline{B}}$, uma estimativa de $\underline{B}^0$, usando o Método dos Mínimos Quadrados.

Então :

$$\begin{aligned} \min L &= \min \|\underline{\varepsilon}\|^2 \\ &= \min \underline{\varepsilon}' \underline{\varepsilon} \\ &= \min [\underline{Y} - \underline{f}^0 - \underline{Z}^0 \underline{B}^0]' [\underline{Y} - \underline{f}^0 - \underline{Z}^0 \underline{B}^0] \\ &\quad \text{denotando } \underline{W} = \underline{Y} - \underline{f}^0, \text{ vem :} \\ &= \min [\underline{W} - \underline{Z}^0 \underline{B}^0]' [\underline{W} - \underline{Z}^0 \underline{B}^0] \\ &= \min \left[\underline{W}' \underline{W} - \underline{W}' \underline{Z}^0 \underline{B}^0 - \underline{B}^{0'} \underline{Z}^{0'} \underline{W} + \underline{B}^{0'} \underline{Z}^{0'} \underline{Z}^0 \underline{B}^0 \right] \end{aligned}$$

como $\tilde{W}' Z^0 B^0$ e $B^{0'} Z^{0'} \tilde{W}$ são escalares, vem :

$$= \min \left[\tilde{W}' \tilde{W} - 2 B^{0'} Z^{0'} \tilde{W} - B^{0'} Z^{0'} Z^0 B^0 \right]$$

Uma condição para que L tenha mínimo é :

$$\frac{\partial L}{\partial B^0} = 0$$

então :

$$- 2 Z^{0'} \tilde{W} - 2 Z^{0'} Z^0 B^0 = 0$$

de onde se conclui :

$$Z^{0'} Z^0 B^0 = Z^{0'} \tilde{W}$$

ou :

$$Z^{0'} Z^0 B^0 = Z^{0'} (\tilde{Y} - \tilde{f}^0) \quad [22]$$

que é conhecido como Sistema de Equações Normais (SEN) do modelo linearizado. Se $(Z^{0'} Z^0)$ é não singular, então, caso exista um mínimo relativo, uma estimativa de B^0 , é dada por:

$$\hat{B}^0 = (Z^{0'} Z^0)^{-1} Z^{0'} (\tilde{Y} - \tilde{f}^0) \quad [23]$$

Apenas a condição $\frac{\partial L}{\partial B^0} = 0$ não garante que L seja um ponto de mínimo. Outra condição necessária para garantir o ponto de mínimo é que a matriz $\frac{\partial^2 L}{\partial (B^0)^2}$ seja positiva definida.

2.11. Cálculo das Novas Estimativas do Vetor $\hat{\theta}$.

Se θ_0 é uma estimativa inicial do vetor $\hat{\theta}$, então uma estimativa melhor que θ_0 é dada por :

$$\theta_1 = \theta_0 + \hat{B}^0 \quad [24]$$

e o processo é repetido colocando θ_1 no lugar de θ_0 , até que algum critério de convergência seja aceito, isto é :

$$\theta_u \rightarrow \hat{\theta}$$

onde u é o número de vezes que o processo foi repetido.

2.12. Critérios de Convergência

Pode-se optar por um dos critérios dados a seguir :

2.12.1. Critério do Erro Relativo

DRAPER & SMITH (1981) sugerem :

$$\left| \frac{\theta_{k(u+1)} - \theta_{ku}}{\theta_{ku}} \right| < e \quad ; \quad k = 1, \dots, p \quad [25]$$

onde e é o limite superior do erro relativo entre $\theta_{k(u+1)}$ e θ_{ku} , sendo $(u + 1)$ a iteração atual e (u) a iteração anterior. Note o leitor que o critério deverá ser satisfeito para cada um dos parâmetros (θ_k) do modelo; nesse caso

$$\theta_k \cong \theta_{k(u+1)}$$

Se algum $\theta_k = 0$, então o critério passa a ser :

$$\left| \theta_{k(u+1)} - \theta_{ku} \right| < e \quad [26]$$

normalmente, para ambos casos, tem-se :

$$10^{-3} < e < 10^{-8}$$

2.12.2. Critério de MARQUARDT

MARQUARDT (1963) sugere :

$$\left| \frac{\theta_{k(u+1)} - \theta_{ku}}{\theta_{ku} + 10^{-3}} \right| < e \quad ; \quad k = 1, \dots, p \quad [27]$$

onde a constante 10^{-3} é utilizada para se evitar problemas quando $\theta_{ku} = 0$. As demais considerações são as mesmas de DRAPER e SMITH.

2.12.3. Critério de DENNIS & SCHNABEL

DENNIS & SCHNABEL (1983) sugerem que além de um dos critérios (DRAPER & SMITH ou MARQUARDT) também se teste o "critério do ponto de mínimo estacionário", definido como :

Se a soma de quadrados do residuo não decresce em 5 iterações consecutivas pare, (e aceite como solução o vetor θ_u) pois encontrou-se um ponto de mínimo estacionário.

2.13. Métodos de GAUSS-NEWTON para Modelos Não-Lineares

Descreve-se o modelo original (seção 2.13.1) e a seguir as modificações introduzidas por HARTLEY (1961, 1964) (seção 2.13.2) e por GALLANT (1985).

2.13.1. Método Original de GAUSS-NEWTON

Considere o modelo não-linear :

$$Y_i = f(X_i ; \theta) + \varepsilon_i$$

Usando a linearização pela série de TAYLOR (seção 2.9.), vem :

$$f(X_i ; \theta) = f(X_i ; \theta) + \sum_{k=1}^p \left[\frac{\partial f(X_i ; \theta)}{\partial \theta_k} \right]_{\theta_k = \theta_{ku}} \cdot (\theta_k - \theta_{ku})$$

onde os índices são definidos por :

$u = 0, 1, \dots$ (até a convergência)

$i = 1, \dots, n$

$k = 1, \dots, p$

Convencionou-se que :

$$f_i^u = f(X_i ; \theta_u)$$

$$B_k^u = \theta_k - \theta_{ku}$$

$$Z_{ki}^u = \left[\frac{\partial f(X_i ; \theta)}{\partial \theta_k} \right]_{\theta_k = \theta_{ku}}$$

Assim o modelo se torna linear :

$$Y_i - f_i^u = \sum_{k=1}^p B_k^u Z_{ki}^u + \epsilon_i$$

onde :

$$Z_u = [Z_{ki}^u] = \begin{bmatrix} Z_{11}^u & Z_{21}^u & \dots & Z_{p1}^u \\ Z_{12}^u & Z_{22}^u & \dots & Z_{p2}^u \\ \vdots & \vdots & & \vdots \\ Z_{1n}^u & Z_{2n}^u & \dots & Z_{pn}^u \end{bmatrix}_p \quad [28]$$

$$B_u = \begin{bmatrix} B_1^u \\ B_2^u \\ \vdots \\ B_p^u \end{bmatrix}_1 \quad [29]$$

$$Y_u = [Y_i - f_i^u] = \begin{bmatrix} Y_1 - f_1^u \\ Y_2 - f_2^u \\ \vdots \\ Y_n - f_n^u \end{bmatrix} \quad [30]$$

valor da variável dependente
valor da função para valores de θ_u

Dessa forma o Sistema de Equações Normais, SEN, (seção 2.10.) é dado por :

$$Z_u' Z_u B_u = Z_u' Y_u \quad [a]$$

cuja solução é :

$$\hat{B}_u = (Z_u' Z_u)^{-1} Z_u' Y_u \quad [31]$$

admitindo que $(Z_u' Z_u)$ seja positiva definida.

Dessa forma a próxima estimativa do vetor θ é dada por :

$$\theta_{u+1} = \theta_u + \hat{B}_u \quad [32]$$

e θ_0 é uma estimativa inicial fornecida pelo usuário (seção 2.21.).

Um critério de parada ou convergência (seção 2.12.) é dado por :

$$\text{Se } \left| \frac{\theta_{ku+1} - \theta_{ku}}{\theta_{ku}} \right| < e \quad ; k = 1, 2, \dots, p \quad [25]$$

então : pare (convergiu), senão repita-se novamente.

Se $\theta_{ku} = 0$ não deve-se usar [25], mas [26].

De [30] verifica-se que a Soma de Quadrados de Resíduo é :

$$SQRes = \| \chi_u \|^2$$

É bem sabido que o Método de GAUSS-NEWTON só é bom quando o problema a ser resolvido é de resíduo zero ou resíduo pequeno (DENNIS & SCHNABEL (1983), AL-BAALI (1984), SEBER & WILD (1989)).

2.13.1.1. Algoritmo de GAUSS-NEWTON

- 1 : Usuário fornece o vetor θ_0 das estimativas iniciais.
- 2 : $u = 0$
- 3 : Calcule [28] e [30].
- 4 : Calcule [31].
- 5 : Use [25] ou [26] para decidir :
 - se convergiu, pare.
 - senão, vá ao passo 6
- 6 : $u = u + 1$ e volte ao passo 3.

A única dificuldade que o leitor pode encontrar é na interpretação (montagem) das expressões [28] e [30], já que as demais são calculadas numericamente.

Exemplo:

Considere o modelo $\underline{Y} = f(\underline{X}; \underline{\theta}) + \varepsilon$, caracterizado por :

$$\underline{Y} = L \underline{1} + B \cdot e^{k\underline{X}} + \varepsilon$$

então :

$$\frac{\partial f}{\partial L} = \underline{1} \quad ; \quad \frac{\partial f}{\partial B} = e^{k\underline{X}} \quad ; \quad \frac{\partial f}{\partial K} = B \underline{X} e^{k\underline{X}}$$

Note que :

$$\frac{\partial f}{\partial L} = Z_{1i}^u$$

$$\frac{\partial f}{\partial B} = Z_{2i}^u$$

$$\frac{\partial f}{\partial K} = Z_{3i}^u$$

Dessa forma [28] é dado por :

$$\underline{Z}_u = [Z_{ki}^u] = \begin{bmatrix} 1 & e^{kx_1} & B X_1 e^{kx_1} \\ 1 & e^{kx_2} & B X_2 e^{kx_2} \\ \vdots & \vdots & \vdots \\ 1 & e^{kx_n} & B X_n e^{kx_n} \end{bmatrix}$$

onde a rigor deveria ser escrito : $e^{k_u x_1}$, $B_u X_1 e^{k_u x_1}$, etc...

A expressão [30] é :

$$\underline{X}_u = [Y_i - f_i^u] = \begin{bmatrix} Y_1 - f_1^u \\ Y_2 - f_2^u \\ \vdots \\ Y_n - f_n^u \end{bmatrix}$$

onde, $Y_i - f_i^u = Y_i - (L + B e^{kx_i})$. Maiores detalhes e exemplos numéricos podem ser encontrados em CHIACCHIO (1991).

2.13.1.2. Método de GAUSS-NEWTON para Modelos Não-Lineares Aplicado a um Modelo Linear

KENNEDY & GENTLE (1980), RATKOWSKY (1983) dentre outros, tem mostrado que o Método de GAUSS-NEWTON para modelos não-lineares quando aplicado a um modelo linear converge em uma única iteração, independente do valor inicial θ_0 .

Seja o modelo linear :

$$Y = X \theta + \varepsilon$$

onde :

${}_n Y_1$ é o vetor das variáveis dependentes.

${}_n X_p$ é a matriz das variáveis independentes, suposta conhecida e fixa.

${}_p \theta_1$ é o vetor de parâmetros que se quer estimar.

${}_n \varepsilon_1$ é o vetor de erros não observáveis, cujos elementos são supostos independentes e normalmente distribuídos.

Substituindo [30] e [31] em [32], vem :

$$\theta_{u+1} = \theta_u + (Z'_u Z_u)^{-1} Z'_u (Y - f^u) \quad [33]$$

Note que como o modelo é linear :

$$Z_u = \frac{\partial f}{\partial \theta} = X_u$$

e

$$f^u = X \theta$$

Então :

$$\begin{aligned} \theta_{u+1} &= \theta_u + (X'_u X_u)^{-1} X'_u (Y - X_u \theta_u) \\ &= \theta_u + (X'_u X_u)^{-1} X'_u Y - (X'_u X_u)^{-1} X'_u X_u \theta_u \\ &= \theta_u + (X'_u X_u)^{-1} X'_u Y - I \theta_u \\ &= \theta_u + (X'_u X_u)^{-1} X'_u Y - \theta_u \\ &= (X'_u X_u)^{-1} X'_u Y \end{aligned}$$

e o processo termina em uma única iteração ($u = 0$), já que $\hat{\theta} = (X' X)^{-1} X' Y$ é o estimador de mínimos quadrados de θ para o modelo linear $Y = X \theta + \varepsilon$.

2.13.2. Método de HARTLEY

O Método Original de GAUSS-NEWTON, embora simples na teoria, na prática é bastante deficiente. Para muitos modelos se a estimativa inicial estiver longe da solução o método não converge, isto é, não se encontra o ponto de mínimo da função $R(X; \theta)$.

HARTLEY (1961) propõe um método iterativo dado por :

$$\theta_{u+1} = \theta_u + v_u (Z'_u Z_u)^{-1} Z'_u (Y - f^u) \quad [34]$$

onde v_u é um escalar que minimiza localmente a função $R(X; \theta)$, conhecida como soma de quadrados de resíduos, no intervalo $[0; 1]$. Embora exista prova teórica da convergência do método, na prática isso não ocorre devido a erros numéricos (aritmética finita ou em ponto flutuante). Na estratégia original v_u é tomado como o ponto de mínimo da parábola que passa por $R(X; \theta(k))$, onde $k = 0, 1/2, 1$.

Denotando respectivamente por $R(0)$, $R(1/2)$ e $R(1)$ essas estimativas, então :

$$v_u = \frac{1}{2} + \frac{(1/4) [R(0) - R(1)]}{R(1) - 2 R(1/2) + R(0)} \quad [35]$$

HARTLEY (1964) afirma que v_u não deve ser calculado usando [35], e sugere uma estratégia mais simples, baseada na expressão :

$$(1/2)^t, \quad t = 0, 1, 2, 3, \dots \quad [36]$$

ressaltando que deve-se tomar para v_u primeiro valor para o qual :

$$R_{u+1}(X; \theta) < R_u(X; \theta)$$

ou seja, tenha havido uma diminuição na soma de quadrados de resíduos.

GALLANT (1985) sugere que [36] seja a sequência :

$$[1 \quad 0,9 \quad 0,8 \quad 0,7 \quad 0,6 \quad 1/2 \quad 1/4 \quad 1/8 \quad \dots] \quad [37]$$

tomando-se para v_u o primeiro valor para o qual tenha havido uma diminuição na soma de quadrados de resíduos.

O Método de HARTLEY só deve ser usado para problemas de resíduos zero ou resíduo pequeno.

Fato : A soma de quadrados de resíduos, $R(X; \theta)$, pode ser calculada pela expressão :

$$SQRes = \| \chi_u \|^2 \quad [38]$$

2.14. Método de MARQUARDT para Modelos Não-Lineares

A literatura afirma que o Método de MARQUARDT é o melhor de sua classe. Como existem diversas modificações, inicia-se a apresentação pelo original.

2.14.1. Método Original de MARQUARDT

MARQUARDT (1963) propõe um método baseado na interpolação entre dois métodos conhecidos na época, a saber :

- Método de GAUSS-NEWTON (ou "Método da Série de TAYLOR").
- Método do GRADIENTE (ou "Steepest Descent").

O Método de GAUSS-NEWTON é conhecido (seção 2.13.), e o Método do GRADIENTE (FLETCHER (1987)) é dado por :

$$\theta_{u+1} = \theta_u - Z'_u (Y - f^u) \quad [39]$$

A estratégia utilizada consiste na interpolação de $\hat{\beta}_t$ e $\hat{\beta}_g$, dados por :

$$\hat{\beta}_t = (Z_u' Z_u)^{-1} Z_u' (Y - f^u) \quad [40]$$

$$\hat{\beta}_g = - Z_u' (Y - f^u) \quad [41]$$

Note o leitor que [40] pode ser considerada como um "fator de correção" em [33]. Da mesma forma [41] é um fator de correção em [39].

A interpolação entre $\hat{\beta}_t$ e $\hat{\beta}_g$ (MARQUARDT (1963)) resulta em :

$$\hat{\beta} = (Z_u' Z_u + \lambda I)^{-1} Z_u' (Y - f^u) \quad [42]$$

onde λ é conhecida como constante de MARQUARDT.

Esse autor demonstra que

$$\hat{\beta} \rightarrow \hat{\beta}_t \text{ se } \lambda \rightarrow 0$$

$$\hat{\beta} \rightarrow \hat{\beta}_g \text{ se } \lambda \rightarrow \infty$$

Então o algoritmo de MARQUARDT é dado por :

$$\beta_{u+1} = \beta_u + (Z_u' Z_u + \lambda_u I)^{-1} Z_u' (Y - f^u) \quad [43]$$

e o problema agora reside na escolha de λ_u .

O seguinte fluxograma retrata a estratégia de MARQUARDT :

FIGURA 1. Fluxograma para o Método de MARQUARDT

Exemplos numéricos e um fluxograma completo do Método de MARQUARDT podem ser encontrados em CHIACCHIO (1991).

2.14.2. Método de MARQUARDT - Estratégia de BARD

BARD (1974) propõe que se substitua a matriz identidade $\mathbb{I}$ por uma matriz $\mathbb{D}$ onde :

$$\mathbb{D}_{kk} = \begin{cases} (\mathbf{Z}'_u \mathbf{Z}_u)_{kk} & \text{se } (\mathbf{Z}'_u \mathbf{Z}_u)_{kk} \neq 0 \\ 1 & \text{se } (\mathbf{Z}'_u \mathbf{Z}_u)_{kk} = 0 \end{cases} \quad [44]$$

A constante de MARQUARDT (λ) segue a mesma estratégia original.

Dessa forma o algoritmo de MARQUARDT-BARD é :

$$\hat{\theta}_{u+1} = \hat{\theta}_u + (\mathbf{Z}'_u \mathbf{Z}_u + \lambda_u \mathbb{D})^{-1} \mathbf{Z}'_u (\mathbf{y} - \hat{\mathbf{f}}^u) \quad [45]$$

2.14.3. Método de MARQUARDT - Estratégia de TABATA & ITO

TABATA & ITO (1975) sugerem que o valor $c = 10$ que incrementa ou decrementa λ , seja uma função do histórico das iterações, e não uma constante como propôs MARQUARDT.

Definiram : C_k , $k = 1, \dots, 5$

$$\begin{aligned} C_1 &= 1.33 \\ C_2 &= 1.78 \\ C_3 &= 3.16 \\ C_4 &= 10 \\ C_5 &= 100 \end{aligned} \quad [46]$$

cuja escolha é função do histórico das iterações :

Histórico	C_k	
partida	C_4	
DI, ID	C_{k-1}	
DDI, IDI, IID	C_{k-1}	se $k > 1$
DDD	C_{k+1}	se $k < 5$
III	C_3	se $k < 3$
outros casos	C_k	
D : Decréscimo na Soma de Quadrados de Resíduos		
I : Acréscimo na Soma de Quadrados de Resíduos		

FIGURA 2. Critérios para a Escolha da Constante de MARQUARDT, TABATA & ITO

Dessa forma o algoritmo de MARQUARDT-TABATA & ITO é :

$$\hat{\beta}_{u+1} = \hat{\beta}_u + (\mathbf{Z}'_u \mathbf{Z}_u + \lambda_u^* \mathbf{I})^{-1} \mathbf{Z}'_u (\mathbf{y} - \mathbf{f}^u) \quad [47]$$

onde λ_u^* denota $C_k \lambda_u$ ou λ_u / C_k .

O Método de MARQUARDT e suas modificações devem ser usadas apenas em problemas de resíduo zero ou resíduo pequeno.

2.15. Método de NEWTON para Modelos Não-Lineares

BARD (1974), BRODLIE (1977) , DENNIS & SCHNABEL (1983), dentre outros, consideram o modelo não-linear $\mathbf{y} = f(\mathbf{x} ; \boldsymbol{\theta}) + \varepsilon$, e citam que o Método de NEWTON é baseado na expansão pela Série de TAYLOR, até 2ª ordem, da função $f(\mathbf{x} ; \boldsymbol{\theta})$ em torno do ponto $\boldsymbol{\theta}_t$. Assim tem-se :

$$f(\mathbf{X} ; \boldsymbol{\theta}) \cong f_t(\mathbf{X} ; \boldsymbol{\theta}) + \boldsymbol{\beta}' \mathbf{g} + \frac{1}{2} \boldsymbol{\beta}' \mathbf{H} \boldsymbol{\beta} \quad [48]$$

com :

$${}_p \mathbf{g}_1 = 2\mathbf{Z}'(\mathbf{Y} - \mathbf{f}) \text{ onde } \mathbf{Z} \text{ é o jacobiano e } \mathbf{f} = f(\mathbf{X} ; \boldsymbol{\theta})$$

$${}_p \mathbf{H}_p = 2(\mathbf{Z}'\mathbf{Z} + \mathbf{S})$$

$${}_p \boldsymbol{\beta}_1 = \boldsymbol{\theta} - \boldsymbol{\theta}_t$$

onde ${}_p \mathbf{S}_p$ é o termo não-linear da matriz hessiana, dada por :

$$s_{km} = 2 \sum_{i=1}^n \left[(Y_i - f_i) \cdot \omega_{kmi} \right] \quad [49]$$

e

$$\omega_{kmi} = \frac{\partial^2 f(X_i ; \boldsymbol{\theta})}{\partial \theta_k \cdot \partial \theta_m} \quad [50]$$

note que a matriz ${}_n \Omega_{(p \times p)}$ tem elementos ω_{kmi} .

Assim o modelo linearizado pode ser escrito como :

$$\mathbf{Y} - \mathbf{f}_t = \boldsymbol{\beta}' \mathbf{g} + \frac{1}{2} \boldsymbol{\beta}' \mathbf{H} \boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad [51]$$

Aplicando o Método dos Mínimos Quadrados :

$$\min L = \min \|\boldsymbol{\varepsilon}\|^2$$

ocorre quando :

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \boldsymbol{\phi}$$

isto é

$$\mathbf{g} + \mathbf{H} \boldsymbol{\beta} = \boldsymbol{\phi}$$

se $\mathbf{H}$ é não singular, então :

$$\boldsymbol{\theta}_{u+1} = \boldsymbol{\theta}_u + (\mathbf{Z}'_u \mathbf{Z}_u + \mathbf{S}_u)^{-1} \mathbf{Z}'_u (\mathbf{Y} - \mathbf{f}^u) \quad [52]$$

é a forma iterativa do Método de NEWTON.

2.15.1. Obtenção do Termo Não-Linear "\$" do Hessiano

Considere o modelo não-linear $\tilde{Y} = f(X; \theta) + \varepsilon$,
 dado por :

$$\tilde{Y} = a_1 - bc^X + \varepsilon$$

onde particularmente :

${}_5\tilde{Y}_1$ é o vetor das observações dependentes,
 contendo 5 observações.

${}_5\tilde{X}_1$ é o vetor das observações independentes,
 contendo 5 observações.

${}_3\tilde{\theta}_1 = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ é o vetor de parâmetros
 desconhecidos, que se quer estimar.

${}_5\tilde{\varepsilon}_1$ é o vetor de erros aleatórios desconhecido
 e não observável.

Dessa forma tem-se :

$$\frac{\partial Y}{\partial a} = 1 \quad ; \quad \frac{\partial Y}{\partial b} = -c^x \quad ; \quad \frac{\partial Y}{\partial c} = -b.X.c^{x-1}$$

$$\frac{\partial^2 Y}{\partial a^2} = 0 \quad ; \quad \frac{\partial^2 Y}{\partial a \partial b} = 0 \quad ; \quad \frac{\partial^2 Y}{\partial a \partial c} = 0 \quad ; \quad \frac{\partial^2 Y}{\partial b^2} = 0$$

$$\frac{\partial^2 Y}{\partial b \partial c} = -Xc^{x-1} \quad ; \quad \frac{\partial^2 Y}{\partial c^2} = -bX(X-1)c^{x-2}$$

Então [50] é :

$$\omega_{kmi} = \begin{bmatrix} \frac{\partial^2 Y_i}{\partial a^2} & \frac{\partial^2 Y_i}{\partial a \partial b} & \frac{\partial^2 Y_i}{\partial a \partial c} \\ \frac{\partial^2 Y_i}{\partial b \partial a} & \frac{\partial^2 Y_i}{\partial b^2} & \frac{\partial^2 Y_i}{\partial b \partial c} \\ \frac{\partial^2 Y_i}{\partial c \partial a} & \frac{\partial^2 Y_i}{\partial c \partial b} & \frac{\partial^2 Y_i}{\partial c^2} \end{bmatrix}$$

11.
12.
13.

21.
22.
23.

31.
32.
33.

Note que:

$$\frac{\partial^2 Y_i}{\partial b \partial a} = \frac{\partial^2 Y_i}{\partial a \partial b}$$

$$\frac{\partial^2 Y_i}{\partial c \partial a} = \frac{\partial^2 Y_i}{\partial a \partial c}$$

$$\frac{\partial^2 Y_i}{\partial c \partial b} = \frac{\partial^2 Y_i}{\partial b \partial c}$$

Dessa forma a matriz $\Omega_{n \times p}$ é :

onde cada sub-matriz ω_{kmi} é calculada usando a i -ésima observação.

Logo a matriz S é dada por :

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{21} & S_{22} & S_{23} \\ S_{31} & S_{32} & S_{33} \end{bmatrix}_3$$

com :

$$S_{km} = \sum_{i=1}^5 \left[(Y_i - f_i) \cdot \omega_{kmi} \right] = (Y - f) \cdot [\Omega]$$

2.15.2. A Proposta de DENNIS, GAY & WELSCH (1981)

Note o leitor que o cálculo de $\mathcal{S}$, a parte não linear da matriz Hessiana $\mathcal{H}$, é extremamente trabalhosa, e indesejável. Além disso o Método de NEWTON frequentemente falha, pois a matriz Hessiana $\mathcal{H}$ se torna singular. Para solucionar o problema os autores propuseram que $\mathcal{S}$ seja estimada usando algum Método QUASI-NEWTON (seção 2.16.), e que o processo iterativo seja :

$$\varrho_{u+1} = \varrho_u + (\mathcal{Z}'_u \mathcal{Z}_u + \mathcal{S}_u + \lambda_u \mathbb{I})^{-1} \mathcal{Z}'_u (\mathcal{Y} - \mathcal{f}^u) \quad [53]$$

implementando também a estratégia de MARQUARDT ao Método de NEWTON, o que fornece uma solução mesmo que $\mathcal{H} = \mathcal{Z}'_u \mathcal{Z}_u + \mathcal{S}_u$ seja singular. Detalhes dessa implementação podem ser encontrados em DENNIS & SCHANABEL (1983) e AL-BAALI (1984).

O método de Newton e suas modificações devem ser usados em problemas de resíduos grandes.

2.16. Métodos QUASI-NEWTON

DAVIDON (1959), FLETCHER & POWELL (1963) apresentam um método para otimizar uma função genérica, com qualquer número de variáveis, conhecido atualmente como método DFP.

BROYDEN (1970), FLETCHER (1970), GOLDFARB (1970) e SHANNO (1970) propuseram independentemente uma nova aproximação, atualmente considerada o melhor Método QUASI-NEWTON, e conhecido como BFGS.

Como a teoria proposta é para otimização não-linear, isto é, minimização ou maximização de uma função qualquer, o caso da Regressão Não-Linear de Mínimos Quadrados será resolvido minimizando a função objetivo "Soma de Quadrados de Resíduos", isto é :

$$\min R(\mathcal{X} ; \varrho) = \min \| \varepsilon \|^2$$

Embora a teoria dos Métodos QUASI-NEWTON seja elegante, ela está fora do escopo desse trabalho. O leitor pode encontrar detalhes em FLETCHER (1987).

2.16.1. Regressão Não-Linear Usando os Métodos QUASI-NEWTON

Seja o modelo não-linear $\underline{Y} = f(\underline{X} ; \underline{\theta}) + \underline{\varepsilon}$,
onde :

${}_n\underline{Y}_1$ é o vetor das observações dependentes.

${}_n\underline{X}_v$ é a matriz das variáveis independentes.

${}_p\underline{\theta}_1$ é o vetor dos parâmetros desconhecidos, que se deseja estimar.

f é a relação funcional, suposta conhecida pelo pesquisador.

${}_n\underline{\varepsilon}_1$ é o vetor de erros aleatório, não observável, desconhecido, suposto ser normal e independentemente distribuído.

O objetivo é obter uma estimativa para o vetor $\underline{\theta}$, $\hat{\underline{\theta}}$, de forma que o erro $\underline{\varepsilon}$ seja mínimo.

Usando a teoria dos mínimos quadrados :

$$\min R(\underline{X} ; \underline{\theta}) = \min \|\underline{\varepsilon}\|^2$$

é dado, usando o Método de NEWTON, pelo processo iterativo :

$$\underline{\theta}_{u+1} = \underline{\theta}_u + (\underline{Z}'_u \underline{Z}_u + \underline{S}_u)^{-1} \underline{Z}'_u (\underline{Y} - \underline{f}^u) \quad [52]$$

Nos Métodos QUASI-NEWTON, essa expressão é dada por :

$$\underline{\theta}_{u+1} = \underline{\theta}_u + \underline{H}_u \underline{Z}'_u (\underline{Y} - \underline{f}^u) \quad [54]$$

onde $\underline{H}_u$ é uma aproximação para a matriz $(\underline{Z}'_u \underline{Z}_u + \underline{S}_u)^{-1}$, obtida a partir de qualquer matriz simétrica e positiva definida, $\underline{H}_0$, através de algum Método QUASI-NEWTON.

2.16.2. Algumas Considerações Teóricas e Práticas

Trabalhos teóricos (FLETCHER (1963), BROYDEN (1967), FLETCHER (1987), dentre outros) mostram que a melhor escolha para H_0 é a matriz identidade I , pois dessa forma, o processo se inicia com o Método do Gradiente (Steepest Descent), que é sabidamente bom quando se está "longe" da solução. Embora teoricamente se mostre que dada H_0 simétrica e positiva definida então H_u também é simétrica e positiva definida, qualquer u , POLAK (1971) sugere que H_u seja periodicamente "resetada", isto é, $H_u = I$. Seu argumento é baseado em testes práticos, que concluem que em aritmética finita, H_u se torna singular ; propõe que o "reset" deve ser dado quando u for múltiplo de 3 do número de parâmetros do modelo.

Deve ficar claro para o leitor que o algoritmo usado para resolver [54] depende :

- Da Equação QUASI-NEWTON escolhida (DFP, BFGS, ou qualquer outra).
- Do Método utilizado para resolver o sub-problema "busca unidimensional" (line-search).

O leitor pode optar por qualquer uma das equações da família QUASI-NEWTON proposta por BROYDEN (1967). Destacam-se, dentre as centenas existentes, as equações DFP e BFGS utilizadas nesse trabalho.

A parte mais trabalhosa do Método QUASI-NEWTON é sem dúvida o sub-problema "busca unidimensional", isto é, calcular um escalar $\alpha > 0$ tal que :

$$R(\mathbf{x} ; \alpha_u \cdot \mathbf{q}_{u+1}) < R(\mathbf{x} ; \mathbf{q}_u)$$

Segundo FLETCHER (1963) a 1ª sugestão foi dada por DAVIDON (1959); esse procedimento pode ser encontrado de forma detalhada em KENNEDY & GENTLE (1980). Outras sugestões são dadas por FLANAGAN, VITALE & MENDELSON (1969), FLETCHER (1980), AL-BAALI (1984), dentre outros.

Nesse trabalho utiliza-se a proposta de HARTLEY (1964), citada também por KENNEDY & GENTLE (1980), cuja expressão é simplesmente :

$$\alpha_k = \frac{1}{2^k} \quad , \quad k = 0, 1, 2, \dots \quad [55]$$

2.16.3. Algoritmo QUASI-NEWTON para Regressão Não-Linear

0) θ_0 : estimativa inicial dada pelo usuário

1) $u = 0$

2) $H_u = \begin{bmatrix} & \\ & \end{bmatrix}_{p \times p}$

3) Z_u ; equação [28]

χ_u ; equação [30]

$g_u = 2 Z_u' \chi_u$; conforme VITALE & TAYLOR

(1968)

$F_u = \| \chi_u \|^2$; equação [38]

4) $C_u = - H_u g_u$

5) $k = 0, 1, 2, \dots$ (Line-search)

$$\alpha_u = \frac{1}{2^k}$$

$$P_u = \alpha_u C_u \quad [56]$$

$$\theta_{u+1} = P_u + \theta_u$$

$$\chi_{u+1}$$

$$F_{u+1} = \| \chi_{u+1} \|^2$$

se $F_{u+1} \geq F_u$ voltar ao passo 5 com $k = k+1$

$$Z_{u+1}$$

$$g_{u+1} = 2 Z_{u+1}' \chi_{u+1}$$

$$Q_u = g_{u+1} - g_u \quad [57]$$

$$T = P_u' Q_u$$

$$\text{se } \left\{ \begin{array}{l} F_{u+1} < F_u \\ \text{e} \\ T > 0 \end{array} \right\} \quad \begin{array}{l} \text{aceita-se } \alpha_u \text{ e vá ao} \\ \text{passo 6} \end{array}$$

caso contrário voltar ao passo 5 com
 $k = k+1$

6) testar convergência ; se convergiu θ_{u+1} são os parâmetros procurados.

7) Reset de POLAK (1971)

$$\text{se } \left\{ \begin{array}{l} u > 0 \\ \text{e} \\ u \text{ é múltiplo de 3 do número} \\ \text{de parâmetros} \end{array} \right\}$$

então faça $u = u+1$ e volte ao passo 2

8) Calcule H_{u+1} pela expressão QUASI-NEWTON desejada.

9) $u = u+1$ e volte ao passo 4

2.16.4. Expressões QUASI-NEWTON Utilizadas Nesse Trabalho

2.16.4.1. DAVIDON, FLETCHER & POWELL (DFP)

$$H_{u+1}^{DFP} = H_u + \frac{P_u P'_u}{P'_u Q_u} - \frac{H_u Q_u Q'_u H'_u}{Q'_u H_u Q_u} \quad [58]$$

onde P_u e Q_u são definidos por [56] e [57] respectivamente.

2.16.4.2. BROYDEN, FLETCHER, GOLDFARB, SHANNO (BFGS)

$$H_{u+1}^{BFGS} = H_u + \left[1 + \frac{Q'_u H_u Q_u}{P'_u Q_u} \right] \frac{P_u P'_u}{P'_u Q_u} - \left[\frac{P_u Q'_u H_u + H_u Q_u P'_u}{P'_u Q_u} \right] \quad [59]$$

Os Métodos QUASI-NEWTON devem ser utilizados em problemas de resíduos grandes.

2.17. Métodos Híbridos para Regressão Não-Linear

2.17.1. Introdução

GILL & MURRAY (1972), citados por FLETCHER (1987), sugerem que para melhorar a estabilidade numérica dos Métodos QUASI-NEWTON seja usada a matriz $B_u = H_u^{-1}$ onde :

$$B_u = Z'_u Z_u \quad [60]$$

Dessa forma o Método QUASI-NEWTON passa a ser :

- a) resolva $B_u \zeta_u = -g_u$ para ζ_u ; note que esse passo deve ser feito usando o algoritmo de GILL & MURRAY (1972), para se evitar a inversão da matriz B_u .
- b) Calcule o escalar α , tal que :
 $R(X; \varrho_u + \alpha \zeta_u) < R(X; \varrho_u)$
isto é, o "Line Search". Note que
 $\varrho_{u+1} = \varrho_u + \alpha \zeta_u$.
- c) Obtenha B_{u+1} a partir de B_u usando a

expressão QUASI-NEWTON corresponde à aproximação desejada (DFP ou BFGS).

Repete-se os passos até que a convergência seja obtida.

2.17.1.1. Expressões QUASI-NEWTON no Contexto de GILL & MURRAY

2.17.1.1.1. DAVIDON, FLETCHER & POWELL (DFP)

$$\mathbb{B}_{u+1}^{\text{DFP}} = \mathbb{B}_u + \left[1 + \frac{\tilde{\mathbf{P}}'_u \mathbb{B}_u \tilde{\mathbf{P}}_u}{\mathbf{Q}'_u \tilde{\mathbf{P}}_u} \right] \frac{\mathbf{Q}_u \mathbf{Q}'_u}{\mathbf{Q}'_u \tilde{\mathbf{P}}_u} - \left[\frac{\mathbf{Q}_u \tilde{\mathbf{P}}'_u \mathbb{B}_u + \mathbb{B}_u \tilde{\mathbf{P}}_u \mathbf{Q}'_u}{\mathbf{Q}'_u \tilde{\mathbf{P}}_u} \right] \quad [61]$$

2.17.1.1.2. BROYDEN, FLETCHER, GOLDFARB, SHANNO (BFGS)

$$\mathbb{B}_{u+1}^{\text{BFGS}} = \mathbb{B}_u + \frac{\mathbf{Q}_u \mathbf{Q}'_u}{\mathbf{Q}'_u \tilde{\mathbf{P}}_u} - \frac{\mathbb{B}_u \tilde{\mathbf{P}}_u \tilde{\mathbf{P}}'_u \mathbb{B}'_u}{\tilde{\mathbf{P}}'_u \mathbb{B}_u \tilde{\mathbf{P}}_u} \quad [62]$$

2.17.2. Método de AL-BAALI

É bem sabido, (DENNIS & SCHANABEL (1983), AL-BAALI (1984), SEBER & WILD (1989)), que os Métodos de GAUSS-NEWTON, MARQUARDT (e suas modificações) são bons quando o problema a ser resolvido é considerado de resíduo

zero ou de residuo pequeno. Se o problema a ser resolvido é de resíduos grandes, os métodos mais adequados são os QUASI-NEWTON.

Isso motivou AL-BAALI a pesquisar métodos híbridos. Dentre os métodos pesquisados considerou o melhor, denominado de HBY2, hoje conhecido como HY2 (XU (1986)).

2.17.2.1. Algoritmo do Método Híbrido HBY2

Se problema é de resíduos grandes
então use Métodos QUASI-NEWTON
senão use Método GAUSS-NEWTON.
Fim.

Esse método será detalhado mais adiante (seção 2.17.3.2.).

2.17.3. Método de XU

2.17.3.1. Introdução

XU (1986) mostrou que o algoritmo HBY2 proposto por AL-BAALI não preserva algumas das propriedades do método BFGS ; provou que o problema se encontra na expressão usada para decidir o tipo de método a ser usado, QN ou GN, e propôs um novo teste (T_u) para a decisão.

Embora conjecturas teóricas mostrem que um novo algoritmo, HY3, deva ser usado, conclui que pela simplicidade de HY2 este deve ser o preferido.

Note que a diferença entre HBY2 e HY2 reside apenas na expressão usada para se decidir se o problema é de resíduos grandes ou não, isto é :

$$T_u \rightarrow \begin{cases} 0 & \Rightarrow \text{problema de residuos grandes} \\ 1 & \Rightarrow \text{problema de residuo zero.} \end{cases}$$

2.17.3.2. Algoritmo do Método HY2

- 1) $u = 0$: escolha θ_0
- 2) calcule $Z'_u Z_u$; se for positiva definida,
faça $B_u = Z'_u Z_u$
- 3) resolva $B_u C_u = -Z'_u (Y - f^u)$ para C_u
- 4) obtenha α_u tal que
 $[R(X; \theta_u + \alpha_u C_u) < R(X; \theta_u)]$;
 calcule $\theta_{u+1} = \theta_u + \alpha_u C_u$
- 5) obtenha Z_{u+1} usando a expressão de BROYDEN
 (1965) ; $B_{u+1} = Z'_{u+1} Z_{u+1}$
- 6) calcule $T_u = \frac{\|Y - f^u\|^2 - \|Y - f^{u+1}\|^2}{\|Y - f^u\|^2}$
- 7) $u = u + 1$
- 8) resolva $B_u C_u = -Z'_u (Y - f^u)$ para C_u
- 9) obtenha α_u tal que
 $R(X; \theta_u + \alpha_u C_u) < R(X; \theta_u)$
- 10) se $T_{u+1} \geq 0,2$ (onde 0,2 serve para evitar
 dificuldades)
 então obtenha Z_{u+1} usando a expressão de
 BROYDEN (1965)
 senão calcule Z_{u+1} através de derivadas.
- 11) se $T_{u+1} < 0,2$
 então obtenha B_{u+1} usando o Método BFGS

$$\text{senão } B_{u+1} = Z'_{u+1} Z_{u+1} \quad (\text{GN})$$

12) volte ao passo 6

2.17.3.2.1. Expressão de BROYDEN (1965)

$$Z_{u+1} = Z_u + \frac{(\varrho_{u+1} - \varrho_u) \cdot [(\underline{y} - \underline{f}^{u+1}) - (\underline{y} - \underline{f}^u) - \dots - Z'_u(\varrho_{u+1} - \varrho_u)]'}{(\varrho_{u+1} - \varrho_u)' (\varrho_{u+1} - \varrho_u)} \quad [63]$$

2.18. Escolha dos Métodos Dentre os Estudados

Como não é possível saber de antemão o tipo de problema que se tem, e partindo-se do pressuposto que estejam disponíveis todos métodos apresentados, sugere-se :

- Método Híbrido (XU)
- Método NEWTON modificado (DENNIS, GAY & WELSCH)
- Método QUASI-NEWTON (BFGS)
- Método de MARQUARDT (original).

Nesse trabalho o usuário poderá optar, dentro da teoria dos mínimos quadrados, pelos seguintes métodos :

- Método de GAUSS-NEWTON (original)
- Método de MARQUARDT (original)
- Métodos QUASI-NEWTON
 - . DFP
 - . BFGS.

A opção GAUSS-NEWTON é justificada pelo fato do usuário poder resolver modelos lineares em apenas uma iteração.

2.18.1. Considerações Práticas

SEBER & WILD (1989) tem feito algumas considerações práticas, admitindo o problema da adequação "modelo-dados" resolvido, expostos a seguir :

- Se o usuário sabe de antemão (baseado em experiências semelhantes) que tipo de problema tem, pode resolver utilizando o método mais adequado.
- Se o usuário não tem experiência alguma deve começar por MARQUARDT.
- se o método de Marquardt converge com λ pequeno então o problema está resolvido, a menos que o modelo não seja estatisticamente adequado.
- se há convergência com λ grande (> 0.1) o usuário deverá verificar a adequação do modelo estatisticamente (ANOVA por exemplo). Pode ser possível melhorar as estimativas usando algum Método QUASI-NEWTON.

Nesses casos o problema pode ser considerado de resíduo zero ou resíduo pequeno.

- Se o método falha ou λ oscila muito sem convergir para valor algum, o problema pode ser de resíduos grandes. Sugere-se nesse caso o uso de Métodos QUASI-NEWTON.
- Se o usuário supõe (baseado em experiências semelhantes) que o modelo é de resíduos grandes, deve começar por algum Método QUASI-NEWTON, sugere-se o BFGS.

2.19. Método dos Mínimos Quadrados Generalizados

É utilizado quando os erros são correlacionados e/ou heterocedásticos.

Assim :

$$\varepsilon \sim N(\phi ; W \sigma^2)$$

onde W é uma matriz simétrica, positiva definida e conhecida.

SEBER & WILD (1989) dentre outros, tem mostrado que o sistema de equações normais para os Métodos dos Mínimos Quadrados Generalizados é dado por :

2.19.1. Método Generalizado de GAUSS-NEWTON

$$\hat{\theta}_{u+1} = \hat{\theta}_u + (Z'_u W^{-1} Z_u)^{-1} Z'_u W^{-1} (Y - \hat{f}^u) \quad [64]$$

e por analogia :

2.19.2. Método Generalizado de MARQUARDT

$$\hat{\theta}_{u+1} = \hat{\theta}_u + (Z'_u W^{-1} Z_u + \lambda I)^{-1} Z'_u W^{-1} (Y - \hat{f}^u) \quad [65]$$

2.19.3. Método QUASI-NEWTON Generalizado

$$\hat{\theta}_{u+1} = \hat{\theta}_u + H_u Z'_u W^{-1} (Y - \hat{f}^u) \quad [66]$$

2.20. Método dos Mínimos Quadrados Ponderados

É utilizado quando os erros são heterocedásticos, isto é :

$$\varepsilon \sim N(\phi ; D\sigma^2)$$

onde $\mathbb{D}$ é uma matriz diagonal positiva definida.

É fácil verificar que o Método dos Mínimos Quadrados Ponderados é um caso particular do Método dos Mínimos Quadrados Generalizados.

Substituindo $\mathbb{W}^{-1}$ por $\mathbb{D}^{-1}$ em [64], [65], [66] tem-se os Métodos Ponderados de GAUSS-NEWTON, MARQUARDT e QUASI-NEWTON, respectivamente.

2.21. O Problema da Estimativa Inicial (θ_0)

É bem sabido que qualquer método utilizado para regressão não-linear necessita de uma boa estimativa inicial θ_0 para que o processo convirja para a estimativa de mínimos quadrados, $\hat{\theta}$.

O valor da estimativa inicial pode ser conhecida :

- por estar ligada ao desenvolvimento teórico do processo ;
- por experiências semelhantes ;

pode ser ainda desconhecida e/ou totalmente empírica.

Algumas técnicas para a obtenção de valores iniciais são mostradas a seguir, sugerindo ao leitor que se um determinado método falhar, ele deve tentar outros até conseguir um que seja satisfatório. Se todos métodos falharem, provavelmente o conjunto modelo/dados é inadequado.

2.21.1. Método do Modelo Aproximado

Consiste em transformar o modelo não-linear em um modelo linear aproximado, do qual se obtém estimativas, que são usadas como "chute" inicial para o modelo não-linear.

Considere por exemplo o modelo não-linear $Y = x^\theta + \varepsilon$, que pode ser transformado no modelo aproximado $Y = x^\theta \cdot \varepsilon$; assim:

$$\ln Y = \theta \ln x + \ln \varepsilon$$

é um modelo linear e a estimativa de θ no modelo linear será usada como θ_0 no Modelo Não-Linear.

Esse método foi usado por FIGUEIRÔA (1986) para ajustar o Modelo de COBB-DOUGLAS.

2.21.2. Métodos de HARTLEY & BROOKER (1965)

Consiste em transformar o problema dado em um problema de residuo zero, dividindo o conjunto de n observações em um conjunto de p observações, onde p é o número de parâmetros do modelo, e cada observação p é a média aritmética das k observações da divisão de n por p .

Exemplo: Suponha (X_i, Y_i) , $i = 1, 2, \dots, 20$ e $p = 3$; então uma alternativa é:

X_1 = média aritmética de $(X_1, \dots, X_6)$; $n = 6$

X_2 = média aritmética de $(X_7, \dots, X_{12})$

X_3 = média aritmética de $(X_{13}, \dots, X_{20})$; $n = 8$

Idem para Y .

Dessa forma tem-se agora um problema de residuo zero, (p equações a p incógnitas) que pode ser resolvido de forma satisfatória usando-se MARQUARDT, com algum "chute" inicial. A estimativa $\hat{\theta}$ obtida dessa forma é θ_0 para o problema inicial.

2.21.3. Método de GALLANT (1985)

É um caso particular do Método de HARTLEY & BROOKER (1965), onde o usuário escolhe dentro da amostra de n pontos, p pontos, e o modelo é transformado em um problema de residuo zero, cuja estimativa $\hat{\theta}$ é usada como θ_0 do problema inicial.

Obviamente que os p pontos escolhidos devem ser representativos da amostra (conjunto de dados).

2.21.4. Método da "MALHA"

Consiste em considerar uma rede (malha) de pontos no espaço dos parâmetros, tomando para valor inicial θ_0 o ponto da rede para o qual o valor da função objetivo ("Soma de Quadrados de Resíduos") é mínimo (citado no manual do SAS, procedimento NLIN).

RATKOWSKY (1983) sugere métodos para obtenção de estimativas iniciais para alguns "modelos de crescimento", muito usados em pesquisa agropecuária.

2.22. Introdução às Medidas de Não-Linearidade

GUTTMAN & MEETER (1965), citam que *"as medidas de não-linearidade são expressões que indicam se o grau de não-linearidade em um modelo de regressão não-linear, é pequeno o suficiente para justificar a utilização dos resultados usuais da teoria dos modelos lineares como aproximação para os não-lineares."*

BEALE (1960) foi o pioneiro a questionar essa validade, propondo a "medida de não-linearidade intrínseca do modelo", isto é, medida que indica se a curvatura da superfície de estimação é acentuada ou não nas proximidades do mínimo.

DRAPER & SMITH (1981) citam que no caso da regressão não-linear as propriedades das estimativas dos parâmetros são válidas apenas assintoticamente, isto é, para grandes amostras; afirmam que as inferências sobre o vetor de parâmetros (usando técnicas de regressão linear) apresentaram resultados apenas aproximados, cuja validade depende do "grau de não-linearidade do modelo" ser pequeno.

BATES & WATTS (1988) estudam as medidas de não-linearidade do modelo através de conceitos geométricos. Citam que a validade das inferências aproximadas (por exemplo, intervalo de confiança para parâmetros), dependem da superfície de estimação ser quase plana (medida de não-linearidade intrínseca) e possuir coordenadas quase uniformes (medida de não-linearidade de efeitos de parâmetros), que são as propriedades fundamentais do plano de estimação no caso linear.

Resumindo, conclui-se que se as medidas de não-linearidade são baixas, então o comportamento do modelo não-linear na estimação se aproxima do comportamento dos modelos lineares, isto é, os estimadores de mínimos quadrados $\hat{\theta}$ serão "quase não viesados", "quase normalmente distribuídos" e as variâncias dos estimadores serão próximos daqueles obtidos pela matriz de variâncias/covariâncias assintóticas $(Z_u' Z_u)^{-1} \sigma^2$, isto é :

$$\hat{\theta} \sim N \left[\theta ; (Z_u' Z_u)^{-1} \sigma^2 \right]$$

RATKOWSKY (1983) apresenta um exemplo didático sobre medidas de não-linearidade. Como o procedimento para obtenção dessas medidas não é trivial (envolve matrizes tridimensionais de derivadas de 2ª ordem do modelo) fornece um programa escrito em FORTRAN para obtê-las. Espera-se que em um futuro próximo existam expressões alternativas a essas medidas de não-linearidade.

2.23. Análise de Variâncias e Teste de Hipótese Sobre o Modelo

DRAPER & SMITH (1981) afirmam que no caso da regressão não-linear, a análise de variâncias não é uma ferramenta útil, já que não se deve concluir a partir do teste F, se o modelo é adequado ou não.

Usar-se-á nesse trabalho a Análise de Variâncias (ANOVA) como um teste de significância aproximado, ou como uma medida de comparação entre modelos.

Assim o quadro de análise é:

Causa de variação	gl	SQ	Q	F
Regressão	p	SQRegr	QMRegr	F _c
Resíduo	n-p	SQRes	QMRes	—
Total	n	SQTo	—	—

onde :

p : é o número de parâmetros do modelo.

n : é o número total de observações.

SQTo : é a soma de quadrados total sem a correção.

SQRes : é a soma de quadrados dos resíduos (erros).

SQRegr : é a soma dos quadrados da regressão.

QMRes : é o quadrado médio dos resíduos, que estima a variância dos erros σ^2 .

F_c : é a estatística que testa (aproximadamente) as hipóteses :

H₀ : Parâmetros tais que E[χ] = 0 [67]

H₁ : Sem restrição nos parâmetros

e de forma usual :

$$SQTo = \| \chi' \chi \|^2 \quad [68]$$

$$SQRes = \| \chi - X \hat{\theta} \|^2 = \| \hat{\varepsilon} \|^2 \quad [69]$$

$$SQRegr = SQTo - SQRes \quad [70]$$

$$QMRegr = \frac{SQRegr}{p} \quad [71]$$

$$QMRes = \frac{SQRes}{n-p} \quad [72]$$

$$F_c = \frac{QMRegr}{QMRes} \quad [73]$$

com o teste de hipóteses :

$$\text{se : } F_0 > F_{\alpha, p, n-n.1-\alpha}$$

então : rejeita-se H_0 , isto é, aceita-se H_1 e conclui-se que o "modelo é adequado", ou que representa razoavelmente bem o conjunto de dados:

senão : aceita-se H_0 , isto é, conclui-se que o "modelo é inadequado".

Da tabela de análise de variâncias, pode-se também obter o "coeficiente de determinação" :

$$R^2 = \frac{SQRegr}{SQT_0} \quad [74]$$

que indica que a regressão explica $(100.R^2)\%$ das variações, isto é, os valores de Y são explicados em $(100.R^2)\%$ pelos valores de X .

Para que se possa comparar o "coeficiente de determinação" R^2 de modelos de regressão distintos faz-se necessária uma correção em R^2 , de tal forma que a nova medida seja independente do número de parâmetros dos modelos a ser comparados.

Essa nova medida é definida como "coeficiente de determinação ajustado" (DRAPER & SMITH (1981)), ou como "coeficiente de comparação" (KOLB (1982)), e é dada por :

$$RR = 1 - \frac{(1-R^2).(n-1)}{(n-p)} \quad [75]$$

2.24. Matriz de Variâncias e Covariâncias Assintóticas

DRAPER & SMITH (1981), DONALDSON & SCHNABEL (1987) sugerem que a matriz de variâncias e covariâncias assintótica seja expressa por :

$$V = (Z'_u Z_u)^{-1} \sigma^2 \quad [76]$$

onde a matriz $\mathbf{Z}_u$, definida em [28], é aquela para a qual se obteve a convergência.

Se o método utilizado é QUASI-NEWTON, então

$$\hat{\mathbf{V}} = \mathbf{H}_u \sigma^2 \quad [77]$$

onde a matriz $\mathbf{H}_u$, definida em [54], é aquela para a qual se obteve a convergência.

Uma estimativa para as matrizes de variância e covariâncias assintóticas, $\hat{\mathbf{V}}$, pode ser obtida usando o QMRes como uma estimativa para σ^2 . Assim :

$$\hat{\mathbf{V}} = \text{QMRes.} (\mathbf{Z}'_u \mathbf{Z}_u)^{-1} \quad [78]$$

ou

$$\hat{\mathbf{V}} = \text{QMRes.} \mathbf{H}_u \quad [79]$$

2.25. Intervalo de Confiança para os Parâmetros

DONALDSON & SCHNABEL (1987) concluíram que se as medidas de não-linearidade são baixas, então o intervalo a $100(1 - \alpha)\%$ de confiança para os parâmetros $\boldsymbol{\theta}$, em analogia aos modelos lineares, é dado por:

$$\hat{\theta}_k \pm t_{(n-p; \alpha/2)} \cdot (\hat{\mathbf{V}}_{kk})^{1/2}, \quad k = 1, 2, \dots, p \quad [80]$$

onde :

$\hat{\boldsymbol{\theta}}$ é a estimativa do vetor de parâmetros $\boldsymbol{\theta}$.

$\hat{\mathbf{V}}_{kk}$ são os elementos da diagonal da matriz de variâncias e covariâncias assintóticas.

$t_{(n-p; \alpha/2)}$ é o valor da distribuição t de student, com $n-p$ graus de liberdade e tal que :

$$P [|t| \leq t_{(n-p; \alpha/2)}] = 1 - \alpha.$$

n é o número total de observações.

p é o número de parâmetros do modelo.

2.26. Matriz de Correlação Assintótica dos Parâmetros

DRAPER & SMITH (1981) citam que a matriz de correlação assintótica dos parâmetros é dada por:

$$C = [C_{ij}] = \frac{Q_{ij}}{(Q_{ii} \cdot Q_{jj})^{1/2}} \quad [81]$$

onde :

$Q = (Z_u' Z_u)^{-1}$ com Z_u definida em [28] e tomada quando a estimativa $\hat{\theta}$ dos parâmetros é obtida.

ou

$Q = H_u$ definida em [54] e tomada na convergência do processo QUASI-NEWTON.

A matriz de correlação assintótica dos parâmetros serve para verificar se o modelo é apropriado ou se necessita de alguma reparametrização. Se C_{ij} é grande (tende para 1), indica que o modelo está superametrizado e que o usuário pode tentar uma reparametrização (retirada de 1 ou mais parâmetros).

2.27. Análise dos Resíduos

A análise dos resíduos tem por objetivo verificar se as pressuposições a respeito do vetor de erros (seção 2.5) não foram violadas, e também detectar pontos influentes e/ou discrepantes.

Uma ou mais observações são ditas discrepantes (outliers) se seus resíduos são muito grandes (em valor absoluto) em relação aos demais; podem se distanciar da média zero dos resíduos em 3, 4 ou mais desvios padrões (DRAPER & SMITH (1966)).

Pontos influentes são observações que, embora não apresentem resíduos grandes, podem alterar significati-

vamente os parâmetros do modelo escolhido.

É bem sabido que o vetor de erros, $\underline{\varepsilon}$, é desconhecido ; uma estimativa, $\hat{\underline{\varepsilon}}$, para esse vetor é dada por :

$$\hat{\underline{\varepsilon}} = \underline{Y} - \hat{\underline{Y}} \quad [82]$$

onde $\underline{Y}$ são os valores observados e $\hat{\underline{Y}}$ são as estimativas de $\underline{Y}$, obtidos através do modelo ajustado, isto é,

$$\hat{\underline{Y}} = f(\underline{X} ; \hat{\underline{\theta}}) \quad [83]$$

Dessa forma o vetor $\hat{\underline{\varepsilon}}$ pode ser considerado como uma quantidade que o modelo ajustado não consegue explicar.

Assim, se o modelo é adequado admite-se que o vetor $\hat{\underline{\varepsilon}}$ represente o verdadeiro vetor de erros $\underline{\varepsilon}$, desconhecido.

Obviamente espera-se que o vetor $\hat{\underline{\varepsilon}}$ tenha componentes próximos de zero, isto é, que o modelo reproduza os dados originais com o menor erro possível.

DRAPER & SMITH (1981) citam que considerando as pressuposições a respeito do vetor de erros $\underline{\varepsilon}$ (seção 2.5.), pode-se escrever :

$$\underline{\varepsilon} \sim N(\underline{\phi} ; \mathbb{I} \sigma^2) \quad [84]$$

COOK & WEISBERG (1982), dentre outros, consideram que a estimativa do erro, $\hat{\underline{\varepsilon}}$, pode estar contida em qualquer intervalo, e sugerem uma normalização no vetor $\hat{\underline{\varepsilon}}$, isto é :

$$\hat{\underline{\varepsilon}}_N = \frac{\hat{\underline{\varepsilon}}}{\sqrt{\text{QMRes}}} \quad [85]$$

dessa forma $\hat{\underline{\varepsilon}}_N \sim N(\underline{\phi} ; \mathbb{I})$ que restringe $\hat{\underline{\varepsilon}}_N$ probabilisticamente ao intervalo $[-5 ; 5]$, mais especificamente, pode-se esperar que 95% dos valores de $\hat{\underline{\varepsilon}}_N$ estejam no intervalo $[-1,96 \leq \hat{\underline{\varepsilon}}_N \leq 1,96]$, isto é :

$$P [-1,96 \leq \hat{\underline{\varepsilon}}_N \leq 1,96] = 0,95 \quad [86]$$

onde $\hat{\varepsilon}_N$ é conhecido como "vetor de erros normalizados".

Embora existam outros tipos de erros (estudentizados internamente, studentizados externamente, DFFITS, etc...) eles não serão estudados neste trabalho. Detalhes podem ser encontrados em COOK & WEISBERG (1982), MACHADO (1991), dentre outros.

2.27.1. Análise Gráfica

É o instrumento mais útil da análise de resíduos pois pode dar indicações diretas de quais das pressuposições, a respeito dos erros, foram violadas. Indica também a presença de pontos influentes e/ou discrepantes.

DRAPER & SMITH (1981) apresentam diversas alternativas para "plotar" os erros (histograma, contra cada variável independente, contra a variável dependente, etc..). Concluem que a melhor é "plotar" $\hat{\varepsilon}_N$ versus $\hat{Y}$, pois não há correlação entre $\hat{\varepsilon}_N$ e $\hat{Y}$, isto é :

$$\text{Cov} [\hat{\varepsilon}_N ; \hat{Y}] = \phi \quad [87]$$

além disso,

$$\hat{\varepsilon}_N' \hat{Y} = \phi \quad [88]$$

matematicamente.

NETER & WASSERMAN (1974), DRAPER & SMITH (1981), MACHADO (1991) dentre outros, apresentam formas características dos gráficos dos resíduos.

Apresenta-se a seguir alguns desses gráficos.

FIGURA 3. Gráficos dos Resíduos.

2.27.2. Análise Estatística

Embora a análise gráfica seja simples, na prática ocorrem dificuldades, pois as conclusões obtidas, pelo julgamento de diferentes observadores, podem ser distintas para o mesmo conjunto de dados.

2.27.2.1. Teste de CHORRILHOS

É uma estatística para testar se os erros são aleatórios, isto é, testa-se a hipótese :

H_0 : os erros são aleatórios. [89]

H_1 : os erros não são aleatórios.

Se o teste concluir que os erros não são aleatórios, o usuário deve, dentre outras opções, selecionar outro modelo (os erros devem ser aleatórios para que o modelo seja aceitável).

FIGUEIREDO (1977) apresenta o "Teste de CHORRILHOS", cuja sequência é :

- calcula-se o vetor de erros $\hat{\varepsilon}$;
 - ordena-se o vetor $\hat{\varepsilon}$ em ordem crescente ;
 - calcula-se a mediana do vetor $\hat{\varepsilon}$.
- se n (número de observações) é ímpar, a mediana está na posição $\frac{n+1}{2}$;
- se n é par, a mediana é a média aritmética entre os valores $n/2$ e seu sucessor ;
- substituir os valores dos erros por 2 símbolos quaisquer, tal que :
 - 1º símbolo para os valores menores que a mediana;
 - 2º símbolo para valores maiores ou iguais a mediana.

Nota : esse passo deve ser feito no vetor original $\hat{\varepsilon}$ (não no vetor ordenado).

- contar o número de símbolos :

n_1 - número de símbolos do 1º tipo

n_2 - número de símbolos do 2º tipo

- calcular o número total de chorrilhos, r .

Fato : Chorrilho é definido como uma sucessão de símbolos idênticos, os quais são seguidos ou precedidos por um símbolo diferente ou nenhum símbolo.

Exemplo :

-	-	+	-	-	+	+	+	+	+	-	-
---	---	---	---	---	---	---	---	---	---	---	---

$$n_1 = 6 \text{ (-)}$$

$$n_2 = 6 \text{ (+)}$$

$$r = 5 \left[\boxed{} \right]$$

- se n_1 ou n_2 são maiores que 20, calcula-se :

$$Z = \frac{r - \left[\frac{2 n_1 n_2}{n_1 + n_2} + 1 \right]}{\sqrt{\frac{2 n_1 n_2 (2 n_1 n_2 - n_1 - n_2)}{(n_1 + n_2)^2 \cdot (n_1 + n_2 - 1)}}} \quad [90]$$

- se n_1 e n_2 são menores que 20, usando a Tabela do ANEXO 4, obtenha r_{inf} e r_{sup} .

- Decisão :

Rejeita-se H_0 se $|z| \geq 1,96$

ou

Rejeita-se H_0 se $r \leq r_{inf}$ ou $r \geq r_{sup}$.

2.27.2.2. Teste de LILLIEFORS

É uma estatística para testar se a distribuição dos erros é normal (pressuposição IV), isto é, testa-se a hipótese :

H_0 : os erros tem distribuição normal. [91]

H_1 : os erros não tem distribuição normal.

Se o teste concluir que os erros não tem distribuição normal, o usuário deve, dentre outras opções, tentar alguma transformação de variáveis.

CAMPOS (1983) descreve o teste de LILLIEFORS (1967) que transcreve-se a seguir :

- calcula-se o vetor de erros $\hat{\varepsilon}$, com n observações.
- calcula-se a média e o desvio padrão dos erros, isto é :

$$\hat{m} = \frac{\sum_{i=1}^n \hat{\varepsilon}_i}{n} \quad [92]$$

$$\hat{s} = \left[\frac{\sum_{i=1}^n (\hat{\varepsilon}_i - \hat{m})^2}{n - 1} \right]^{1/2} \quad [93]$$

- ordena-se os erros em ordem crescente.
- para cada valor do erro $\hat{\varepsilon}_i$ ordenado, calcula-se :

$$Z_i = \frac{\hat{\varepsilon}_i - \hat{m}}{\hat{s}} \quad [94]$$

$$Q(Z_i) = P [Z \leq Z_i] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_i} e^{-t^2/2} dt$$

$$S_i = \frac{i}{n}$$

$$D = \text{Máximo } [|Q(Z_i) - S_i| ; |Q(Z_{i-1}) - S_{i-1}|] \quad [95]$$

com $|Q(Z_0) - S_0| = -100$

- Decisão :

Rejeita-se H_0 se $D \geq D_{(n)}$ onde $D_{(n)}$ é dado na Tabela do ANEXO 1.

Para que esse procedimento possa ser executado por um computador, se faz necessária a inclusão do "cálculo do nível de significância observado", $Q(Z_i)$.

ABRAMOWITZ & STEGUN (1970), sugerem dentre outros, a seguinte sequência :

- Dado Z_i

$$\text{- calcular } f(Z_i) = \frac{e^{-Z_i^2/2}}{\sqrt{2\pi}} \quad [96]$$

onde $e \cong 2,7182\dots$

$$\text{- calcular } t = \frac{1}{1 + r \cdot |Z_i|} \quad [97]$$

onde $r = 0,2316419$

- calcular R

$$R = f(Z_i) \cdot \left[\sum_{j=1}^5 b_j \cdot t^j \right] \quad [98]$$

com:

$$\begin{aligned} b_1 &= 0,31938153 \\ b_2 &= -0,356563782 \\ b_3 &= 1,781477937 \\ b_4 &= -1,821255978 \\ b_5 &= 1,330274429 \end{aligned}$$

$$e \quad Q(Z_i) = \begin{cases} R & \text{se } Z_i < 0 \\ 1 - R & \text{se } Z_i \geq 0 \end{cases} \quad [99]$$

Fato :

2.27.2.3. Outras Estatísticas da Normalidade

Se o teste de LILLIEFORS conclue que os erros tem distribuição normal, então a pressuposição da normalidade dos erros foi atendida (seção 2.5.).

Se o teste conclue que os erros não tem distribuição normal, é conveniente que se pesquise para saber as causas da não normalidade. Na maioria das vezes essas causas podem ser reveladas pelas estatísticas dadas a seguir.

- Média

É esperado que os resíduos, $\hat{\varepsilon}_i$, tenham média zero. Uma estimativa dessa estatística é dada por:

$$\hat{m} = \frac{\sum_{i=1}^n \hat{\varepsilon}_i}{n}$$

- Assimetria

É sabido que a curva normal é simétrica em relação à média, ou que o momento de assimetria, m_a , é zero.

Se $m_a < 0$ então a curva tem uma simetria à esquerda.

Se $m_a > 0$ então a curva tem uma assimetria à direita.

Compare com a normal dada na seção 2.27.2.2., onde $m_a = 0$.

$$m_a = \frac{\frac{\sum_{i=1}^n (\hat{\epsilon}_i - \hat{m})^3}{n}}{\left[\frac{\sum_{i=1}^n (\hat{\epsilon}_i - \hat{m})^2}{n} \right]^{3/2}} \quad [100]$$

- Curtose

O coeficiente de curtose, m_c , (ou achatamento) para a curva normal é 3; outros valores e respectivos acrônimos estão ilustrados abaixo.

Assim, se a curva pode ser considerada simétrica, o coeficiente de curtose é dado por:

$$m_c = \frac{\frac{\sum_{i=1}^n (\hat{\epsilon}_i - \hat{m})^4}{n}}{\left[\frac{\sum_{i=1}^n (\hat{\epsilon}_i - \hat{m})^2}{n} \right]^2} \quad [101]$$

2.27.2.4. Teste de ANSCOMBE para Homocedasticidade

ANSCOMBE (1961) sugere que para testar a homocedasticidade, isto é, os erros tem a mesma variância, deve-se fazer uma regressão linear simples :

$$\hat{\varepsilon}_i^2 = a + b \hat{Y}_i \quad [102]$$

onde

$\hat{\varepsilon}_i$ são os erros (resíduos) observados ;
 $\hat{Y}_i$ são as estimativas de Y .

ambas do modelo de regressão $Y = f(X ; \theta) + \varepsilon$.

O teste consiste em verificar, pelo teste t de Student, se $\hat{b}$ é igual a zero estatisticamente.

Assim as hipóteses :

$$\begin{aligned} H_0 &: \text{os erros são homocedásticos.} \\ H_1 &: \text{os erros são heterocedásticos.} \end{aligned} \quad [103]$$

são testados usando a estatística :

$$T = \frac{|\hat{b}|}{\sqrt{V[\hat{b}]}} \quad [104]$$

Conclusão :

Se $T < t_{(n-2; \alpha/2)}$, aceita-se H_0 : os erros são homocedásticos.

Uma forma computacional para o teste de ANSCOMBE (1961) é dada pela sequência :

$$\hat{b} = \frac{n \sum \hat{Y}_i \hat{\varepsilon}_i^2 - (\sum \hat{Y}_i) \sum (\hat{\varepsilon}_i^2)}{n \sum \hat{Y}_i^2 - (\sum \hat{Y}_i)^2} \quad [105]$$

$$\hat{\sigma}^2 = \frac{\left[\sum \hat{\varepsilon}_i^4 - \frac{(\sum \hat{\varepsilon}_i^2)^2}{n} \right] - \hat{b} \left[\sum \hat{Y}_i \hat{\varepsilon}_i^2 - \frac{(\sum \hat{Y}_i) \cdot (\sum \hat{\varepsilon}_i^2)}{n} \right]}{n - 2} \quad [106]$$

$$V [\hat{b}] = \frac{\hat{\sigma}^2}{\sum \hat{Y}_i^2 - \frac{(\sum \hat{Y}_i)^2}{n}} \quad [107]$$

$$T = \frac{|\hat{b}|}{\sqrt{V[\hat{b}]}} \quad [108]$$

onde :

n é o número total de observações.

$\sum$ deve ser lido como $\sum_{i=1}^n$.

2.27.2.5. Teste de COOK & WEISBERG para Homocedasticidade

COOK & WEISBERG (1983) sugerem que para testar as hipóteses :

H_0 : os erros são homocedásticos.

H_1 : os erros são heterocedásticos.

[103]

seja usada a estatística :

$$S = \frac{\left[\sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}}) \cdot \hat{\epsilon}_i^2 \right]^2}{2 \hat{\sigma}^4 \cdot \sum_{i=1}^n (\hat{Y}_i - \bar{\hat{Y}})^2} \quad [109]$$

onde :

$$\hat{\sigma}^4 = \text{QMRes}^2 .$$

$$\bar{\hat{Y}} = \frac{\sum_{i=1}^n \hat{Y}_i}{n}$$

n é o número total de observações.

e pelo teste da razão de verossimilhança conclui-se que :

Se $S < \chi^2_{(p, \alpha)}$
 aceita-se H_0 : os erros são homocedásticos

onde p é o número de parâmetros do modelo $Y = f(X; \theta) + \varepsilon$.

2.28. Matriz de Projeção Ortogonal

SEBER & WILD (1989) descrevem que, em analogia aos modelos lineares, a matriz de projeção ortogonal, P , é dada por :

$$P_{n \times n} = Z_u (Z_u' Z_u)^{-1} Z_u' \quad [110]$$

onde Z_u é definida em [28] e tomada na convergência do processo iterativo de regressão.

Fato : Para os Métodos QUASI-NEWTON, tem-se

$$P_{n \times n} = Z_u H_u Z_u' \quad [111]$$

A matriz P , projetor ortogonal, tem diversas propriedades importantes, dentre elas :

- P é simétrica, isto é, $P = P'$;
- P é idempotente, isto é, $P^2 = P$;
- $r [P] = r [X] = r [Z_u] = \text{tr} [P]$
- P_{ii} pode ser usado para indicar se a observação i é um ponto influente ou não, através do critério :

Se :

$$P_{ii} > \frac{2p}{n} \text{ então } X_i \text{ é um ponto influente.}$$

com p sendo o número de parâmetros do modelo e n , o número total de observações.

- $P = Z_u Z_u^+$, onde Z_u^+ é a inversa de Moore-Penrose de Z_u .

Note que usando a última propriedade, é fácil demonstrar as demais.

COOK & WEISBERG (1983), dentre outros, tem mostrado que :

$$\begin{aligned}\hat{V} [\hat{\varepsilon}_i] &= (1 - P_{ii}) \cdot \text{QMRes} \\ &\text{ou} \\ \hat{V} [\hat{\varepsilon}] &= (I - P) \cdot \text{QMRes} \quad [112]\end{aligned}$$

Uma forma computacional de se obter os elementos da diagonal de P , é dada por :

$$P_{ii} = \ell_i Z_u (Z_u' Z_u)^{-1} \cdot \ell_i' Z_u \quad [113]$$

onde

$\ell_i Z_u$ é a linha i da matriz Z_u

$\ell_i' Z_u$ é o transposto da linha i da matriz Z_u .

2.29. Simulação

Simulação é uma técnica que é utilizada (dentre outras opções) para se verificar quão adequado é o "algoritmo da regressão", ou qualquer método de regressão.

Partindo de um modelo, um conjunto de dados, e um vetor de parâmetros θ^* , conhecidos, calcula-se o valor da variável dependente Y , usando $Y = f(X; \theta^*)$.

A seguir gera-se um vetor de erros independentes, ε , com distribuição normal, de média zero e variância conhecida, que será adicionado a Y , obtendo assim $Y = (X; \theta^*) + \varepsilon$.

Usando o "algoritmo de regressão", tenta-se obter uma estimativa, $\hat{\theta}$ do vetor θ^* .

Obviamente espera-se que o intervalo de confiança de $\hat{\theta}$ contenha o verdadeiro valor θ^* usado na simulação.

2.29.1. Geração de Números Aleatórios com Distribuição Normal

KNUTH (1969) sugere a seguinte expressão para gerar números aleatórios :

$$S_{i+1} = \text{frc} [9821 S_i + 0,211327] \quad [114]$$

com $i = 1, 2, 3, \dots, 10^5$ e frc indicando que se deve tomar a parte fracionária da expressão entre colchetes, e S_0 é um valor inicial dado pelo usuário, tal que $0 < S_0 < 1$.

O autor cita que esse algoritmo pode gerar 1 milhão de números aleatórios uniformemente distribuídos entre 0 e 1, a partir de S_0 , que passam pelo teste expectral.

BOX & MÜLLER (1958) sugerem o seguinte algoritmo para obtenção de um par de variáveis aleatórias independentes de mesma distribuição normal, com média m e desvio padrão s :

$$\varepsilon_i = m + s [(-2 \ln S_i)^{1/2} \cos (2\pi S_{i+1})] \quad [115]$$

$$\varepsilon_{i+1} = m + s [(-2 \ln S_i)^{1/2} \sin (2\pi S_{i+1})]$$

Fato : usar radianos para as funções trigonométricas.

GODOI (comunicação pessoal) sugere que o algoritmo de BOX & MÜLLER (1958), seja escrito como :

$$U1 = 2\pi S_i$$

$$U2 = (-2 \ln S_{i+1})^{1/2}$$

$$\varepsilon_i = m + s \cdot U2 \cdot \cos (U1) \quad [116]$$

2.30. Método da Máxima Verossimilhança para Modelos Não-Lineares

TURNER, MONROE & LUCAS (1961) mostraram que se o vetor da variável dependente $\underline{Y}$, é normal, independentemente distribuído e tem variância constante, os estimadores de máxima verossimilhança são iguais aos estimadores de mínimos quadrados.

Partem da suposição de que se o modelo $\underline{Y} = f(\underline{X} ; \underline{\theta}) + \underline{\varepsilon}$, é adequado, e $\underline{Y}$ tem distribuição normal, então o erro também tem distribuição normal, isto é :

$$\underline{\varepsilon} \sim N(\underline{0} ; \sigma^2)$$

Dessa forma o valor da densidade de probabilidade para certo valor Y_i é :

$$f(Y_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(-1/\sigma^2) \|\underline{Y}_i - f(\underline{X} ; \underline{\theta})\|^2} \quad [117]$$

A (densidade de) probabilidade do modelo $\underline{Y} = f(\underline{X} ; \underline{\theta}) + \underline{\varepsilon}$ ter gerado tais valores de $\underline{Y}_i$, é dado pela função de verossimilhança :

$$L(\underline{\theta} ; \underline{Y}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{(-1/2\sigma^2) \|\underline{Y} - f(\underline{X} ; \underline{\theta})\|^2} \quad [118]$$

Logo a estimativa de máxima verossimilhança coincide com a estimativa de mínimos quadrados, pois maximizar a função de verossimilhança $L(\underline{\theta} ; \underline{Y})$ é o mesmo que minimizar a função de mínimos quadrados $\|\underline{Y} - f(\underline{X} ; \underline{\theta})\|^2$.

Note que a estimativa de máxima verossimilhança depende da distribuição de probabilidade da variável dependente, $\underline{Y}$, fato que não ocorre na estimativa de mínimos quadrados.

HOFFMANN & VIEIRA (1987) apresentam o método de máxima verossimilhança aplicado ao modelo particular :

$$\underline{Y} = a\underline{1} + b\underline{c}^{\underline{X}} + \underline{\varepsilon}$$

2.31. Regressão Não Linear pelo Critério dos Desvios Absolutos ou Regressão Não Linear " L_1 "

É uma opção robusta ao Método dos Mínimos Quadrados, e consiste em minimizar a Soma dos Absolutos dos Erros :

$$\min \sum_{i=1}^n |\varepsilon_i| = \min \sum_{i=1}^n |Y_i - f(X_i; \theta)| \quad [119]$$

No caso da regressão linear, NARULA & STANGENHAUS (1988) apresentam de forma bastante didática o procedimento adotado por KARST (1958).

No caso não-linear 3 opções tem-se destacado :

- Os Métodos de Programação Não-Linear (BAZARAA & SHETTY (1979)) ;
- O Método de Aproximação de TISHLER & ZANG ;
- O Método Iterativo de SCHLOSSMACHER.

Aqui, apresenta-se apenas o Método de SCHLOSSMACHER, pois os problemas de otimização não-linear estão fora do escopo desse trabalho.

2.31.1. Método de SCHLOSSMACHER

SCHLOSSMACHER (1973) propõe um processo iterativo para calcular a regressão de desvios absolutos ou regressão " L_1 ", partindo de uma regressão de mínimos quadrados ponderados.

Seja o modelo $Y = f(X; \theta) + \varepsilon$; pelo método dos mínimos quadrados ponderados :

$$\min \|\varepsilon\|^2 = \|\mathbb{D} (Y - f(X; \theta))\|^2 \quad [120]$$

onde $\mathbb{D}$ é uma matriz diagonal semi-positiva definida.

Esse problema, usando a estratégia de MARQUARDT, tem solução (seção 2.20.) dada por :

$$\hat{\theta}_{u+1} = \hat{\theta}_u + (Z_u' D^{-1} Z_u + \lambda_u I)^{-1} Z_u' D^{-1} (Y - \hat{f}^u) \quad [121]$$

SCHLOSSMACHER mostrou que se D_{u+1} for escolhido como sendo $1 / |Y - \hat{f}^u|$ então o processo "converge" para regressão de desvios absolutos ou "regressão L_1 ".

2.31.2. Algoritmo de SCHLOSSMACHER

1. $u = 0$; $D = I$; forneça a estimativa inicial $\hat{\theta}_u$
2. Calcule $\hat{\theta}_{u+1}$ usando a expressão [121]
3. Calcule $R = Y - \hat{f}^{u+1}$
4. Se $|R_k| > 10^{-\sigma}$, $k = 1, 2, \dots, n$
então $D_{kk} = 1 / |R_k|$
senão : $D_{kk} = 0$
onde D_{kk} é o elemento da diagonal de D
5. Critério de convergência (seção 3.4.2.3.)
6. $u = u + 1$ e volta ao passo 2

2.31.3. Algumas Considerações

DIELMAN & PFAFFENBERGER (1982) apresentam algumas propriedades dos estimadores da regressão L_1 (linear). Mostram que existem alguns resultados assintóticos para inferência na regressão L_1 (linear), mas não citam sua validade para as regressões não-lineares.

Nesse trabalho a regressão L_1 é considerada apenas como uma opção robusta em relação à regressão L_2 , e as comparações limitam-se aos resultados da regressão L_2 ($\hat{\theta}$; $\sum_{i=1}^n |\hat{\epsilon}_i|$) em relação aos mesmos resultados da regressão L_1 .

O leitor poderá notar visualmente, através do gráfico gerado pelos resultados das regressões L_1 e L_2 , que a regressão L_1 muitas vezes apresenta melhores resultados.

3. DESCRIÇÃO DO SISTEMA COMPUTACIONAL

3.1. Escolha do Modelo

O problema da escolha do modelo é resolvido considerando que o usuário é capaz de fornecer, ou selecionar, uma função que represente seus dados de forma satisfatória.

Dessa forma criou-se 2 opções, a saber :

- Biblioteca de modelos ;
- Função do usuário.

A biblioteca contém modelos usuais (lineares, linearizáveis e não-lineares) que podem ser acessados através do "nome da função".

Se o modelo é específico, e não consta da biblioteca, o usuário pode fornecer seu modelo matemático no menu "função do usuário".

Obviamente o leitor poderá fornecer um modelo da biblioteca no menu "função do usuário", e o mesmo será calculado.

Deve ser ressaltado, que na opção "função do usuário" não é necessário fornecer as derivadas do modelo em relação aos parâmetros, pois essas derivadas são calculadas numericamente.

No caso da "biblioteca de modelos", existe um arquivo contendo :

- O modelo propriamente dito ;
- As derivadas analíticas do modelo em relação aos parâmetros.

Dessa forma o leitor deverá optar, sempre que possível, pelos modelos da biblioteca pois, além de maior precisão, são mais rápidos.

3.1.1. Modelos da Biblioteca

Esses modelos foram extraídos de KOLB (1982), do arquivo *Models* da HP71, de diversos artigos dos periódicos Chemical Engineering, Biometrics, JASA, etc..., e também de algumas teses e dissertações apresentadas na ESALQ/USP.

Deve-se ressaltar que no computador o usuário pode verificar a expressão matemática do modelo, selecionando o nome da função e a seguir pressionando a tecla de ajuda F1.

3.1.1.1. Relação do Modelos da Biblioteca

Modelo	Expressão Matemática
Reta Através de Origem	$Y = ax + \epsilon$
Parábola Através da Origem	$Y = ax + bx^2 + \epsilon$
Hipérbole	$Y = a + \frac{b}{x} + \epsilon$
Hipérbole de Segunda Ordem	$Y = a + \frac{b}{x} + \frac{c}{x^2} + \epsilon$
Potência	$Y = ax^b + \epsilon$
Potência com "Off Set"	$Y = a + bx^c + \epsilon$

Modelo	Expressão Matemática
Potência com 2 Termos	$Y = ax^b + cx^d + \epsilon$
Potência com 3 Termos	$Y = ax^b + cx^d + ex^f + \epsilon$
Potência Modificada	$Y = ab^x + \epsilon$
Potência Modificada com "Off Set"	$Y = a + bc^x + \epsilon$
Recíproco de Reta	$Y = \frac{1}{a + bx} + \epsilon$
Logística	$Y = \frac{a}{1 + bc^x} + \epsilon$
Raiz	$Y = ab^{(1/x)} + \epsilon$
Super-Geométrica	$Y = ax^{bx} + \epsilon$
Raiz Geométrica	$Y = ax^{(b/x)} + \epsilon$
Linear-Hiperbólica	$Y = a + bx + \frac{c}{x} + \epsilon$
Recíproco de Hipérbole	$Y = \frac{x}{a + bx} + \epsilon$
Exponencial	$Y = a e^{bx} + \epsilon$ $e = 2,7182 \dots$
Exponencial Assintótica	$Y = a(1 - e^{bx}) + \epsilon$ $e = 2,7182 \dots$
Exponencial Dupla	$Y = a(e^{bx} - e^{cx}) + \epsilon$ $e = 2,7182 \dots$
Raiz Exponencial	$Y = a e^{(b/x)} + \epsilon$ $e = 2,7182 \dots$
Linear Exponencial	$Y = a x e^{bx} + \epsilon$ $e = 2,7182 \dots$
Logarítmica	$Y = a + b \cdot \ln(cx) + \epsilon$
Recíproca de Logarítmica	$Y = \frac{1}{a + b \ln(cx)} + \epsilon$
Função de Hoerl	$Y = ab^x x^c + \epsilon$
Função Raiz de Hoerl	$Y = ab^{(1/x)} x^c + \epsilon$

Modelo	Expressão Matemática
Distribuição Normal	$Y = a e^{-(x-b)^2/c} + \epsilon$ $e = 2,7182 \dots$
Distribuição Log-Normal	$Y = a e^{(\ln(x)-b)^2/c} + \epsilon$ $e = 2,7182 \dots$
Distribuição Beta	$Y = ax^b (1-x)^c + \epsilon$
Distribuição Gamma	$Y = a \left[\frac{x}{b} \right]^c \cdot e^{-(x/b)} + \epsilon$ $e = 2,7182 \dots$
Distribuição de Cauchy	$Y = \frac{1}{a(x+b)^2 + c} + \epsilon$
Senóide	$Y = a \text{ sen}(b + cx) + \epsilon$
Senóide com "Off Set"	$Y = a \text{ sen}(b + cx) + d + \epsilon$
Cossenóide	$Y = a \text{ cos}(b + cx) + \epsilon$
Cossenóide com "Off Set"	$Y = a \text{ cos}(b + cx) + d + \epsilon$
Seno-Fourier com 2 Termos	$Y = b + c \text{ sen}(ax) + \epsilon$
Seno-Fourier com 3 Termos	$Y = b + c \text{ sen}(ax) +$ $+ d \text{ sen}(2ax) + \epsilon$
Seno-Fourier com 4 Termos	$Y = b + c \text{ sen}(ax) + d \text{ sen}(2ax)$ $+ e \text{ sen}(3ax) + \epsilon$
Cosseno-Fourier 2 Termos	$Y = b + c \text{ cos}(ax) + \epsilon$
Cosseno-Fourier 3 termos	$Y = b + c \text{ cos}(ax) +$ $+ d \text{ cos}(2ax) + \epsilon$
Cosseno-Fourier 4 Termos	$Y = b + c \text{ cos}(ax) + d \text{ cos}(2ax)$ $+ e \text{ cos}(3ax) + \epsilon$
Fourier com 2 Termos	$Y = \frac{b}{2} + c \text{ cos}(ax) +$ $+ d \text{ sen}(ax) + \epsilon$
Fourier com 3 Termos	$Y = \frac{b}{2} + c \text{ cos}(ax) + d \text{ sen}(ax)$ $+ e \text{ cos}(2ax) +$ $+ f \text{ sen}(2ax) + \epsilon$

Modelo	Expressão Matemática
Fourier com 4 Termos	$Y = \frac{b}{2} + c \cos(ax) + d \sin(ax) + e \cos(2ax) + f \sin(2ax) + g \cos(3ax) + h \sin(3ax) + \epsilon$
Cosseno hiperbólico	$Y = a + \cosh(b + cx) + \epsilon$
Secante hiperbólica	$Y = a + \operatorname{sech}(b + cx) + \epsilon$
Reta	$Y = a + bx + \epsilon$
Parábola	$Y = a + bx + cx^2 + \epsilon$
Cúbica	$Y = a + bx + cx^2 + dx^3 + \epsilon$
Polinômio do 4º Grau	$Y = a + bx + cx^2 + dx^3 + ex^4 + \epsilon$
⋮	⋮
Polinômio do 19º grau	$Y = a + bx + cx^2 + \dots + \dots + tx^{19} + \epsilon$
Prunty-2 ⁽³⁾	$Y = a + bx + c[(x-d)^2 + e]^{1/2} + \epsilon$
Prunty-3	$Y = a + bx + c[(x-d)^2 + e]^{1/2} + f[(x-g)^2 + h]^{1/2} + \epsilon$
Mitscherlich-1	$Y = a[1 - 10^{-c(x+b)}] + \epsilon$
Mitscherlich-2	$Y = a[1 - 10^{-c(x_1+b)}] \cdot [1 - 10^{-d(x_2+e)}] + \epsilon$
Mitscherlich-3	$Y = a[1 - 10^{-c(x_1+b)}] \cdot [1 - 10^{-d(x_2+e)}] \cdot [1 - 10^{-f(x_3+g)}] + \epsilon$
2ª Lei de Mitscherlich	$Y = a[1 - 10^{-c(x+b)}] \cdot [10^{-d(x+b)^2}] + \epsilon$
Plano	$Y = a + bx_1 + cx_2 + \epsilon$
Superfície do 2º Grau	$Y = a + bx_1 + cx_2 + dx_1x_2 + ex_1^2 + fx_2^2 + \epsilon$

.....
⁽³⁾ Biometrics (1983).

Modelo	Expressão Matemática
Espaço do 2º Grau	$Y = a + bx_1 + cx_2 + dx_1x_2 +$ $+ ex_1^2 + fx_2^2 + gx_1^2x_2 +$ $+ hx_1x_2^2 + ix_1^2x_2^2 + jx_1^3 +$ $+ kx_2^3 + \epsilon$
Hiperplano (3 variáveis)	$Y = a + bx_1 + cx_2 + dx_3 + \epsilon$
Superfície do 3º Grau	$Y = a + bx_1 + cx_2 + dx_3 +$ $+ ex_1x_2 + fx_1x_3 + gx_2x_3 +$ $+ hx_1^2 + ix_2^2 + jx_3^2 + \epsilon$

FIGURA 4. Relação dos Modelos da Biblioteca

3.1.1.2. Um Exemplo do Arquivo de Modelos da Biblioteca

Considere o modelo Prunty-2 :

$$Y = a + bx + c [(x - d)^2 + e]^{1/2}$$

que representa 2 retas com intercepto em d.

Se $e \rightarrow 0$ o intercepto é abrupto: se $e \rightarrow 1$ o intercepto é suave.

Para calcular o jacobiano (matriz Z), necessita-se da derivadas⁽⁴⁾ :

$$\frac{\partial Y}{\partial a} = 1 \quad ; \quad \frac{\partial Y}{\partial b} = x \quad ; \quad \frac{\partial Y}{\partial c} = [(x-d)^2 + e]^{1/2}$$

⁽⁴⁾ O autor agradece ao Prof. Dr. Ary O. Chiacchio, do Departamento de Matemática do IMECC/UNICAMP pelo apoio, conferências e sugestões.

$$\frac{\partial Y}{\partial d} = \frac{-c (x - d)}{[(x - d)^2 + e]^{1/2}} \quad ; \quad \frac{\partial Y}{\partial e} = \frac{c}{2[(x - d)^2 + e]^{1/2}}$$

De posse do modelo e suas derivadas, pode-se criar o campo principal do arquivo .DBF :

```
{! P, X, G, M ;
G[1] = 1
G[2] = X[1]
M = X[1] - P[4]
G[3] = INDC (M * M + P[5] >= 0 ) *
          SQRT (ABS (M * M + P[5] )) + 0,001
G[4] = -P[3] * M/G[3]
G[5] = P[3]/(2 * G[3])
M = P[1] + P[2] * G[2] + P[3] * G[3]}
```

que, particularmente, representa o modelo Prunty-2 da biblioteca.

3.1.2. Modelos do Usuário

Se o modelo é específico, o leitor pode optar pela "função do usuário". Nesse caso, o computador apresenta :

- O nome das variáveis independentes disponíveis : X[1], X[2] ,..., X[20]
- O nome dos parâmetros disponíveis :
a, b, ...,t
- O nome das funções matemáticas disponíveis (através de uma tela de ajuda) : +, -, *, ..., sin, ...

e o usuário deve entrar sua fórmula usando as regras da álgebra.

Isso é possível devido ao "Interpretador de Fórmulas", que a partir de uma *string* ("função do usuário") retorna um valor numérico (CHIACCHIO (1990)).

Considere o modelo : $Y = a + b^*x^c + \varepsilon$, que deve ser expresso como :

$$a + b * x[1] * x[1]^c$$

já que existe apenas uma variável independente $x \rightarrow x[1]$, e 3 parâmetros : a, b, c.

Seja o modelo :

$$Y = a + b \sin x_1 + c \cos x_2 + d \sin x_1 \cos x_2 + \varepsilon$$

Assim o modelo deve ser expresso por :

$$a + b * \sin(x[1]) + c * \cos(x[2]) + d * \sin(x[1]) * \cos(x[2])$$

Considere o ajuste de uma poligonal (HOFFMANN & VIEIRA (1987)), dado pelo modelo $Y = f(x; \theta) + \varepsilon$, onde :

$$Y = \begin{cases} a + b x & \text{se } x \leq 0 \\ a + (b + c) x & \text{se } x > 0 \end{cases}$$

então o modelo do usuário é :

$$a + b * (x[1]) + c * (x[1]) * \text{INDC} (x[1] > 0)$$

onde a função "*indicador*", dada por $\text{INDC} (x[1] > 0)$, retorna

$$\begin{cases} 0 & \text{se } x[1] \leq 0 \\ 1 & \text{se } x[1] > 0 \end{cases}$$

De forma genérica, a função INDC (a operador b) retorna :

$$\begin{array}{ll} 0 & \text{se (a operador b) for falso} \\ 1 & \text{se (a operador b) for verdadeiro} \end{array}$$

onde os operandos (a, b) podem ser compostos por expressões complexas .

Veja por exemplo a expressão G[3] do modelo Prunty-2 na seção 3.1.1.2. ; é fácil concluir que o segmento abaixo :

```
IF M * M + P[5] >= 0
  THEN G[3] = SQRT (M * M + P[5]) + 10-3
  ELSE G[3] = 10-3
```

foi substituído pelo seu equivalente INDC ().

Isso é necessário visto que um arquivo .DBF não pode conter a sequência IF ... THEN ... ELSE

3.2. Hardware

3.2.1. Necessário

Computador padrão IBM-PC, com :

- Pelo menos 640 kbytes de RAM
- Unidade de disco rígido
- Pelo menos um drive de 5 1/4
- Monitor monocromático
- Teclado.

Embora o software "rode" em um XT, deve-se optar por máquinas mais velozes (286, 386, 486 ou algum microprocessador mais recente).

3.2.2. Opcional

- Impressora padrão Epson 80 colunas
- Monitor policromático (EGA, VGA)

3.3. Software Necessário

- Programa executável "REGRE", incluindo todos arquivos de suporte.

3.4. Descrição do Software

O sistema é composto por 12 arquivos com aproximadamente 10.000 linhas de programa.

A linguagem Clipper foi utilizada devido a dois fatores principais :

- Facilidade de construção de telas gráficas de entrada e/ou saída de dados, conhecidas como janelas ;
- Possibilidade de acesso a outras linguagens de programação, como por exemplo, a linguagem C. (de onde se extraiu as funções trigonométricas, hiperbólicas, transcendent-tes, etc...)

A descrição do software será limitada aos procedimentos de regressão não-linear, posto que é o tema central desse trabalho.

3.4.1. Descrição Gráfica das Opções do Software para Regressão Não-Linear

Descreve-se graficamente as conexões (opções) entre :

- Métodos
- Cálculos
- Estatísticas
- Gráficos

que são disponíveis na versão atual do programa.

m é t o d o s	Regressão L_2			Regressão L_1	
	Gauss-Newton	Marquardt	Quasi-Newton	Schlossmacher	Schlossmacher modificado
			DFP BFGS		

c á l c u l s	Gauss-Newton	Marquardt	DFP BFGS	L_1 original	L_1 modificado
---------------------------------	--------------	-----------	-------------	----------------	------------------

e s t a t í s t í c a s	ANOVA
	I C Parametros
	Mat Var/Covar
	Mat Correlação
	Análise dos resíduos
	Pressuposições

g r á f í c o s	Resíduos
	Função

3.4.2. Descrição dos Cálculos

3.4.2.1 Método de GAUSS-NEWTON

Passo	Sequência de cálculos	expressão/seção	pg
1	Jacobiano	[28]	27
2	SEN	[a]	28
3	Solução do SEN	[31]	28
4	Atualiza parâmetros	[32]	28
5	Critério de convergência	[25] ou [26]	25
6	SQRes	[38]	32
7	Verifica mínimo estacionário	2.12.3	26
8	volta ao passo 1	-	-

Fatos :

No cálculo do Jacobiano, tem -se :

- Modelo da biblioteca : as derivadas são fornecidas analiticamente, e o programa calcula o valor numérico ;
- Modelo do usuário : as derivadas são aproximadas numericamente, através da expressão :

$$\frac{\partial f}{\partial \theta_k} \approx \frac{f(\mathbf{X} ; \theta_k + h) - f(\mathbf{X} ; \theta_k)}{h} \quad [122]$$

para $k = 1, 2, \dots, p$ e $h = 10^{-8}$, podendo ser alterado pelo usuário.

No cálculo do Sistema de Equações Normais (SEN) tem-se que calcular $\mathbf{A} = \mathbf{Z}' \mathbf{Z}$ e $\mathbf{Q} = \mathbf{Z}' (\mathbf{Y} - \mathbf{f})$. O segmento a seguir mostra como calcular esses produtos sem ter $\mathbf{Z}'$ explicitamente (truque necessário para se "economizar" memória no computador) :

```

FOR I = 1 TO número de parâmetros
  FOR J = I TO número de parâmetros
    FOR K = 1 TO número de observações
       $A[I, J] = A[I, J] + Z[K, I] * Z[K, J]$ 
      IF I # J THEN  $A[J, I] = A[I, J]$ 
    NEXT K
  NEXT J
NEXT I
FOR J = 1 TO número de parâmetros
  FOR I = 1 TO número de observações
     $C[J] = C[J] + Z[I, J] * W[I]$  !  $C = Z'(Y - f)$ 
  NEXT I
NEXT J

```

Na solução do Sistema de Equações Normais (SEN), usa-se a Inversa Generalizada Condicional (IEMMA (1988)), GODOI (comunicação pessoal) posto que o SEN pode ser singular ou numericamente singular (problemas mal condicionados). Versões futuras deverão implementar a Inversa de Moore-Penrose, que é sabidamente melhor que a Inversa Condicional (IEMMA (1988)).

3.4.2.2. Método de MARQUARDT

Passo	Sequência de cálculos	expressão/seção	pg
1	$u = 0$ $\lambda_u = 10^{-9}$ $SQRes_u = SQTo$	-	-
2	Jacobiano	[28]	27
3	SEN	-	-
4	Solução do SEN	[42]	33
5	$SQRes_{u+1} = \ \Delta u \ ^2$	[38]	32
6	Se $SQRes_{u+1} \leq SQRes_u$	-	-
6.a	então vá para o passo 7	-	-
6.b	senão incremente λ_u e volte ao passo 3	-	-
7	Atualiza parâmetros	[43]	34
8	Critério de convergência	[25] ou [26]	25
9	Verifica mínimo estacionário	2.12.3	26
10	$\lambda_{u+1} = \lambda_u / 10$	-	-
11	$u = u + 1$	-	-
12	volta ao passo 2	-	-

Alguns detalhes da montagem parcial do SEN podem ser encontradas na seção 3.4.2.1. . Para concluir a montagem do SEN de Marquardt, não é necessária a matriz λI ; o artifício utilizado consiste em somar diretamente o valor de λ aos elementos da diagonal de $A = Z'Z$.

3.4.2.3. Métodos QUASI-NEWTON

O procedimento completo e detalhado está apresentado na seção 2.16.3.

3.4.2.4. Método de SCHLOSSMACHER Original

O procedimento completo está na seção 2.31.2.

A fim de tornar o procedimento mais computacional, usou-se a expressão [121], como a seguir :

$$\begin{aligned} \varrho_{u+1} = \varrho_u + (Z'_u D_u^{-1/2} D_u^{-1/2} Z_u + \lambda_u I)^{-1} \cdot \\ \cdot Z'_u D_u^{-1/2} D_u^{-1/2} (Y - f^u) \end{aligned} \quad [123]$$

onde

$$D_{kk}^{-1/2} = \frac{1}{\sqrt{|R_k|}} \quad \text{se } |R_k| > 10^{-6} \quad [124]$$

ou

$$D_{kk}^{-1/2} = 0 \quad \text{se } |R_k| < 10^{-6} \quad [125]$$

e o produto $D_u^{-1/2} Z_u$ é armazenado na matriz Z_u e $D_u^{-1/2}(Y - f^u)$ é armazenado no vetor $(Y - f^u)$.

Considere :

$$\sum_{i=1}^n |Y_i - f_i^u| = R_u \quad [126]$$

$$\sum_{i=1}^n |Y_i - f_i^{u+1}| = R_{u+1} \quad [127]$$

então o critério de convergência é dado por :

$$|R_u - R_{u+1}| < e \quad [128]$$

onde e é um limite de erro dado pelo usuário.

Ocorre que nem sempre se obtém o critério de convergência em iterações consecutivas. Dessa forma, o critério foi modificado para :

$$|R_m - R_{u+1}| < e \quad [129]$$

onde R_m é a menor R_u obtida.

Se $R_{u+1} < R_m$ e $|R_m - R_{u+1}| < e$, então $R_m = R_{u+1}$ e o processo continua até que [129] seja satisfeita.

3.4.2.5. Método de SCHLOSSMACHER Modificado

Verificou-se empiricamente que se a expressão [124] for substituída por :

$$D_{kk}^{-1/2} = \frac{1}{|R_k|} \quad \text{se } |R_k| > 10^{-6} \quad [130]$$

mantidas as demais considerações, o processo converge mais rapidamente. Embora não se prove teoricamente a convergência, na prática o processo tem se mostrado satisfatório.

3.4.3. Descrição das Estatísticas

3.4.3.1. Análise de Variâncias (ANOVA)

Usa-se a análise usual (seção 2.23.), com duas restrições :

- Se $SQRegr < 0$ então $SQRegr = 0$
- Se $SQRes = 0$ então $SQRes = 10^{-10}$

A primeira é justificada considerando que longe da solução $\hat{\theta}$, a $SQRes$ é maior que a $SQTo$, gerando $SQRegr < 0$, o que invalidaria a ANOVA.

A segunda deve-se ao fato de que o usuário pode necessitar resolver um problema exato ou um problema de geometria analítica, como por exemplo na estimativa do valor inicial θ_0 (seção 2.21.). Nesse caso $SQRes = 0$ implicando em $F \rightarrow \infty$ (divisão por 0).

3.4.3.2. Intervalo de Confiança dos Parâmetros

Usa-se a sugestão de DONALDSON & SCHNABEL (1987), descrita na seção 2.25.

Para o cálculo do intervalo, usa-se 2 vetores de forma que o primeiro guarde o limite inferior e o segundo o limite superior.

O valor "t de Student" é pesquisado em tabela armazenada na memória do computador; valores que não constam da tabela são obtidos por interpolação harmônica, isto é :

gl		t Student	
n_2			t_2
n			t
n_1			t_1

de onde :

$$t = t_2 + \frac{\left[t_1 - t_2 \right] \cdot \left[\frac{1}{n} - \frac{1}{n_2} \right]}{\frac{1}{n_1} - \frac{1}{n_2}} \quad [131]$$

sendo que os extremos correspondem a valores tabelados.

3.4.3.3. Matriz de Variâncias/Covariâncias dos Parâmetros

Usou-se a sugestão de DRAPER & SMITH (1981), dada pelas equações [78] ou [79] , definidas na seção 2.24.

É utilizado um arquivo texto para sua apresentação.

3.4.3.4. Matriz de Correlação dos Parâmetros

Usou-se a definição de DRAPER & SMITH (1981), dada pela expressão [81].

Utilizou-se um arquivo texto para sua apresentação.

3.4.3.5. Análise dos Resíduos

Usou-se um arquivo texto que contém uma tabela como abaixo :

$\tilde{Y}$	$\hat{\tilde{Y}}$	$\hat{\varepsilon}$	$\hat{\varepsilon}_N$	diag(P)
$\approx$				$\approx$

onde :

$\tilde{Y}$ são os valores observados da variável dependente

$\hat{\tilde{Y}}$ são os valores estimados para a variável dependente

$\hat{\varepsilon}$ é a estimativa dos erros

$\hat{\varepsilon}_N$ é a estimativa normalizada dos erros (expressão [85])

diag(P) são os elementos da diagonal da matriz de projeção (seção 2.28).

Deve ser ressaltado que "observações que excedem seu valor crítico, são seguidos de um '*' " (equação [86] e seção 2.28.).

3.4.3.6. Verificação das Pressuposições do Modelo

São testadas as seguintes hipóteses :

- Aleatoriedade dos Erros : teste de Chorri-
lhos (seção 2.27.2.1.)
- Normalidade dos Erros : teste de Lilliefors
(seção 2.27.2.2.)
- Outras estatísticas da normalidade (seção
2.27.2.3.) :
 - Média
 - Mediana
 - Variância
 - Assimetria
 - Curtose
- Homo/Heterocedasticidade
 - Teste de ANSCOMBE (seção 2.27.2.4.)
 - Teste de COOK & WEISBERG (seção 2.27.2.5.)

É conveniente ressaltar que o programa calcula essas estatísticas usando as definições apresentadas na seção 2.27.2.

Os resultados são apresentados usando um arquivo texto.

3.4.4. Gráficos

Utilizou-se a biblioteca gráfica CLBC para Clipper 5.0 , que utiliza algumas subrotinas em C.

3.4.4.1. Gráficos dos Resíduos Normalizados

É bem sabido que deve-se plotar os resíduos colocando no eixo das abscissas o valor de $\hat{Y}$ e no eixo das ordenadas o valor de $\hat{\varepsilon}_N$, pois $\hat{Y} \hat{\varepsilon}_N = 0$, isto é, não há correlação entre $\hat{Y}$ e $\hat{\varepsilon}_N$ (seção 2.27.1.)

O leitor pode comparar o gráfico obtido com aqueles apresentados na figura 3 e concluir a respeito do ajuste.

3.4.4.2. Gráficos da Função

Coloca-se em um sistema de coordenadas cartesianas, os valores x e y observados (variável independente e variável dependente) representadas por "*", e os valores calculados de $\hat{Y} = f(X; \hat{\theta})$ representados por ".".

Infelizmente o sistema utilizado só permite fazer gráficos bidimensionais.

No caso da regressão L_1 , plota-se ambas as regressões L_2 e L_1 em um mesmo gráfico para efeito de comparações. Nesse caso :

- pontos L_2 : "..."
- pontos L_1 : "——"
- dados : "*"

4. DISCUSSÃO E RESULTADOS

4.1. Introdução

Entre o resultado matemático e sua representação aritmética há uma grande lacuna. Ainda hoje não se conhece um programa de computador capaz de manipular dados aritméticos de forma matemática.

Considere $\sqrt{5} \cdot \ln \pi$; cada máquina dará uma resposta que é dependente do número de algarismos de sua mantissa (single, double, extend, etc ...), do tipo de operação interna (Binário, Hexadecimal, Binário codificado em decimal, etc...), dos tipos de aproximação para as operações (expansão em série para funções trigonométricas e transcendentais, algoritmos especiais, etc...), etc...

O caso ainda se torna mais grave quando ocorrem operações não definidas em um particular conjunto solução. Tais operações "impróprias" tem significado bastante conhecido em outros conjuntos (ex : $X/0$ só tem sentido no limite, $\sqrt{-5}$ só tem resposta no conjunto dos complexos, etc...)

Cada programa apresenta uma resposta diferente a essas "anomalias".

Para tentar resolver esses problemas, o "Institute of Electrical and Electronics Engineers", IEEE, propõe em 1984 algumas regras, descritas em "IEEE PROPOSAL

FOR HANDLING MATH EXCEPTIONS". definindo dessa forma uma estratégia para solucionar essas 'anomalias'.

Assim , dentre outras, define-se :

$$\frac{1}{0} = \text{inf}$$

$$\frac{0}{0} = \text{NaN} \quad (\text{Not a Number ou não é número})$$

$$\text{sen}(\text{inf}) = \text{NaN}$$

$$\sqrt{-3} = \text{NaN} \quad (\text{se o modo atualmente definido é real})$$

Uma descrição mais completa do padrão pode ser encontrada em 'HP71 Reference Manual', 1984, páginas 338 - 345.

É conveniente ressaltar que o programa não "aborta" se ocorre um erro "tratável", mas continua a execução com as regras adotadas.

Exemplos :

$$2 * \text{NaN} = \text{NaN} ; \sqrt{3} . \text{inf} = \text{inf} ; \text{inf} - \text{inf} = \text{NaN} ; \text{etc...}$$

Infelizmente não se encontrou na versão 5.0 do Clipper esse padrão tão útil⁽⁵⁾ .

Espera-se também algoritmos mais eficientes, capazes de manipular em sua forma original :

- Números racionais, irracionais e transcendentos
- Expansão, compressão e simplificação de literais (ex : $(a + b)^2 = a^2 + 2ab + b^2$)
- Cálculo diferencial e integral, com literais.

Alguns "pacotes" já manipulam algumas expressões dessa forma (Mathematica, HP48SX), contudo não se teve acesso a eles.

⁽⁵⁾ O autor agradece por qualquer informação que leve à utilização do padrão citado.

4.2. Conjuntos de Dados e Modelos Adotados para Análise

4.2.1. Dados e Modelos de SILVA (1978)

Dados simulados :

x	y
0	36,6
0,2	53,6
0,4	69,9
0,6	86,2
0,8	101,0
1,0	110,0
1,2	119,0
1,4	128,0
1,6	132,0
1,8	133,0
2,0	131,0
2,2	129,0
2,4	120,0

onde x representa as doses de nutriente utilizado e y a média dos tratamentos observados.

Modelos propostos :

$$Y = a \left[1 - 10^{-c(x+b)} \right] + \epsilon$$

$$Y = a + bx + cx^2 + \epsilon$$

$$Y = a \left[1 - 10^{-c(x+b)} \right] \cdot 10^{-d(x+b)^2} + \epsilon$$

4.2.2. Dados e Modelo de LANDRIANI et alli (1983)

Dados da infusão endovenosa da droga VM26 anticâncer :

t	Y
0,5	1,234
1	2,093
1,5	2,323
2	2,916
2,5	3,364
3	2,722
4,16	3,570
6	3,875
8	5,009
15	6,012
19	6,998
23	7,542
25	6,51
25,5	6,131
26	6,010
27	5,271
28	4,635
30	3,645
32	3,642
42	1,234
48	0,991

onde t é o tempo em horas após a infusão da droga e Y é a concentração da droga no sangue.

Modelo proposto :

$$Y = \frac{a}{b} [e^{kb} - 1] \cdot e^{-bt} + \frac{c}{d} [e^{kd} - 1] \cdot e^{-dt} + e$$

onde :

$$k = \begin{cases} t & \text{se } t < 24 \\ 24 & \text{se } t \geq 24 \end{cases}$$

e

$$e = 2,7182...$$

4.3. Análise dos Dados de SILVA (1978)

Nos quadros 1, 2, 3, 4 e 5 a coluna SAS (Statistical Analysis System) foi introduzida para se comparar os resultados obtidos pela Regressão de Mínimos Quadrados (L_2) com um padrão aceito internacionalmente (Procedure NLIN, método DUD). Não se encontrou um padrão para a Regressão de Desvios Absolutos (L_1).

Os resultados apresentados foram obtidos em computador PC 386, 30 MHz.

4.3.1. Modelo : 1ª Lei de MITSCHERLICH

Partindo das estimativas iniciais fornecidas pela autora :

$$\theta_0 = \begin{bmatrix} 100 \\ 0.8 \\ 0.4 \end{bmatrix}$$

obteve-se os resultados apresentados no Quadro 1.

Para a Regressão de Mínimos Quadrados, todos métodos convergiram para a mesma estimativa, embora com tempo de execução e número de iterações distintos.

A análise de resíduos mostra que os erros não tem distribuição normal, são homocedásticos e tem alguns pontos influentes e/ou aberrantes.

A discrepância observada para os pontos influentes deve-se à maneira pela qual são calculadas as

matrizes que formam a matriz de projeção.

Algumas outras estatísticas, tais como :

- Intervalo de confiança dos parâmetros a 95% de probabilidade
- Matriz de variâncias/covariâncias assintótica
- Matriz de correlação dos parâmetros

também foram fornecidas pelo programa e seus valores são semelhantes aos obtidos pelo SAS.

Obteve-se também alguns testes sobre as pressuposições do modelo, que constam de forma resumida no Quadro 1.

Na Figura 5 apresenta-se os gráficos das funções ajustadas.

Para a Regressão de Desvios Absolutos (L_1) considerou-se como limite, 100 iterações. Nesse caso, apenas a regressão " L_1 " modificada convergiu. Os resultados estão no Quadro 1.

Na Figura 6 apresenta-se os gráficos das funções ajustadas.

Apresenta-se no Anexo 5 os resultados, tais quais obtidos, pelos Métodos de MARQUARDT e SCHLOSSMACHER modificado.

O leitor pode verificar que para a regressão de mínimos quadrados (L_2), há no Anexo 5 :

- Análise de variâncias com teste de hipótese sobre a validade do modelo.
- Estimativa por ponto dos parâmetros.
- Número de iterações efetuadas.
- Coeficientes de determinação e comparação.
- Lambda (constante de Marquardt).
- Norma do vetor correção ($\| GRD \|$).
- Intervalo de confiança dos parâmetros a 95% de probabilidade.
- Matriz de variâncias e covariâncias assintóticas.

- Gráfico da função.
- Análise dos resíduos ($\hat{Y}$, $\hat{e}$, $\hat{e}$ normalizado, diagonal de projeção).
- Gráfico dos resíduos normalizados.
- Matriz de correlações dos parâmetros (assintótica).
- Verificação das pressuposições do modelo (Aleatoriedade, normalidade, homocedasticidade).

Para a regressão de desvios absolutos (L_1),
tem-se :

- Resumo da regressão L_2 (mínimos quadrados).
- Estimativas por ponto (L_1 e L_2).
- $\sum |Y|$, $\sum |\hat{e}|$ (menor e atual).
- Gráficos da função.

MÉTODOS								
		GAUSS- NEWTON	MARQUARDT	DFP	BFGS	SAS	L1 Original	L1 Modificado
PARÂMETROS	a	136,8929	136,8929	136,8929	136,8929	136,8929	138,1879	138,1875
	b	0,1924	0,1924	0,1924	0,1924	0,1924	0,1583	0,1583
	c	0,5896	0,5896	0,5896	0,5896	0,5896	0,5948	0,5948
SQRes		390	390	390	390	390	429,7385	429,7361
$\sum e $		56,8827	56,8827	56,8827	56,8827	-	53,5331	53,5331
Iterações		14	31	16	17	18	100	97
Tempo (seg)		5,93	23,62	7,58	12,97	-	322,25	343,12
R ²		0,969	0,969	0,969	0,969	-	-	-
F		1300	1300	1300	1300	-	-	-
Normalidade		Não	Não	Não	Não	-	-	-
Ponto Influyente		1º	1º	-	1º e 2º	-	-	-
Ponto Aberrante		13º	13º	13º	13º	-	-	-
Homo/Heterocedasticidade		Homo	Homo	Homo	Homo	-	-	-

QUADRO 1. Resumo das Análises dos Dados de SILVA (1978) Usando a 1ª Lei de Mitscherlich

FIGURA 5. Gráficos das regressões L2 para os dados de SILVA (1978) usando a 1ª Lei de Mitscherlich.

Original

Modificado

FIGURA 6. Gráfico das regressões L_1 para os dados de SILVA (1978) usando a 1ª Lei de Mitscherlich.

4.3.2. Modelo : Parábola

Usando como estimativa inicial :

$$\theta_0 = \begin{bmatrix} 0,1 \\ 0,1 \\ 0,1 \end{bmatrix}$$

obteve-se os resultados apresentados no Quadro 2.

Para a Regressão de Mínimos Quadrados, todos métodos convergiram para a mesma estimativa, embora com tempo de execução e número de iterações distintos.

A análise dos resíduos mostra que os erros não tem distribuição normal, são homocedásticos e tem alguns pontos influentes.

A discrepância observada para a presença/ausência de pontos influentes deve-se à maneira pela qual são calculadas as matrizes que formam a matriz de projeção.

Algumas outras estatísticas, tais como :

- Intervalo de confiança dos parâmetros a 95% de probabilidade
- Matriz de variâncias/covariâncias
- Matriz de correlação dos parâmetros

também foram fornecidas pelo programa, e seus valores não puderam ser checados com os valores do SAS, pois para o SAS a matriz de coeficientes ($Z'Z$) é singular.

Obteve-se também alguns testes sobre as pressuposições do modelo que constam de forma resumida no Quadro 2.

Na Figura 7 apresentam-se os gráficos das funções ajustadas.

Para a Regressão de Desvios Absolutos (L_1) considerou-se como limite, 100 iterações.

Na Figura 8 apresentam-se os gráficos das funções ajustadas.

É conveniente ressaltar que o modelo quadrático, mais simples e linear, se ajustou melhor aos dados que a 1ª Lei de Mitscherlich.

		MÉTODOS							
		GAUSS- NEWTON	MARQUARDT	DFP	BFGS	SAS	L1 Original	L1 Modificado	
PARÂMETROS	a	33,9351	33,9351	33,9351	33,9351	33,9351	33,7428	33,0800	
	b	106,7157	106,7157	106,7157	106,7157	106,7157	104,8845	108,4999	
	c	-29,1008	-29,1008	-29,1008	-29,1008	-29,1008	-27,9936	-29,4999	
SQRes		31,9835	31,9835	31,9835	31,9835	31,9835	41,2803	40,5572	
$\sum e $		17,8773	17,8773	17,8773	17,8773	-	17,0797	17,1400	
Iterações		2	3	4	4	2	100	100	
Tempo (seg)		0,82	1,32	4,94	5,28	-	412,11	331,81	
R^2		0,997	0,997	0,997	0,997	-	-	-	
F		15900	15900	15900	15900	-	-	-	
Normalidade		Não	Não	Não	Não	-	-	-	
Ponto Influyente		1º e 13º	1º e 13º	-	-	-	-	-	
Ponto Aberrante		-	-	-	-	-	-	-	
Homo/Heterocedasticidade		Homo	Homo	Homo	Homo	-	-	-	

QUADRO 2. Resumo das Análises dos Dados de SILVA (1978) Usando Parábola

FIGURA 7. Gráficos das regressões L2 para os dados de SILVA (1978) usando parábola.

Original

Modificado

FIGURA 8. Gráfico das regressões L_1 para os dados de SILVA (1978) usando Parábola.

4.3.3. Modelo : 2ª Lei de MITSCHERLICH

Partindo das estimativas iniciais fornecidas pela autora :

$$\theta_0 = \begin{bmatrix} 230 \\ 0,43 \\ 0,05 \\ 0,122 \end{bmatrix}$$

obtem-se os resultados apresentados nos Quadros 3 e 4.

Observando o Quadro 3 (máximo de 25 iterações) nota-se que os resultados são discrepantes, devido a não convergência do processo iterativo (o algoritmo de GAUSS-NEWTON falha com *overflow*, isto é, alguma operação excede 10^{308}).

Uma observação : o algoritmo DFP diz que "converge" na 7ª iteração com SQRes = 717, 4752 ; tal resultado não é coerente, e assim efetuou-se mais 25 iterações DFP, obtendo-se :

$$\hat{\theta} = \begin{bmatrix} 307,5440 \\ 0,2785 \\ 0,1774 \\ 0,0303 \end{bmatrix}$$

com SQRes = 69,38 e o processo não convergiu.

Na Figura 9 apresenta-se os gráficos das funções ajustadas para 25 iterações.

No Quadro 4 (máximo de 100 iterações) os resultados ainda são discrepantes, devido a não convergência do processo iterativo; naturalmente o algoritmo de GAUSS-NEWTON falha com *overflow*.

Após a "convergência" na 7ª iteração do DFP, efetuou-se mais 100 iterações DFP (utilizando como θ_0 dados do Quadro 4), obtendo-se :

$$\hat{\theta} = \begin{bmatrix} 712,4240 \\ 0,3227 \\ 0,0668 \\ 0,0399 \end{bmatrix}$$

com $SQRes = 37,90$ sem haver convergência do processo iterativo.

Uma última tentativa de 1.000 iterações (por Marquardt) foi feita, e o processo não convergiu.

Observou-se que aumentando o número de iterações o parâmetro a cresce indefinidamente, enquanto o parâmetro c decresce; b e d se mantêm praticamente constantes.

Outras tentativas com novos valores iniciais foram feitas, e nem assim obteve-se convergência.

Na Figura 11 apresenta-se os gráficos das funções ajustadas para 100 iterações.

É prudente notar que as estatísticas fornecidas pelo SAS são distintas das obtidas pelo software apresentado.

Como a Regressão de Mínimos Quadrados (L_2) não converge não tem sentido falar em Regressão L_1 (por Schlossmacher), já que a 1ª iteração L_1 é uma regressão L_2 ; mesmo assim apresenta-se os resultados obtidos nos Quadros 3 e 4.

Os gráficos das regressões L_1 estão nas Figuras 10 e 12.

O mais curioso é que SILVA (1978) obteve na convergência :

$$\hat{\theta} = \begin{bmatrix} -151,206 \\ 0,46 \\ -0,212 \\ 0,070 \end{bmatrix}$$

utilizando o Método de HARTLEY (1961).

Partindo-se das estimativas finais da autora e usando o Método de GAUSS-NEWTON, o programa converge em 4

iterações, confirmando os resultados da autora: obteve-se também :

- $SQ_{Res} = 12,9$
- $R^2 = 0.999$
- $F = 26400$
- 12 e 132 são pontos influentes
- Erros não normais
- Erros são homocedásticos

dentre outras estatísticas.

A partir dos resultados obtidos, sugere-se que a 2ª Lei de Mitscherlich deve ser usada com cautela. MAGNANI (1991) sugere uma reparametrização nesse modelo.

MÉTODOS								
		GAUSS- NEWTON	MARQUARDT	DFP	BFGS	SAS	L1 Original	L1 Modificado
PARÂMETROS	a	-	243,6887	148,0627	244,8732	260,8949	3620,4625	1949,4039
	b	-	0,2612	0,1127	0,2596	0,2992	0,3575	0,3742
	c	-	0,2405	0,6604	0,2375	0,2189	0,0119	0,0220
	d	-	0,02487	0,0135	0,0249	0,0283	0,0431	0,0404
SQRes		-	92,4427	717,4752	94,7190	168,96	36,4886	48,5135
$\sum e $		-	29,4410	86,2000	28,2981	-	15,3440	16,4834
Iterações		-	25	7	25	25	25	25
Tempo (seg)		-	30,48	5,22	33,40	-	305,60	280,29
R^2		-	0,993	0,943	0,993	-	-	-
F		-	3710	476	3620	-	-	-
Normalidade		-	Não	Não	Não	-	-	-
Ponto Influyente		-	-	-	-	-	-	-
Ponto Aberrante		-	-	-	-	-	-	-
Homo/Heterocedasticidade		-	Homo	Homo	Homo	-	-	-

QUADRO 3. Resumo das Análises dos Dados de SILVA (1978) Usando a 2ª Lei de Mitscherlich - Max. 25 Iterações

FIGURA 9. Gráficos das regressões L2 para os dados de SILVA (1978) usando a 2ª Lei de Mitscherlich com 25 iterações.

Original

Modificado

FIGURA 10. Gráfico das regressões L1 para os dados de SILVA (1978) usando a 2ª Lei de Mitscherlich com 100 iterações.

MÉTODOS								
		GAUSS- NEWTON	MARQUARDT	DFP	BFGS	SAS	L1 Original	L1 Modificado
PARÂMETROS	a	-	456,3067	148,0627	1058,3556	994,0432	36338,4600	2606,0624
	b	-	0,3060	0,1127	0,3318	0,3275	0,3616	0,3787
	c	-	0,1102	0,6604	0,0437	0,0474	0,0011	0,0163
	d	-	0,0358	0,0135	0,0422	0,0433	0,0442	0,0408
SQRes		-	47,8118	717,4752	33,3604	39,8225	35,1956	48,2340
$\sum e $		-	21,3668	86,2007	18,3218	-	14,9236	16,3773
Iterações		-	100	7	100	97	100	100
Tempo (seg)		-	67,78	5,22	73,93	-	677,51	1165,46
R^2		-	0,996	0,943	0,997	-	-	-
F		-	7180	476	10200	-	-	-
Normalidade		-	Não	Não	Não	-	-	-
Ponto Influyente		-	-	-	-	-	-	-
Ponto Aberrante		-	-	-	-	-	-	-
Homo/Heterocedasticidade		-	Homo	Homo	Homo	-	-	-

QUADRO 4. Resumo das análises dos Dados de SILVA (1978) Usando a 2ª Lei de Mitscherlich - Max. 100 Iterações

FIGURA 11. Gráficos das regressões L_2 para os dados de SILVA (1978) usando a 2ª Lei de Mitscherlich com 100 iterações.

Original

Modificado

FIGURA 12. Gráfico das regressões L₁ para os dados de SILVA (1978) usando a 2ª Lei de Mitscherlich com 100 iterações.

4.4. Análise dos Dados de LANDRIANI et alii (1983)

Partindo-se das estimativas iniciais fornecidas pelos autores :

$$\underline{\theta}_0 = \begin{bmatrix} 0,1 \\ 0,3 \\ 1 \\ 0,1 \end{bmatrix}$$

apenas o SAS consegue convergir para as estimativas finais dadas pelos autores: o Método BFGS converge para os valores apresentados no Quadro 5. Os demais métodos param com *overflow*.

Partindo-se de :

$$\underline{\theta}_0 = \begin{bmatrix} 0,1 \\ 0,1 \\ 0,1 \\ 0,1 \end{bmatrix}$$

chega-se aos resultados apresentados no Quadro 5, onde nota-se que apenas o Método de GAUSS-NEWTON falha (*overflow*).

Para a regressão L_2 (Mínimos Quadrados), apenas o Método BFGS apresenta "resultados discrepantes, permutando os valores dos parâmetros a com c e b com d ".

Usando o Método de GAUSS-NEWTON, com as estimativas iniciais iguais as estimativas finais obtidas por MARQUARDT, o método converge e confirma as estimativas obtidas por MARQUARDT.

Da mesma forma, confirma-se os resultados obtidos para o Método BFGS.

É conveniente ressaltar que graficamente não se distingue as estimativas obtidas (Figura 13).

Pode-se ver também no Quadro 5 que os erros não são normais, são homocedásticos e não há pontos

influentes e/ou aberrantes.

Para regressão L_1 (desvios absolutos), o processo não convergiu com 100 iterações.

A Figura 14 apresenta os gráficos obtidos para a função de LANDRIANI et alli (Regressão L_1).

Nota : Limitou-se a regressão L_1 a 100 iterações devido ao tempo dispensado para execução dos cálculos que foram efetuados em um computador PC 386, 30 MHz.

MÉTODOS							
	GAUSS- NEWTON	MARQUARDT	DFP	BFGS	SAS	L1 Original	L1 Modificado
PARÂMETROS	a	2,7148	2,7149	0,5869	2,7139	3,0483	3,0024
	b	1,9121	1,9121	0,0770	1,9121	2,1224	2,0090
	c	0,5869	0,5869	2,7149	0,5869	0,6011	0,5850
	d	0,0770	0,0770	1,9121	0,0770	0,0783	0,07753
SQRes	-	2,2618	2,2618	2,2618	2,2618	2,3382	2,3271
$\sum e $	-	5,8676	5,8676	5,8676	-	5,7017	5,7254
Iterações	-	36	116	37	38	100	100
Tempo (seg)	-	98,16	544,14	247,38	-	5382,75	6585,73
R ²	-	0,971	0,971	0,971	-	-	-
F	-	799	799	799	-	-	-
Normalidade	-	Não	Não	Não	-	-	-
Ponto Influyente	-	Não	Não	Não	-	-	-
Ponto Aberrante	-	Não	Não	Não	-	-	-
Homo/Heterocedasticidade	-	Homo	Homo	Homo	-	-	-

QUADRO 5. Resumo d **álises dos Dados de LANDRIANI et alli (1983)**

FIGURA 13. Gráficos das regressões L2 para os dados de LANDRIANI *et alii* (1983).

FIGURA 14. Gráfico das regressões L1 para os dados de LANDRIANI *et alii* (1983).

5. CONCLUSÕES E SUGESTÕES PARA OUTRAS PESQUISAS

O sistema proposto para análise de regressão não-linear, (extensivo aos casos linear e intrinsecamente linear) tem se mostrado satisfatório, com resultados precisos (conforme observado nos problemas de teste) e relativa facilidade de uso (por ser baseado em menus).

Dentre os métodos propostos para a Regressão de Mínimos Quadrados (L_2), a saber : GAUSS-NEWTON, MARQUARDT, DFP, BFGS, apenas o Método de GAUSS-NEWTON deve ser usado com cautela, pois é numericamente instável.

Não se pode afirmar que para regressão não-linear exista um método melhor (de ajuste) que outro, e sugere-se que cada caso deve ser tratado isoladamente.

Apresenta-se também algumas inovações tais como a análise dos resíduos e os testes sobre as pressuposições dos modelos, que permitem ao estatístico concluir sobre a adequação ou não do conjunto modelo/dados.

Quanto a Regressão de Desvios Absolutos (L_1) não se dispõe de um pacote para checar sua precisão; usou-se para isso alguns exemplos da literatura, extraídos de KARST (1958) e STANGENHAUS (1988); o pacote aqui apresentado, reproduziu as estimativas obtidas pelos autores (caso linear apenas).

Para outras pesquisas sugere-se :

- A implementação do Método Híbrido de XU
- O estudo e a implementação de regressões com restrição nos parâmetros.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAMOWITZ, M. & STEGUN, I.A., ed. **Handbook of mathematical functions.** Washington, National Bureau of Standards, 1972. 1046p. (Applied Mathematics ser., 55).
- AL-BAALI, M. Methods for nonlinear least squares. Dundee, 1984. 115p. (Ph-D. - University of Dundee).
- ANSCOMBE, F.J. Examination of residuals. In: BERKELEY SYMPOSIUM OF MATHEMATICAL STATISTICS AND PROBABILITY, 4., Berkeley, 1960. **Proceedings,** Berkeley, University of California, 1961. V.1, p.1-36.
- ARMSTRONG, R.D. & FROME, E.L. A comparison of two algorithms for absolute deviation curve fitting. **JASA,** Boston, **71** (354): 328-30, 1976.
- BARD, Y. **Nonlinear parameter estimation,** New York, Academic Press, 1974. 341p.
- BATES, D.M. & WATTS, D.G. **Nonlinear regression analysis and its applications,** New York, John Wiley, 1988. 364p.
- BAZARAA, M.S. & SHETTY, C.M. **Nonlinear programming theory and algorithms.** New York, John Wiley, 1979. 560p.

- BEALE, E.M.L. Confidence regions in nonlinear estimation. **Journal of the Royal Statistical Society. Ser.B.** London, 22: 41-76, 1960.
- BOOTH, G.W.; BOX, G.E.P.; MULLER, M.E.; PETERSON, T.I. **Forecasting by generalized regression methods, nonlinear estimation**, s.l., IBM Corp., 1959. n.p. (IBM Share Program, 687 WLNL1.).
- BOX, G.E.P. & MULLER, M.E. A note on the generation of random normal deviates. **Annals of Mathematical Statistics**, Ann Arbor, 29: 610-1, 1958.
- BRODLIE, K.W. Unconstrained minimization. In: JACOBS, D., ed. **The state of the art in numerical analysis**, New York, Academic Press, 1977. cap.3, p.229-68.
- BROYDEN, C.G. A class of methods for solving nonlinear simultaneous equations. **Mathematics of Computation**, Menasha, 19: 577-93, 1965.
- BROYDEN, C.G. Quasi-newton methods and their application to function minimization. **Mathematics of Computation**, Menasha, 21: 368-81, 1967.
- BROYDEN, C.G. The convergence of a class of double-rank minimization algorithms. **Journal of the Institute of Mathematics its Applications**, London, 6: 76-90, 222-31, 1970.
- CAMPOS, H. **Estatística experimental não-paramétrica**. 4.ed. Piracicaba, ESALQ, 1983. 349p.
- CHIAICCHIO, E.J. **Avalia-String: um interpretador de fórmulas**. Piracicaba, ESALQ/Depto. Mat. Estat., 1990. 38p.

- CHIAICCHIO, E.J. **Um algoritmo para estimativa de mínimos quadrados de parâmetros não-lineares.** Piracicaba, ESALQ/Depto. Matem. Estat., 1991. 60p.
- COOK, R.D. & WEISBERG, S., colab. **Residuals and influence in regression.** New York, Chapman and Hall, 1982. 230p. (Monographs on Statistics and Applied Probability).
- COOK, R.D. & WEISBERG, S. Diagnostics for heteroscedasticity in regression. **Biometrika**, Cambridge, **70** (1): 1-10, 1983.
- CORTARELLI, A. Estudo da curva de crescimento de machos de raça nelore, através de quatro modelos estocásticos. Jaboticabal, 1973. 179p. (Doutorado - Faculdades Médica, Veterinária e Agrícola de Jaboticabal/UNESP).
- CURRY, H.B. The method of steepest descent for non-linear minimization problems. **Quarterly Applied Mathematics**, Providence, **2**: 258-61, 1944.
- DAVIDON, W.C. **Variable metric method for minimization.** s.l., AEC, 1959. (A.E.C. Research and Development Report ANL, 5990).
- DAVIS, D.S. **Nomography and empirical equations.** New York, Reinhold Publ., 1943.
- DENNIS JR., J.E. Some computational techniques for the nonlinear least squares problem. In: BYRNE, G.D. & HALL, C.A., ed. **Numerical solution of systems of nonlinear equations.** New York, Academic Press, 1973.
- DENNIS JR., J.E. & SCHNABEL, R.B. **Numerical methods for unconstrained optimization and nonlinear equations.** Englewood Cliffs, Prentice Hall, 1983. 378p.

- DENNIS JR., J.E.; GAY, D.M.; WELSCH, R.E. An adaptive nonlinear least squares algorithm. **ACM Transactions on Mathematical Software**, New York, **7**: 348-68, 1981.
- DIELMAN, T. & PFAFFENBERGER, R. LAV (Least Absolute Value) estimation in the regression model : a review. **TIMS Studies in the Management Sciences**, Providence, **19**: 31-52, 1982.
- DONALDSON, J.R. & SCHNABEL, R.B. Computacional experience with confidence regions and confidence intervals for nonlinear least squares. **Technometrics**, Richmond, **29**(1): 67-82, 1987.
- DRAPER, N.R. & SMITH, H. **Applied regression analysis**, New York, John Wiley, 1966. 407p.
- DRAPER, N.R. & SMITH, H. **Applied regression analysis**, 2.ed. New York, John Wiley, 1981. 710p.
- FELTRIN, M.A.V. A análise harmônica no estudo das precipitações pluviiais no município de Bandeirantes, PR. Piracicaba, 1980. 102p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- FIGUEIREDO, G.C. **Teste baseado no número total de chorrilhos**, Piracicaba, ESALQ/Depto. Matem. Estatíst., 1977. 10p.
- FIGUEIRÔA, M.L. Inferência sobre parâmetros da função de Cobb-Douglas. Piracicaba, 1986. 127p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).

- FLANAGAN, P.D.; VITALE, P.A.; MENDELSON, J. A numerical investigation of several one-dimensional search procedures in nonlinear regression problems. **Technometrics**, Richmond, **11**(2): 265-84, 1969.
- FLETCHER, R. A new approach to variable metric algorithms. **The Computer Journal**, London, **13**: 317-22, 1970.
- FLETCHER, R. **A modified Marquardt subroutine for nonlinear least squares**. London, A.E.R.E., 1971. n.p. (A.E.R.E. Harwell Report AERE, R 6799).
- FLETCHER, R. **Practical methods of optimization**, 2.ed. New York, John Wiley, 1987. 436p.
- FLETCHER, R. & POWELL, M.J.D. A rapidly convergent descent method for minimization. **Computer Journal**, London, **6**: 163-8, 1963.
- GALLANT, A.R. **Nonlinear statistical models**, New York, John Wiley, 1987. 610p.
- GILL, P.E. & MURRAY, W. Quasi-Newton methods for unconstrained optimization. **Journal of the Institute of Mathematics its Applications**, London, **9**: 91-108, 1972.
- GOLDFARB, D. A family of variable metric methods derived by variational means. **Mathematics of Computation**, Menasha, **24**: 23-6, 1970.
- GUTTMAN, I. & MEETER, D.A. On Beale's measures of nonlinearity. **Technometrics**, Richmond, **7**: 623-37, 1965.
- HARTLEY, H.O. The estimation of non-linear parameters by internal least squares. **Biometrika**, Ser 1, Cambridge, **35**: 32-45, 1948.

- HARTLEY, H.O. The modified Gauss-Newton method for the fitting of nonlinear regression function by least squares. **Technometrics**, Richmond, **3**: 269-80, 1961.
- HARTLEY, H.O. Exact confidence regions for the parameters in nonlinear regression law. **Biometrika**, Cambridge, **51**: 347-53, 1964.
- HARTLEY, H.O. & BROOKER, A. Non-linear least squares estimations. **Annals of Mathematical Statistics**, Ann Arbor, **36**: 638-50, 1965.
- HEWLETT PACKARD, CO. **Curve fitting pac owner's manual for the HP-71**. Corvallis, 1984a. 125p.
- HEWLETT PACKARD, CO. **HP-71 Reference manual**, 3.ed. Corvallis, 1984b. 417p.
- HEWLETT PACKARD, CO. **HP-48SX owners manual**, Corvallis, 1989. pt.1.
- HOFFMANN, R. & VIEIRA, S. **Análise de regressão**, 2.ed. S. Paulo, Hucitec; EDUSP, 1987. 379p.
- IEMMA, A.F. **Matrizes para estatística**; um texto para profissionais de ciências aplicadas. Piracicaba, ESALQ/Depto. Matem. Estat., 1990. 339p.
- JOHNSON, E.E. Empirical equations for approximating tabular F values. **Technometrics**, Richmond, **15**(2): 379-84, 1973.
- KARST, O.J. Linear curve fitting using least deviations. **JASA**, Boston, **53**: 118-32, 1958.
- KENNEDY JR., W.J. & GENTLE, J.E. **Statistical computing**, New York, Marcel Dekker, 1980.

- KNUTH, D. Semi-numerical algorithms. In: _____. **The art of computer programming.** Redwood City, Addison Wesley, 1969. v.2, sect. 3.4.4.
- KOLB, W.M. **Curve fitting for programmable calculators.** Bowie. 1982. 127p.
- LANDRIANI, G.S.; GUARDABASSO, V.; ROCCHETTI, M. NL-FIT: a microcomputer program for non-linear fitting. **Computer Programs in Biomedicine**, Amsterdam, **16**: 35-42, 1983.
- LEVENBERG, K. A method for the solution of certain nonlinear problems in least squares. **Quarterly Applied Mathematics**, Providence, **2**: 164-8, 1944.
- MACHADO, A.A. **Modelos lineares.** Piracicaba, s.ed., 1991. 275p.
- MAGNANI, R. Alguns aspectos da aplicação da segunda Lei de Mitscherlich e da equação de regressão quadrática à maturação da cana-de-açúcar. Piracicaba, 1985. 73p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- MAGNANI, R. Avaliação de propriedades estatísticas da segunda Lei de Mitscherlich. Piracicaba, 1991. 152p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- MARQUARDT, D.W. An algorithm for least squares estimation of nonlinear parameters. **Journal Society for Industrial and Applied Mathematics**, Philadelphia, **11**(2): 431-41, 1963.

- MELO, F.I.O. Aplicação do método modificado de Gass-Newton para estimar os parâmetros da equação de Mitscherlich. Piracicaba, 1976. 74p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- MIRSHAWKA, V. **Probabilidades e estatística para engenharia**. São Paulo. Nobel, 1983. v.2, 483p.
- NARULA, S.C. & STANGENHAUS, G. **Análise de regressão L_1** . São Paulo. IMPA, 1988. 67p.
- NETER, J. & WASSERMAN, W. **Applied linear statistical models, regression, analysis of variance and experimental designs**. Homewood, Richard D. Irwin, 1974. 842p.
- POLAK, E. **Computational methods in optimization; a unified approach**. New York, Academic Press, 1971.
- POWELL, M.J.D. A hybrid method for nonlinear equations. In: RABINOWITZ, P., ed. **Numerical methods for nonlinear algebraic equations**. London, Gordon and Breach, 1970.
- PRUNTY, L. Curve fitting with smooth functions that are piecewise-linear in the limit. **Biometrics**, Washington, **39**: 857-66, 1983.
- QUINTELA, L.A. & PIMENTEL, M.S.L. **Clipper com linguagem C**. São Paulo, M. Graw-Hill, 1990. 81p.
- RATKOWSKY, D.A. **Nonlinear regression modelling**. New York, Marcel Dekker, 1983. 276p.
- REYES, A.E.L. & CHIACCHIO, E. J. Sistema para análise de regressão - manual do usuário. CIAGRI/USP. No Prelo.

- SAS INSTITUTE INC. SAS user's guide; statistics Proc. NLIN. 5.ed. Cary, 1985. cap.25. p.575-606.
- SCHLOSSMACHER, E.J. An iterative technique for absolute deviations curve fitting. *JASA*, Boston, **68**(344): 857-9, 1973.
- SEAL, H.L. The historical development of the Gauss linear model. *Biometrika*, Cambridge, **54**: 1-24, 1967.
- SEBER, G.A.F. & WILD, C.J. *Nonlinear regression*, New York. John Wiley, 1989. 768p.
- SHANNO, D.F. Conditioning of Quasi-Newton methods for function minimization. *Mathematics on Computation*, Menasha, **24**: 647-56, 1970.
- SILVA, M.A.P. Segunda aproximação de Mitscherlich aplicada à adubação mineral. Piracicaba, 1978. 96p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- SILVA, M.A.P. Métodos de estimação dos parâmetros do modelo $Y = A(1 - 10^{-cx}).10^{-kx^2} + \epsilon$ aplicado a dados de maturação da cana-de-açúcar. Piracicaba, 1980. 85p. (Doutorado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- SILVEIRA JR., P. Estudo de alguns modelos exponenciais no crescimento de bovinos da raça Ibagé. Piracicaba, 1976. 174p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- SMITH, F.B. & SHANNO, D.F. An improved Marquardt procedure for nonlinear regressions. *Technometrics*, Richmond, **13**: 63-74, 1971.

- SOFTCAD INFORMATICA. **Clipper com gráficos usando CLBC.** São Paulo. Livros Técnicos e Científicos. 1991. 187p.
- SPENCE, R. **Clipper 5.0 release 5.01.** São Paulo. Makron Books, 1991. 767p.
- TABATA, T. & ITO, R. Effective treatment of the interpolation factor in Marquardt's non-linear least squares fit algorithm. **The Computer Journal**, London, **18**: 250-1. 1975.
- TISHLER, A. & ZANG, I. An absolute deviations curve-fitting algorithm for nonlinear models. **TIMS Studies in the Management Sciencies**, Providence, **19**: 81-103. 1982.
- TURNER, M.E.; MONROE, R.S.; LUCAS, H.L. Generalized asymptotic regression and nonlinear data analysis. **Biometrics**, Washington, **17**: 120-43, 1961.
- UDO, M.C.T. Estudo de funções que se ajustam à maturação da cana-de-açúcar. Piracicaba, 1983. 74p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- VESPUCCI, M.T. An efficient code for minimization of highly nonlinear and large residual least squares functions. **Optimization**, Berlin, **18**(6): 825-55, 1987.
- VITALE, P. & TAYLOR G. A note on the application of Davidon's method to nonlinear regression problems. **Technometrics**, Richmond, **10**(4): 843-9, 1968.
- WAGNER, H. M. Linear programming techniques for regression analysis. **JASA**, Boston, **54**: 206-12, 1959.
- WOLFRAM RESEARCH INC. MATHEMATICA; a system for doing mathematics by computer. **BYTE**, Perterborough, Aug. 1990.

XU, C. Hybrid methods for nonlinear least squares and related problems. Dundee. 1986. 205p. (Ph-D. - University of Dundee).

ANEXO 1. Limites Superiores para o Teste de LILLIEFORS,
($\alpha = 0,05$)

n	$10^3 \cdot D$
4	381
5	337
6	319
7	300
8	285
9	271
10	258
11	249
12	242
13	234
14	227
15	220
16	213
17	206
18	200
19	195
20	190
21	186
22	182
23	179
24	176
25	173
26	170
27	168
28	166
29	164
30	161
> 30	$886/\sqrt{n}$

Fonte : CAMPOS (1983)

ANEXO 2. Limites da Distribuição de χ^2 ($\alpha = 0,05$)

gl	$10^2 \cdot \chi^2$
1	384
2	599
3	781
4	949
5	1107
6	1259
7	1407
8	1551
9	1692
10	1831
11	1968
12	2103
13	2236
14	2368
15	2500
16	2630
17	2759
18	2887
19	3014
20	3141
21	3267
22	3392
23	3517
24	3642
25	3765
26	3889
27	4011
28	4134
29	4256
30	4377
40	5576
50	6750
60	7908
70	9053
80	10188
90	11314
100	12434

Fonte : CAMPOS (1983)

ANEXO 3. Limites da Distribuição Bilateral t de Student
 $(\alpha = 0,05)$

gl	$10^2 \cdot t$
1	1271
2	430
3	318
4	278
5	257
6	245
7	236
8	231
9	226
10	223
11	220
12	218
13	216
14	214
15	213
16	212
17	211
18	210
19	209
20	209
21	208
22	207
23	207
24	206
25	206
26	206
27	205
28	205
29	205
30	204
50	201
100	198
∞	196

Fonte : MOSTELLER (1983)

ANEXO 4. Limites Críticos de r para o Teste de CHORRILHOS
 $(\alpha = 0,05)$

		max																			
n2	n1	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	r inf	r sup		
		2	-	-	-	-	-	-	-	2	2	2	2	2	2	2	2			2	
3	-	-	-	-	-	-	-	-	6	6	6	6	6	6	6	6	6				
	-	-	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3				
4	-	-	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8				
	2	2	2	3	3	3	3	3	3	3	3	3	3	4	4	4	4				
5	9	9	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10				
	2	3	3	3	3	3	3	4	4	4	4	4	4	4	4	5	5				
6	10	10	11	11	12	12	12	12	12	12	12	12	12	12	12	12	12				
	-	3	3	3	4	4	4	4	4	5	5	5	5	5	5	5	6				
7	-	11	12	12	13	13	13	13	13	14	14	14	14	14	14	14	14				
	-	-	3	4	4	5	5	5	5	5	5	6	6	6	6	6	6				
8	-	-	13	13	14	14	14	14	14	15	15	15	16	16	16	16	16				
	-	-	-	4	5	5	5	5	6	6	6	6	6	6	7	7	7				
9	-	-	-	14	14	15	15	16	16	16	16	16	17	17	17	17	17				
	-	-	-	-	5	5	6	6	6	6	7	7	7	7	8	8	8				
10	-	-	-	-	15	16	16	16	16	17	17	18	18	18	18	18	18				
	-	-	-	-	-	6	6	7	7	7	7	7	8	8	8	8	9				
11	-	-	-	-	-	16	17	17	17	18	18	18	19	19	19	20	20				
	-	-	-	-	-	-	7	7	7	7	8	8	8	9	9	9	9				
12	-	-	-	-	-	-	17	18	18	19	19	19	20	20	20	21	21				
	-	-	-	-	-	-	-	7	8	8	8	8	9	9	9	9	10				
13	-	-	-	-	-	-	-	19	19	20	20	20	21	21	21	22	22				
	-	-	-	-	-	-	-	-	8	9	9	9	9	10	10	10	10				
14	-	-	-	-	-	-	-	-	20	20	21	21	22	22	22	23	23				
	-	-	-	-	-	-	-	-	-	9	9	10	10	10	10	11	11				
15	-	-	-	-	-	-	-	-	-	-	9	10	10	11	11	11	12				
	-	-	-	-	-	-	-	-	-	-	-	10	10	11	11	11	12				
16	-	-	-	-	-	-	-	-	-	-	-	-	11	11	11	12	12				
	-	-	-	-	-	-	-	-	-	-	-	-	23	24	25	25	25				
17	-	-	-	-	-	-	-	-	-	-	-	-	-	11	12	12	13				
	-	-	-	-	-	-	-	-	-	-	-	-	-	25	25	26	26				
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12	13	13				
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	26	26	27				
19	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	13	13				
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	27	27				
20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	14				
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	28				

Fonte : FIGUEIREDO (1977)

**ANEXO 5. Resultados Obtidos pelo Método de MARQUARDT e
SCHLOSSMACHER Modificado para os Dados de SILVA
(1978) Usando a 1ª Lei de Mitscherlich**

PARAMETROS

Estimativas

a: +1.368E+02

b: +1.924E-01

c: +5.896E-01

A N A L I S E

	CV	GL	SQ	QM	F
Regressao	003		+1.52E+05	+5.07E+04	+1.30E+03
Residuo	010		+3.90E+02	+3.90E+01	
Total	013		+1.52E+05		
Tot.Corrigido			+1.26E+04		

Iteracao: 031 ||GRD||= +4.5508E-11

R²= 0.969 RR= 0.963 Lambda= +1E-10

Teste de Hipotese
 Probabilidade > F : +3.330E-04
 H1: Modelo adequado !!

Convergiu !!
 Tempo (seg) : 40.64

Par.	Valor
a	136.89291562457
b	0.19241043011
c	0.58967930512

Intervalos de Confianca dos Parametros

Parametro	Lim.Inferior	Estimativa	Lim.Superior
a	+1.25318E+02	+1.36892E+02	+1.48467E+02
b	+6.65723E-02	+1.92410E-01	+3.18248E-01
c	+3.76805E-01	+5.89679E-01	+8.02553E-01

Fim de Arquivo

Matriz de correlacoes

Par. a b c
a +1.00E+00 +5.63E-01 -8.71E-01
b +5.63E-01 +1.00E+00 -8.13E-01
c -8.71E-01 -8.13E-01 +1.00E+00
Fim de Arquivo

Matriz de variancias e covariancias

Par. a b c
a +2.69E+01 +1.65E-01 -4.32E-01
b +1.65E-01 +3.19E-03 -4.38E-03
c -4.32E-01 -4.38E-03 +9.12E-03
Fim de Arquivo

Analise de Residuos

Y	Y	e	Erro Norm.	Diag.Projecao
+3.66E+01	+3.14E+01	+5.12E+00	0.82	+7.58E-01*
+5.36E+01	+5.65E+01	-2.94E+00	-0.47	+2.59E-01
+6.99E+01	+7.56E+01	-5.75E+00	-0.92	+2.12E-01
+8.62E+01	+9.02E+01	-4.01E+00	-0.64	+2.20E-01
+1.01E+02	+1.01E+02	-3.15E-01	-0.05	+2.01E-01
+1.10E+02	+1.09E+02	+2.24E-01	0.04	+1.62E-01
+1.19E+02	+1.16E+02	+2.77E+00	0.44	+1.26E-01
+1.28E+02	+1.21E+02	+6.86E+00	1.10	+1.08E-01
+1.32E+02	+1.24E+02	+7.11E+00	1.14	+1.13E-01
+1.33E+02	+1.27E+02	+5.25E+00	0.84	+1.39E-01
+1.31E+02	+1.29E+02	+1.08E+00	0.17	+1.80E-01
+1.29E+02	+1.31E+02	-2.57E+00	-0.41	+2.31E-01
+1.20E+02	+1.32E+02	-1.28E+01	-2.06*	+2.85E-01

Nota: * na coluna Erro Normalizado : possivel ponto aberrante (outliers)

* na coluna da Diag.Projecao : possivel ponto influente

Fim de Arquivo

Gráfico da Função

Verificacao das Pressuposicoes do Modelo

Teste de Chorrilhos - Aleatoriedade dos Erros

Numero de Chorrilhos : 4

Nro. de simbolos do tipo 1 : 7

Nro. de simbolos do tipo 2 : 6

Aceita-se H_0 : os erros sao aleatorios

Teste de Lilliefors - Normalidade dos Erros

Estatistica D : 0.988

Aceita-se H_1 : erros nao sao normais

Outras Estatisticas da Normalidade

Estatistica	Valor esperado	Valor obtido
Media	0.000	0.000
Mediana	-	0.224
Variancia	-	32.501
Assimetria	0.000	-0.669
Curtose	3.000	2.987

Teste de Cook & Weisberg : Homocedasticidade

Estatistica S : 0.470

Aceita-se H_0 : os erros sao homocedasticos

Teste de Anscombe : Homocedasticidade

Estatistica T : 0.849

Aceita-se H_0 : os erros sao homocedasticos

Fim de Arquivo

Regressao L1 Iteracao: 097

PARAMETROS		
	L1	L2
a:	+1.38E+02	+1.36E+02
b:	+1.58E-01	+1.92E-01
c:	+5.94E-01	+5.09E-01

menor SseI = +5.3533E+01
 atual SseI = +5.3533E+01

L1 Convergiu

Regressao L2

S.Quad.Total.....: +1.52E+05
 S.Quad.Total Corrig.: +1.26E+04
 S.Quad. Residuos.....: +3.98E+02
 S.Absoluto do Erro.....: +5.68E+01
 Q.Medio do Residuo.....: +3.98E+01
 Valor de F.....: 1381.68
 Coef. de Correlacao.: 0.969
 Coef. de Comparacao.: 0.963
 Lambda.....: +1E-12

Tempo (seg) : 461.82

<ESC>-Continua <F3>-Imprime <F6>-Graf.Feila <F7>-Graf.Inpr.

Par.	Valor
a	138.168
b	0.15834738142
c	0.59484649542

Gráfico da Função

... ou você usa um PC 486 DX2 60 Mhz ou ...