



Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Adriana de Souza Costa

Uso de medidas não lineares no ajuste do modelo exponencial

Campina Grande
Dezembro de 2012

Adriana de Souza Costa

Uso de medidas não lineares no ajuste do modelo exponencial

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Orientadora:

Ana Patricia Bastos Peixoto

Campina Grande

Dezembro de 2012

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

C837u

Costa, Adriana de Souza.

Uso de medidas não lineares no ajuste de modelo exponencial [manuscrito] / Adriana de Souza Costa. – 2012.

39 f. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2012.

“Orientação: Profa. Ma. Ana Patrícia Bastos Peixoto, Departamento de Estatística”.

1. Linearidade . 2.Delineamento experimental.
3.Medidas de curvatura. I. Título.

21. ed. CDD 551.519.76

Adriana de Souza Costa

Uso de medidas não lineares no ajuste do modelo exponencial

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 07/12/12

Banca Examinadora:

Ana Patricia Bastos Peixoto

Prof. Msc. Ana Patricia Bastos Peixoto
Orientadora

DNéstor

Profa. Dra. Divanilda Maia Esteves
Universidade Estadual da Paraíba

Ricardo Alves de Olinda

Prof. Dr. Ricardo Alves de Olinda
Universidade Estadual da Paraíba

Dedicatória

A DEUS, por tudo que mim proporcionou para que conseguisse chegar até aqui.

Aos meus pais, Assis e Marineide,

As minhas irmãs, Andréia, Mércia e Michelle,

por todo amor, cuidado, apoio, confiança e paciência.

Agradecimentos

Em primeiro lugar, a Deus, por ter me concedido a vida.

A toda a minha família, e em especial meu pai Assis e minha mãe Marineide, minhas irmãs, Andréia, Mércia e Michelle, por todo amor, carinho, apoio e dedicação, sem vocês nunca teria conseguido realizar mais este sonho.

A minha orientadora Ana Patricia Bastos Peixoto, pela paciência, amizade e dedicação.

A UEPB por proporcionar condições para meu aprendizado, durante esses quatro anos. Especialmente, aos professores Ricardo Alves, Juarez, Thiago, Diana Maia, Ana Cristina, Edwirde, Ruth que repassaram os seus conhecimentos e tanto contribuíram para minha formação.

Às minhas amigas Wanessa Luana, Arielly Arethuza, Valnely Melo, Nathielly Lima e Samara Rilda, por toda a amizade, confiança, incentivo, por todos os momentos vividos juntas que jamais esquecerei.

Gostaria de agradecer também a todos os meus amigos que sempre me deram forças para chegar até aqui, Josevandro, Gilmara, Jaqueline, Tarciana, Jéssica, Vanusa, Vilma, Dorinha Pereira, Joseane Santos, Alânia Raquel, Fábio Avelino, Fernanda Matias e Erasnilson, enfim a todos.

A todos vocês o meu muito obrigada.

Resumo

O presente estudo teve por objetivo avaliar estatisticamente as principais características do modelo não linear exponencial analisando-se as propriedades estatísticas dos estimadores do modelo realizou-se um estudo da não linearidade do modelo considerando-se as medidas de curvatura. O delineamento experimental utilizado foi o inteiramente casualizado, com 20 repetições, e os tratamentos foram dispostos em esquema fatorial com uma concentração de sacarose, três concentrações de sorbitol, mais uma testemunha, totalizando 10 tratamentos ou ensaios em branco, ver tabela 1. Em todos os tratamentos analisados foi verificado que o modelo apresenta um comportamento próximo do linear, pois os valores das medidas de curvaturas encontradas foram menores que 0,5, o que evidencia um bom ajuste segundo a metodologia utilizada.

Palavras-chaves: Linearidade, Sacarose, Curvaturas, Delineamento Experimental.

Abstract

The present study aimed to evaluate statistically the main characteristics of the non-linear exponential model and analyzed the statistical properties of estimators of the model was a study of non-linearity of the model considering the measures of curvature. The experiment was completely randomized with 20 replications and the treatments were arranged in factorial with a sucrose concentration, three concentrations of sorbitol, another witness, totaling 10 treatments or blanks. In all treatments was found that the analyzed model presents a near linear behavior because the values of the measures of curvatures found were lower than 0.5, which shows a good fit according to the methodology used.

Key-words: Linearity, Saccharose, Curvatures, Desing Experiment.

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 11
2	Fundamentação Teórica	p. 13
2.1	Marco Histórico	p. 13
2.2	Modelos de regressão linear	p. 14
2.3	Modelos de regressão não linear	p. 17
2.3.1	A geometria dos mínimos quadrados não lineares	p. 18
2.4	Medidas de não linearidade	p. 19
2.5	Medidas de curvatura de Bates e Watts	p. 21
2.6	Vício de Box	p. 23
3	Aplicação	p. 25
3.1	Material e métodos	p. 25
3.2	Resultados e discussão	p. 27
4	Conclusão	p. 34
5	Referências Bibliográficas	p. 35
	Apêndice A - Primeiro anexo	p. 36

Lista de Figuras

1	Representação da curva $(2^\beta \ 3^\beta)^T$ com β variando em espaçamentos iguais a 0,5 (RATKOWSKY, 1983).	p. 21
2	Representação da curva $(2^{\log\phi} \ , \ 3^{\log\phi})^T$ com ϕ variando em espaçamentos iguais a 1,0 (RATKOWSKY, 1983).	p. 23
3	Ajuste do modelo exponencial em experimentos de conservação <i>in vitro</i> de <i>Passiflora giberti</i> N.E. Brown	p. 29
4	Bandas de confiança para a resposta esperada e para estimativas do modelo exponencial em experimentos de conservação <i>in vitro</i> de <i>Passiflora giberti</i> N.E. Brown	p. 30

Lista de Tabelas

- 1 Dosagens de sacarose e sorbitol de acordo com os tratamentos, utilizadas no experimento de conservação *in vitro* de *Passiflora giberti* N.E.Brown. p. 25
- 2 Tamanho da parcela (X), forma da parcela e número de parcelas totais para os ensaios de uniformidade de plantas de *P. giberti* N. E. Brown aos 60 dias de conservação *in vitro*. p. 26
- 3 Estimativas (E) dos parâmetros (P) a e b baseadas na inferência por aproximação linear, erro padrão da estimativa (E.P.E), valor de p para o teste t , intervalo de confiança (IC) de 95% para os parâmetros, vício de Box e coeficiente de determinação (R^2) para o modelo exponencial, nos 10 tratamentos (T) analisados p. 28
- 4 Medidas de curvatura quadrática média (RMS) para o modelo $CV = aX^{-b}$. p. 31
- 5 Estimativas (E) dos parâmetros (P) a e b baseadas na inferência por aproximação linear, erro padrão da estimativa (E.P. E), valor de p para o teste t , intervalo de confiança (IC) de 95% para os parâmetros, vício de Box e coeficiente de determinação (R^2) para o modelo exponencial, nos 10 tratamentos (T) analisados p. 32

1 Introdução

A análise de regressão consiste em verificar a existência da relação funcional significativa entre variáveis, ou seja, obter uma equação que explique a variação da variável dependente pela variação dos níveis da variável independente (CHARNET et al., 1999). Os modelos de regressão utilizados são classificados em duas categorias mais gerais: os modelos de regressão lineares e os não lineares. A diferença primordial desses dois modelos está na forma como os parâmetros aparecem nos mesmos, se os parâmetros aparecem de forma linear na equação matemática que determina a relação entre as variáveis, ele será classificado na categoria dos lineares, caso contrário na dos não lineares.

Nos modelos de regressão lineares e não lineares o critério de estimação mais comumente utilizado é o critério de mínimos quadrados. Quando o modelo é linear, o estimador de mínimos quadrados tem uma fórmula explícita, entretanto no caso não linear, isso geralmente não ocorre e para obter estimativas desses parâmetros deve-se recorrer a procedimentos iterativos, por exemplo, o método de Gauss-Newton. A inferência nos modelos não lineares ocorre por aproximação em Série de Taylor na região próxima às estimativas, e essa aproximação pode ser considerada apropriada ou não dependendo do modelo a ser estudado, do delineamento experimental e do conjunto de dados. Os estimadores dos parâmetros em um modelo de regressão linear possuem propriedades ótimas sob a suposição de erros normais, tais como: normalidade, não viés, eficiência, variância mínima, intervalos de confiança exatos, etc. Entretanto, no caso dos modelos não lineares essas propriedades só são válidas assintoticamente. do

Quando o tamanho da amostra cresce a aproximação para um modelo linear vai se tornando cada vez melhor, no entanto, para amostras pequenas é necessário quantificar essa aproximação. Bates e Watts (1980), quantificaram tal aproximação utilizando-se do conceito geométrico de curvatura, que pode servir, por exemplo, na escolha entre dois modelos considerados em um determinado experimento. Baseando-se nos conceitos de curvatura, Bates e Watts (1980) definem medidas quantitativas de não linearidade. Essas medidas são definidas a partir da especificação do modelo e do conjunto de dados,

constituído-se uma ferramenta de grande valor para a avaliação da não linearidade. O uso dessas medidas nem sempre é explorado em trabalhos práticos na literatura brasileira, sendo assim, o objetivo principal deste trabalho é apresentar um modelo não linear e investigar se o grau de não linearidade deste modelo é pequeno o suficiente para justificar a utilização dos resultados usuais da teoria dos modelos lineares como aproximações para os não lineares, para tanto será utilizado o *software* R, para obtenção dessas medidas.

2 Fundamentação Teórica

O conteúdo desta seção relata os principais aspectos da utilização dos modelos não lineares e medidas que quantificam a não linearidade dos modelos, utilizando-se de artigos práticos e teóricos relacionados ao objetivo da pesquisa.

2.1 Marco Histórico

Por meio da necessidade de se desvendar os mistérios da hereditariedade, Mendel (1822-1884) chegou a um estudo matemático-probabilístico para explicar as características das ervilhas. Outros pesquisadores também contribuíram para o estudo, concluindo que os caracteres herdados de uma geração para a outra o faziam por intermédio de fatores particulares.

Para Mendel, a partir dos resultados dos cruzamentos com ervilhas verdes e amarelas, percebia-se que algumas características eram dominantes sobre outras, e tudo era explicado com cálculos probabilísticos. Ele observou que, quando os pais eram mais altos do que a média, os filhos tendiam a ser menores do que eles e, quando os pais eram mais baixos que a média, os filhos tendiam a ser maiores do que eles. O Sir Francis Galton (1822-1911), antropólogo, biometrista e estatístico inglês, em seus estudos sobre hereditariedade, comparava as alturas dos filhos com as dos pais. Comparando as séries de alturas de pais e filhos, ele observou que a altura dos filhos de pais com alturas extremas tendia a regredir (voltar) na direção da altura média (da população de pais), a medida que as gerações se sucediam. Ao usar nesse problema a técnica de interpolar uma reta a um conjunto de dados para fazer previsões, o nome Regressão ficou ligado a esta técnica.

Um modelo é classificado como não linear se pelo menos uma das derivadas parciais da função esperança em relação ao parâmetro é função de parâmetros desconhecidos. Até o início da década de 70, as principais técnicas desenvolvidas para os modelos de regressão não linear se restringiam à suposição de normalidade para a variável resposta. E, mesmo

após a extensão da distribuição da variável resposta para a família exponencial de distribuições, os modelos normais não lineares (MNNL) continuaram recebendo um tratamento especial, surgindo diversos artigos científicos na mesma década e em décadas posteriores, destacando-se o livro de Ratkowsky (1983), que descreve vários MNNL, segundo diversos aspectos.

2.2 Modelos de regressão linear

A análise de regressão é uma das técnicas de estimação mais usadas na estatística, com aplicações em diversas áreas. Um dos objetivos do modelo de regressão é explicar a existência da relação funcional de uma ou mais variáveis de interesse em função de outras variáveis explicativas. Outros objetivos da análise de regressão são estimação de parâmetros, predição e controle, assim é possível determinar como duas ou mais variáveis se relacionam, além de estimar a função que determina a relação entre as variáveis e usar a equação ajustada para prever valores da variável dependente.

Sempre é interessante conhecer os efeitos que algumas variáveis exercem, ou parecem exercer, sobre outras. Mesmo que não exista relação casual entre as variáveis pode-se relacioná-las por meio de uma expressão matemática, que pode ser útil para se estimar o valor de uma das variáveis quando se conhece os valores das outras (HOFFMANN, 1998).

Genericamente, tais relações funcionais podem ser representadas por

$$\mathbf{Y} = f(X_1, X_2, \dots, X_n), \quad (2.1)$$

em que Y representa a variável dependente e os X_i com $i = 1, 2, \dots, n$ representam as variáveis explanatórias. É de se esperar que essa variável \mathbf{Y} sofra influência dos valores de várias outras variáveis independentes (X_1, X_2, \dots, X_n) , e que exista uma função que expresse tal dependência. Geralmente não consegue-se observar todas as n variáveis, observando-se um número $i < n$ delas. Logo, o modelo (2.1) pode ser reescrito como

$$\mathbf{Y} = f(X_1, X_2, \dots, X_k) + h(X_{k+1}, X_{k+2}, \dots, X_k),$$

em que, todas as variáveis $X_{k+1}, X_{k+2}, \dots, X_k$, serão consideradas como casuais (erros) e associadas a uma variável aleatória U .

O objetivo de um modelo de regressão é então, encontrar uma relação entre os dados de forma que uma variável possa ser definida em relação a outra. Sabe-se, que a relação encontrada por estes modelos não é perfeita como uma função matemática, apresentando

erros nos valores observados. Estas diferenças são chamadas de erros de dispersão. O modelo de regressão será então mais adequado quanto menores forem estes erros de dispersão. Genericamente, escreve-se um modelo de regressão como

$$\mathbf{Y} = f(X_1, X_2, \dots, X_n) + U.$$

Para Bates e Watts (1988), a regressão linear fornece estimativas e outros resultados de inferências para os parâmetros $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ no modelo

$$Y_n = (X_{n1}, X_{n1}, \dots, X_{n1}) \boldsymbol{\beta} + \varepsilon_n, \quad (2.2)$$

em que $(x_{n1}, x_{n2}, \dots, x_{np}) \boldsymbol{\beta}$ depende dos parâmetros e das variáveis preditoras ou variáveis regressoras x_{np} e ε_i é o erro aleatório.

O modelo linear (2.2) pode ser escrito na forma matricial como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}. \quad (2.3)$$

Especificamente os métodos que estabelecem relações lineares para os parâmetros de um modelo entre uma ou mais variáveis são denominados métodos de regressão linear. Seja o modelo

$$\mathbf{Y}_i = \beta_0 + \beta_1 \mathbf{X}_i + \varepsilon_i, \quad i = 1, \dots, n \quad (2.4)$$

Se for considerado o modelo funcional apresentado de forma geral na Equação (2.4), então o modelo de regressão é denominado de linear simples.

Nessa forma de modelagem, as derivadas parciais da variável dependente Y_i em relação aos parâmetros β_0 e β_1 não são funções dos próprios parâmetros. Por essa razão, os modelos são considerados formalmente por modelos lineares. Nesse modelo, X_i é denominado de variável independente e ε_i é o erro não observável associado a i -ésima observação. Para se estimar os parâmetros desse modelo (β_0 e β_1), que são o intercepto ou coeficiente linear e o coeficiente angular ou coeficiente de regressão, nenhuma pressuposição é feita acerca da distribuição dessa variável aleatória (ε_i). Todavia, se inferências forem realizadas, esses erros deverão assumir modelo específico de probabilidade e devem ser homocedásticos (FERREIRA, 2005). O objetivo da regressão é determinar estimadores de β_0 e β_1 de tal forma que as distâncias médias entre a reta de regressão e os valores observados sejam minimizadas. Definindo-se $\varepsilon_i = Y_i - \beta_0 + \beta_1 X_i$ a partir do modelo (2.4), então um critério interessante para determinar esses estimadores seria minimizar a soma de quadrados desses

resíduos ao longo de todos os n pares (X_i, Y_i) . Essa soma de quadrados é definida por

$$\mathbf{S}(\boldsymbol{\beta}) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \beta_0 + \beta_1 X_i)^2. \quad (2.5)$$

Esse método é chamado de mínimos quadrados e é facilmente obtido derivando-se essa última expressão em relação aos parâmetros β_0 e β_1 e igualando-se essas derivadas a zero.

Sendo assim, ao se usar as estimativas de mínimos quadrados supõe-se que (BATES e WATTS, 1988):

- i) A função esperança é correta;
- ii) A resposta é a função esperança mais o erro;
- iii) O erro é independente da função esperança;
- iv) Cada erro tem uma distribuição normal;
- v) Cada erro tem média zero;
- vi) Os erros têm variâncias iguais;
- vii) Os erros são independentemente distribuídos.

A abordagem geométrica na estatística fornece entre outras coisas, uma compreensão dos métodos dos mínimos quadrados lineares e da análise de variância usual, e subsequentemente, da regressão não linear. O método dos mínimos quadrados permite calcular um vetor \mathbf{X} que é o mais próximo possível da solução do sistema. O processo de cálculo de $\mathbf{S}(\boldsymbol{\beta})$ envolve duas etapas:

- i) Usar o vetor $\boldsymbol{\beta}$ do parâmetro p -dimensional e a matriz \mathbf{X} derivada $N \times P$ para obter o vetor da resposta esperada N -dimensional $\eta(\boldsymbol{\beta}) = \mathbf{X}\hat{\boldsymbol{\beta}}$;
- ii) Calcular a distância quadrada no $\eta(\boldsymbol{\beta})$ à resposta observada \mathbf{Y} , $\|\mathbf{y} - \eta(\boldsymbol{\beta})\|^2$.

A representação geométrica de mínimos quadrados lineares nos permite formular um esquema muito simples para determinar os parâmetros estimados $\hat{\boldsymbol{\beta}}$. Já que a superfície esperada é linear, tudo que deve-se fazer para determinar o ponto na superfície que é mais próximo ao ponto \mathbf{y} , isto é, projetar \mathbf{y} por sobre o plano esperado. Isto nos dá $\hat{\boldsymbol{\eta}}$, e o $\hat{\boldsymbol{\beta}}$ é então simplesmente o valor de $\boldsymbol{\beta}$ correspondente ao $\hat{\boldsymbol{\eta}}$.

2.3 Modelos de regressão não linear

A definição de um modelo não linear apresentada pela maioria dos autores, como Draper e Smith (1998), Bates e Watts (1988), Ratkowsky (1983) e outros, é que pelo menos uma derivada parcial da variável dependente, com relação a algum parâmetro presente no modelo, depende de algum parâmetro. Um modelo é não linear quando uma variável dependente \mathbf{Y} não pode ser escrita como funções lineares de seus parâmetros.

Considere o modelo

$$\mathbf{Y}_i = \beta_1 X_{2i}^{\beta_2} + U_i \quad (2.6)$$

Derivando-se o modelo em relação aos seus parâmetros tem-se

$$\frac{\partial y_i}{\partial \beta_1} = X_{2i}^{\beta_2} \quad \frac{\partial y_i}{\partial \beta_2} = \beta_1 X_{2i}^{\beta_2} \ln(X_{2i})$$

Como as derivadas do modelo dependem dos parâmetros pode-se dizer que é um modelo não linear. Mas fazendo-se uma transformação de variáveis, podem se tornar modelos de regressão lineares em novos parâmetros.

Segundo Gujarati (2006) alguns modelos podem parecer não lineares nos parâmetros, mas são intrinsecamente lineares, pois, as devidas transformações, podem se tornar modelos de regressão lineares em novos parâmetros. Os modelos não lineares são linearizados para facilitar a obtenção das estimativas dos parâmetros (BATES e WATTS, 1988).

Por exemplo, aplicando-se o logaritmo neperiano em ambos os lados da igualdade (2.6), o modelo passou a:

$$\mathbf{W}_i = \alpha + \beta_2 \ln X_{2i} + U_i \quad (2.7)$$

Com $\alpha = \ln \beta_1$ e $\mathbf{W}_i = \ln Y_i$, e as seguintes derivadas parciais em relação aos novos parâmetros

$$\frac{\partial W_i}{\partial \alpha} = 1 \quad \frac{\partial W_i}{\partial \beta_2} = \ln X_{2i}$$

Diz-se que um modelo de regressão é intrinsecamente não linear nos parâmetros, quando ele não é linear e nem intrinsecamente linear.

Considerando-se o modelo

$$Y_i = f(x_i + \beta) + \varepsilon_i \quad i = 1, 2, \dots, n$$

As derivadas parciais de \mathbf{Y}_t em relação aos parâmetros dependem dos mesmos:

$$\frac{\partial f}{\partial \beta_0} = 1 \quad ; \quad \frac{\partial f}{\partial \beta_1} = e^{\beta_2 X} \quad e \quad \frac{\partial f}{\partial \beta_2} = \beta_1 X e^{\beta_2 X}$$

O modelo não satisfaz a definição de um modelo de regressão linear e não há jeito de transformá-lo em um modelo linear nos parâmetros, neste caso o modelo é chamado de intrinsecamente não linear.

2.3.1 A geometria dos mínimos quadrados não lineares

A maioria dos algoritmos para computação das estimativas de mínimos quadrados $\hat{\beta}$ e a maioria dos métodos inferenciais para modelos não lineares são baseados em métodos iterativos que consideram uma aproximação linear local para o modelo (CHAMBERS, 1973; BATES e WATTS, 1980). Existem vários métodos iterativos para a obtenção das estimativas de mínimos quadrados dos parâmetros de um modelo de regressão não linear.

Nem sempre os modelos não lineares são expressos numa forma paramétrica adequada, que facilite a convergência rápida dos processos iterativos utilizados na estimação dos seus parâmetros, sendo necessário procurar, em muitos casos, uma parametrização mais apropriada.

Sejam y_1, \dots, y_n variáveis aleatórias independentes, e suponha que os dados a serem analisados sejam representados pelo vetor $y = (y_1, \dots, y_n)^T$. A estimativa do vetor de parâmetros β é calculada pelo método dos mínimos quadrados que, nesse caso, coincide com o método de máxima verossimilhança, pois o modelo tem respostas normais independentes com a mesma variância. A estimativa de β é, então, calculada, minimizando-se a função quadrática

$$S(\beta) = \sum_{i=1}^n \{y_i - \eta_i(\beta)\}^2.$$

Derivando-se $S(\beta)$ em relação a β , obtém-se

$$\frac{\partial S(\beta)}{\partial \beta} = 2 \sum_{i=1}^n \{y_i - \eta_i(\beta)\} \frac{\partial \eta}{\partial \beta}.$$

A estimativa $\hat{\beta}$ do vetor de parâmetros β é obtida igualando-se $\frac{\partial S(\beta)}{\partial \beta}$ a zero para $i = 1, \dots, p$. Em geral, as equações $\frac{\partial S(\beta)}{\partial \beta_r} = 0$ não são lineares e têm que ser resolvidas, numericamente, por processos iterativos.

A abordagem sugerida por Gauss Newton, foi utilizar uma aproximação linear à função

esperança, para reiterativamente melhorar o β como valor inicial para $\hat{\beta}$ e continuar melhorando as estimativas, até não existir mais mudança. Esse método iterativo usa uma expansão em série de Taylor para aproximar o modelo de regressão não linear com termos lineares e, então, aplica mínimos quadrados para estimar os parâmetros. Iterações desses passos geralmente conduzem a uma solução para o problema de regressão não linear. O processo é repetido um determinado número de vezes, até que o vetor de estimativas não se altere mais dentro de precisão pré-estipulada.

O método que se descreveu acima é o método de Gauss-Newton que é a base de muitos outros métodos. Alguns desses métodos introduzem modificações para obter uma convergência mais rápida como o caso do método de Marquardt (1963), bastante usado em investigação pesqueira. Outros métodos usam o desenvolvimento de Taylor de segunda ordem (método de Newton-Raphson), procurando-se assim uma melhor aproximação. Outros ainda, combinam as duas modificações.

Dentre os muitos existentes, um critério para verificar a convergência poderia ser

$$\sum_{r=1}^p \left(\frac{\beta_r^{(m+1)} - \beta_r^{(m)}}{\beta_r^{(m)}} \right)^2 < \xi,$$

tomando-se para ξ um valor suficientemente pequeno. A convergência em geral, depende dos valores iniciais para os parâmetros do vetor β . Isso pode evitar que problemas relacionados com a estrutura paramétrica do modelo, tais como, não linearidade acentuada, prejudiquem a convergência do processo iterativo.

2.4 Medidas de não linearidade

À medida que o tamanho da amostra cresce, o comportamento dos estimadores de mínimos quadrados se torna cada vez mais próximo do comportamento dos estimadores lineares, mas, até recentemente, não havia prontamente métodos aplicáveis para quantificar o comportamento não linear. As expressões que indicam a adequação de uma aproximação linear e seus efeitos nas inferências são chamadas medidas de não linearidade (BATES e WATTS, 1980).

As propriedades estatísticas de modelos não lineares, seu comportamento numérico no processo de estimação (convergência), bem como a validade de inferências assintóticas em amostras finitas são funções da não linearidade do modelo, ou, em outras palavras, do quão adequado é a aproximação linear (RATKOWSKY, 1983;).

Ratkowsky (1983) comparou algumas formas paramétricas para diversos modelos normais não lineares por meio de simulações e utilizou as medidas de Box e de Bates e Watts. Para se ter uma ideia mais clara dos conceitos de não linearidade intrínseca e de não linearidade paramétrica, serão comparados, a seguir, um modelo linear e um modelo não linear para o caso de $n = 2$ e $p = 1$. Considere, inicialmente, o modelo linear simples $Y_i = \beta x_i + \varepsilon_i, i = 1, 2$, em que x é uma covariável qualquer e β um parâmetro desconhecido. Nesse caso, o espaço de estimação tem dimensão igual a um, sendo formado pelos pontos

$$x\beta = (x_1, x_2)\beta, \beta \in \mathbb{R},$$

ou seja, é uma reta em R^2 . Além disso, para qualquer conjunto de soluções β^1, β^2, \dots , tais que $\beta^{(i+1)} - \beta^{(i)} = \Delta$, onde Δ é uma constante arbitrária, as soluções possíveis para $X\beta$ serão tais que

$$x\beta^{(i+1)} - x\beta^{(i)} = (x_1, x_2)\Delta, i = 1, 2, \dots,$$

ou seja, se as soluções para β forem igualmente espaçadas, então os valores ajustados correspondentes serão, também, igualmente espaçados.

Considere agora o modelo normal não linear $y_i = x_i^\beta + \varepsilon_i, i = 1, 2$ e os dados apresentados em Ratkowsky (1983)

$$\mathbf{y} = (2, 5, 10)^T \text{ e } \mathbf{x} = (2, 3)^T.$$

Nesse caso, o espaço de estimação não é mais uma reta, e sim uma curva ao redor da estimativa de máxima verossimilhança $\hat{\beta} = 2,05$. A curva correspondente aos pontos $(2^\beta, 3^\beta)^T$ com β variando em espaçamentos iguais a 0,5. Note que os pontos do espaço de estimação não são igualmente espaçados como ocorre no modelo linear.

Assim, quanto mais essa curva se afasta da reta tangente em $\hat{\beta}$ maior será o que Bates e Watts (1980) chamam de não linearidade intrínseca do modelo, e quanto mais desiguais forem os espaçamentos entre os pontos do espaço de estimação, maior será o que ambos autores chamam de não linearidade aparente causada pela parametrização do modelo.

Portanto, a não linearidade de um modelo pode ser devida a duas causas. A primeira é a curvatura real do modelo ou intrínseca, como definem Bates e Watts(1980), que é invariante com qualquer tipo de reparametrização. A segunda é a curvatura devida à forma como os parâmetros aparecem no modelo. Essa última pode ser eliminada ou pelo menos reduzida por meio da reparametrização. Para ilustrar esse fato, considere o modelo

normal não-linear descrito anteriormente com a seguinte reparametrização:

$$y_i = x_i^{\log \phi} + \varepsilon_i, i = 1, 2,$$

em que $\phi = \exp(\boldsymbol{\beta})$. A Figura 2 mostra os pontos da curva $(2^{\log \phi}, 3^{\log \phi})^T$ com espaçamentos iguais a 1,0 para ϕ . Nota-se que os espaçamentos entre os pontos correspondentes são praticamente iguais, indicando que o grau de não linearidade aparente foi, substancialmente, reduzido com essa reparametrização. Entretanto, a curvatura do espaço de estimação continua com a mesma forma anterior, como era de se esperar.

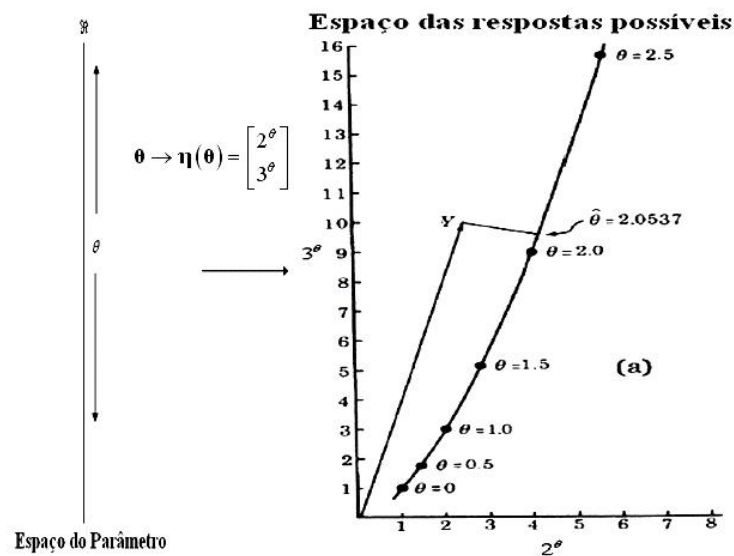


Figura 1: Representação da curva $(2^\beta, 3^\beta)^T$ com β variando em espaçamentos iguais a 0,5 (RATKOWSKY, 1983).

2.5 Medidas de curvatura de Bates e Watts

Um tratamento explicitamente mais geométrico para a não linearidade foi apresentado por Bates e Watts (1980), que propuseram a quantificação da não linearidade em duas componentes.

Considere o modelo de regressão normal não linear $Y = g(\beta; X) + \varepsilon = \mu(\beta) + \varepsilon$. Uma reta no espaço paramétrico passando por $\hat{\beta}$, pode ser expressa, usando um parâmetro escalar b , por

$$\beta(b) = \hat{\beta} + b\mathbf{h},$$

em que $\mathbf{h} = (h_{(1)}, \dots, h_{(p)})^T$ é um vetor de valores não-nulos. Essa reta gera uma curva,

sobre o espaço de estimação, definida por

$$\mu_h(b) = \mu(\hat{\beta} + b\mathbf{h}).$$

A aceleração da curva μ_h é estimada por

$$\ddot{\mu}_h = \mathbf{h}^T \widehat{\mathbf{W}} \mathbf{h}$$

em que \mathbf{W} é uma matriz de dimensão $n \times (p \times p)$ com i -ésima face expressa como $W_i = \frac{\partial^2 \mu_i}{\partial \beta_r \partial \beta_s}$, $i = 1, \dots, n$ e $r, s = 1, \dots, p$. Portanto, cada elemento indexado por i é da forma $\mathbf{h}^T \widehat{\mathbf{W}}_i \mathbf{h}$, para $i = 1, \dots, n$. Bates e Watts (1980), descreveram as seguintes curvaturas:

Curvatura intrínseca definida por:

$$K_h^{IN} = \|\ddot{\mu}^{IN}\| / \|\dot{\mu}\|^2$$

Curvatura devida à parametrização definida por

$$K_h^{PE} = \|\ddot{\mu}^{PE}\| / \|\dot{\mu}\|^2$$

A reparametrização não mudará o formato da superfície esperada. Entretanto, pode tornar as propriedades dos modelos não lineares mais similares aquelas dos modelos lineares com relação a questão das variações iguais em β não produzem variações iguais em $\eta(\beta)$. Para ilustrar esse fato, considere o modelo normal não linear descrito anteriormente com a seguinte reparametrização:

$$Y_i = x_i^{\log \phi} + \varepsilon_i, \quad i = 1, 2,$$

em que $\phi = \exp(\beta)$.

Essas curvaturas podem ser padronizadas de modo que fiquem invariantes a mudanças de escala. Após padronizadas as medidas relativas podem ser usadas não somente para comparar diferentes parametrizações de um determinado modelo, mas, também, diferentes conjuntos de dados para o mesmo modelo ou para modelos diferentes.

Para efetuar os cálculos não há, em geral, fórmulas explícitas, sendo necessário recorrer a algum processo iterativo. Souza (2008) descreve a obtenção de por meio de um processo iterativo proposto por Bates e Watts (1980).

Segundo Bates e Watts (1980), as reparametrizações não mudam a forma da superfície esperada, nem o efeito da curvatura intrínseca, é invariante com qualquer tipo de repara-

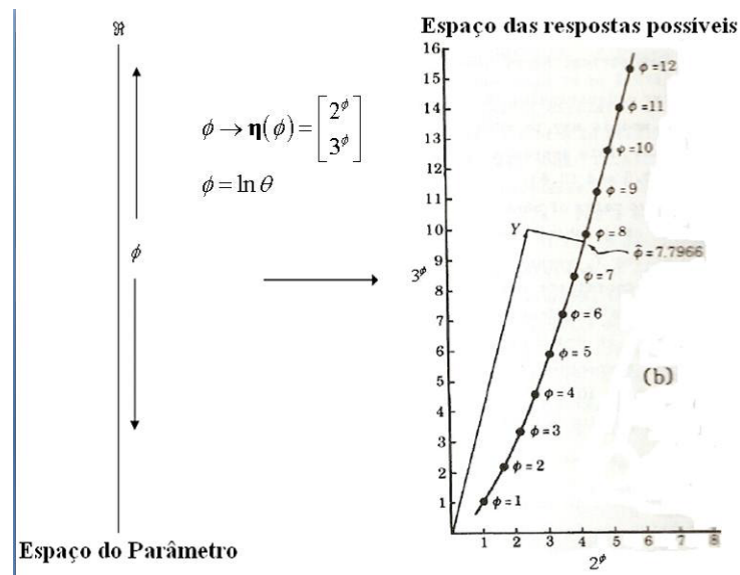


Figura 2: Representação da curva $(2^{\log \phi}, 3^{\log \phi})^T$ com ϕ variando em espaçamentos iguais a 1,0 (RATKOWSKY, 1983).

metrização, que é uma propriedade do local da solução. Assim, uma boa reparametrização pode mudar consideravelmente a curvatura de efeito parâmetro, sendo eliminada ou pelo menos reduzida, tornando mais válidos os resultados das inferências que foram obtidos pela aproximação linear e pela teoria assintótica.

2.6 Vício de Box

O processo inferencial em modelos não lineares é baseado em argumentos assintóticos e é fortemente influenciado pela não linearidade do modelo. Para uma avaliação da qualidade das inferências, vários pesquisadores propuseram medidas para quantificar a não linearidade e avaliar as suas consequências. O vício, geralmente, é ignorado com a justificativa de que é negligenciável quando comparado com o erro padrão das estimativas dos parâmetros (CORDEIRO e VASCONCELOS, 1997). Entretanto, à medida que a amostra se torna menor ou a não linearidade mais intensa, o vício nas estimativas passa a ser demasiado e, nessas situações, a disponibilidade de fórmulas para conhecê-lo é muito útil.

Box (1971) chegou à seguinte fórmula para o cálculo de viés em casos univariados homocedásticos

$$B(\hat{\beta}) = -\frac{\sigma^2}{2} \left(\sum_{i=1}^n F_i F_i^T \right)^{-1} \sum_{u=1}^n F_u \text{tr} \left[\left(\sum_{i=1}^n F_i F_i^T \right)^{-1} H_u \right],$$

em que $F_i (= F_u)$ é o vetor $(p \times 1)$ da primeira derivada do modelo e H_u é a matriz $(p \times p)$ de segunda derivada, ambos com respeito aos elementos de β , avaliados em $x_i, i = 1, \dots, n$. Na prática, são usados $\hat{\theta}$ e $\hat{\sigma}^2$ no lugar das quantidades desconhecidas e tr indica a operação traço. Para fins de interpretação, é esclarecedor explicar a aproximação do viés com regressão linear (Cook et al., 1986).

Na avaliação de modelos não lineares, é interessante apresentar as estimativas dos vícios em termos percentuais da estimativa, ou seja, $\%B(\hat{\beta}) = 100 \cdot B(\hat{\beta})/\beta$, valores acima de 1%, em valor absoluto, indicam comportamento não linear. A importância de se avaliar o vício reside em indicar quais parâmetros do modelo mais contribuem para o afastamento do comportamento linear.

Gills e Ratkowsky (1978), via simulação, concluíram que a medida de vício proposta por Box (1971) não só estima o vício de maneira muito aproximada como também fornece uma boa indicação do comportamento não linear do modelo.

3 Aplicação

Encontram-se nesta seção as principais metodologias que serviram de base para este trabalho, tanto na parte da descrição dos dados utilizando-se de modelos não lineares quanto nas inferências realizadas via aproximação e simulação.

3.1 Material e métodos

Os dados utilizados para a realização deste trabalho foram provenientes de um experimento conduzido no Laboratório de Biotecnologia Vegetal da Embrapa Mandioca e Fruticultura, em Cruz das Almas, Bahia. Segmentos nodais de um (1) centímetro de comprimento de plantas de maracujazeiro *Passiflora giberti* N. E. Brown., foram cultivados *in vitro* em meio de cultura MS suplementado com 30 g L^{-1} de sacarose (MS padrão) por 60 dias, o objetivo alvo na elaboração e obtenção desses dados foi o de estimar qual o tamanho de parcela considerado ideal para esse tipo de experimento e para tanto utilizou-se 20 unidades básicas.

Tabela 1: Dosagens de sacarose e sorbitol de acordo com os tratamentos, utilizadas no experimento de conservação *in vitro* de *Passiflora giberti* N.E.Brown.

Tratamentos	Sacarose.g L^{-1}	Sorbitol.g L^{-1}
T1	30	0
T2	0	10
T3	0	20
T4	0	40
T5	15	10
T6	15	20
T7	15	40
T8	30	10
T9	30	20
T10	30	40

Para estimação do tamanho de parcela cada tratamento foi considerado como um

ensaio em branco. Desse modo, as 20 unidades básicas, foram dispostas convenientemente em linhas e colunas nas grades, simularam-se 15 diferentes tamanhos de parcelas, formados por X_1 unidades básicas na linha e X_2 unidades básicas na coluna. Os tamanhos de parcela foram simulados pelo agrupamento de unidades de modo que X_2 correspondesse a X (tamanho da parcela em unidades básicas), conforme pode ser observado na Tabela 2. Para as parcelas simuladas de diferentes formas, mas com o mesmo tamanho, foi

Tabela 2: Tamanho da parcela (X), forma da parcela e número de parcelas totais para os ensaios de uniformidade de plantas de *P. giberti* N. E. Brown aos 60 dias de conservação *in vitro*.

Simulações	Tamanho X	Forma $X_1 \times X_2$	Número de Parcelas
1	1	1x1	20
2	2	2x1	10
3	2	1x2	8
4	3	3x1	5
5	3	1x3	4
6	3	2+1	4
7	3	1+2	4
8	4	2x2	4
9	5	4+1	2
10	6	2x3	2
11	6	3x2	2
12	7	4+3	2
13	8	2x4	2
14	8	4x2	2
15	10	2x5	2

considerada a média aritmética dos coeficientes de variação, provenientes de formas de diferentes parcelas.

O modelo exponencial foi utilizado para estimar o tamanho ótimo de parcela considerando-se a relação entre o coeficiente de variação (CV) e o tamanho da parcela com X unidades básicas é explicado por

$$CV = \alpha X^{-\beta} \quad (3.1)$$

em que α e β são os parâmetros a serem estimados. A partir da função de curvatura dada por esse modelo, determinou-se o valor da abscissa no qual ocorre o ponto de máxima curvatura, dada por: $X_0 = \exp\{[1/(2\beta + 2)] \log[(\alpha\beta)^2(2\beta + 1)/(\beta + 2)]\}$, em que X_0 é o valor da abscissa no ponto de máxima curvatura, o qual corresponde à estimativa do tamanho ótimo da parcela experimental.

Foi empregado o procedimento iterativo Gauss-Newton para a obtenção das estimativas dos parâmetros utilizando-se da função *nls* desenvolvida para ajustes de modelos não lineares do programa R versão 2.13.0.

As curvaturas médias de Bates e Watts são obtidas por

$$\gamma_{rms}^{ep} = \frac{1}{p(p+2)} \sum_{i=1}^p \left[2 \sum_{j=1}^p \sum_{m=1}^p c_{ijm}^2 + \left(\sum_{j=1}^p c_{ijj} \right)^2 \right], \quad (3.2)$$

$$\gamma_{rms}^{in} = \frac{1}{p(p+2)} \sum_{i=p+1}^k \left[2 \sum_{j=1}^p \sum_{m=1}^p c_{ijm}^2 + \left(\sum_{j=1}^p c_{ijj} \right)^2 \right] \quad (3.3)$$

Os valores de curvatura foram comparadas com o raio do disco de confiança $100(1-\alpha)$, portanto, uma curvatura será considerada pequena se $\gamma_{rms}\sqrt{F} < 1/2$ em que $F = F_{\alpha,p,n-p}$.

O vício relativo assintótico (VR_a) foi definido em relação ao erro padrão da estimativa (ep), calculado por

$$VR_a = 100 \times \frac{B(\hat{\theta}_i)}{ep(\hat{\theta}_i)}, \quad (3.4)$$

em que $ep(\hat{\theta}_i)$ é o erro padrão de θ_i e $B(\hat{\theta}_i)$.

Estudos de simulação foram conduzidos para investigar as propriedades dos estimadores de mínimos quadrados dos parâmetros sob o modelo em estudo. Foram gerados $N = 1000$ experimentos aleatórios, assumindo-se como verdadeiro o modelo exponencial, sendo a distribuição dos erros considerada independente, normal e de variância constante. Os valores simulados dos parâmetros e da variância foram obtidos pelo ajuste de mínimos quadrados dos modelos aos dados. De acordo com as 1000 estimativas foram construídos histogramas de distribuição de frequência e os intervalos de confiança *bootstrap*. Também foram encontrados os primeiros quatro momentos amostrais centrados na média dessas 1000 estimativas, com o objetivo da realização do teste de assimetria e curtose assintótico.

3.2 Resultados e discussão

O modelo não linear que relaciona o coeficiente de variação com o tamanho da parcela ajustou-se consideravelmente bem aos dados de acordo com o coeficiente de determinação R^2 , pois o modelo teve ajuste superior a 70% para todos os tratamentos, exceto pra o

tratamento 9 que foi de 57,64% e apresentou todas as estimativas dos parâmetros significativas com valor de p inferior a 0,05. Os intervalos de confiança para os parâmetros, bem como o erro padrão das estimativas foram calculados considerando-se um nível de confiança de 95% (Tabela 3). Ressalta-se, porém, que o R^2 não é um critério adequado para discussão de ajuste de modelos, pois, geralmente, em ajuste de modelos não lineares, é comum a obtenção de valores assintóticos altos e similares.

Tabela 3: Estimativas (E) dos parâmetros (P) a e b baseadas na inferência por aproximação linear, erro padrão da estimativa (E.P.E), valor de p para o teste t , intervalo de confiança (IC) de 95% para os parâmetros, vício de Box e coeficiente de determinação (R^2) para o modelo exponencial, nos 10 tratamentos (T) analisados

T	P	E	E.P.E.	Valor-p	IC (95%)	Vício de Box	R^2
T1	a	71,2507	7,6445	< 0,0001	[53,1742; 89,3272]	0,1332	0,8170
	b	0,5402	0,0957	< 0,0001	[0,3137; 0,7666]	0,2015	
T2	a	21,1537	2,1158	< 0,0001	[16,1505; 26,1569]	0,1245	0,7570
	b	0,3740	0,0781	< 0,0001	[0,1893; 0,5588]	-0,0246	
T3	a	57,1405	6,0175	< 0,0001	[42,9113; 71,3696]	0,1274	0,8355
	b	0,5614	0,0957	< 0,0001	[0,3349; 0,7879]	0,2211	
T4	a	13,4459	1,1716	< 0,0001	[10,6754; 16,2164]	0,0842	0,8946
	b	0,6865	0,0890	< 0,0001	[0,4760; 0,8969]	0,2633	
T5	a	80,0169	6,3275	< 0,0001	[65,0548; 94,9791]	0,0659	0,9321
	b	1,0180	0,1144	< 0,0001	[0,7476; 1,2884]	0,5210	
T6	a	51,0479	3,6669	< 0,0001	[42,3771; 59,7189]	0,0695	0,7371
	b	0,2406	0,0512	< 0,0001	[0,1195; 0,3619]	-0,1287	
T7	a	28,523	4,3660	< 0,0001	[18,1994; 38,8471]	0,2617	0,7100
	b	0,6570	0,1520	< 0,0001	[0,2977; 1,0164]	0,7289	
T8	a	93,4211	9,1974	< 0,0001	[71,6726; 115,1698]	0,1133	0,8113
	b	0,5115	0,08575	< 0,0001	[0,3088; 0,7143]	0,1378	
T9	a	45,9112	5,7239	< 0,0001	[32,3763; 59,4461]	0,2026	0,5764
	b	0,2922	0,9195	< 0,0001	[0,0748; 0,5096]	-0,2297	
T10	a	20,7143	1,5191	< 0,0001	[17,1221; 24,3066]	0,0592	0,9284
	b	0,7226	0,07761	< 0,0001	[0,5391; 0,9062]	0,2108	

Na figura 3, tem-se o ajustamento do modelo exponencial para experimentos de conservação *in vitro* de *Passiflora giberti* N.E. Brown, vemos que há presença de alguns pontos fora da curva indicando que os dados não se ajustaram tão bem quanto esperado, porém de acordo com as medidas de não linearidade o modelo tem propriedades próximas das dos modelos lineares.

Como podemos ver na figura 4, temos bandas de confiança para resposta esperada delimitando-se uma região relativamente pequena entre elas, podemos perceber um ajuste razoavelmente bom do modelo exponencial para um modelo linear, existem alguns pontos

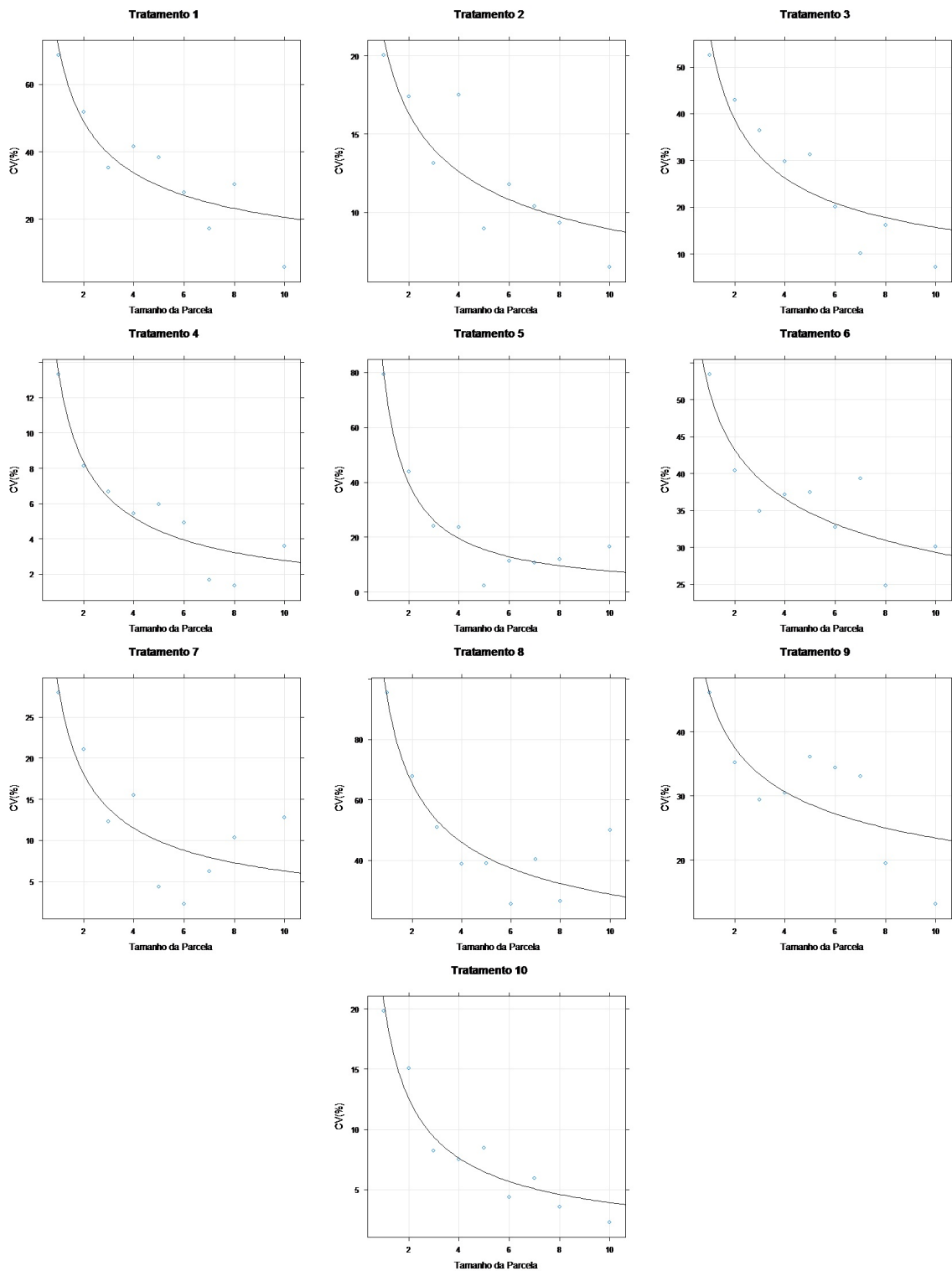


Figura 3: Ajuste do modelo exponencial em experimentos de conservação *in vitro* de *Passiflora giberti* N.E. Brown

que podemos ser considerar atípicos que estão fora da reta , mais apenas um estudo futuro sobre estes poderá fazer análises para saber se contribuíram ou não para o ajuste do modelo.

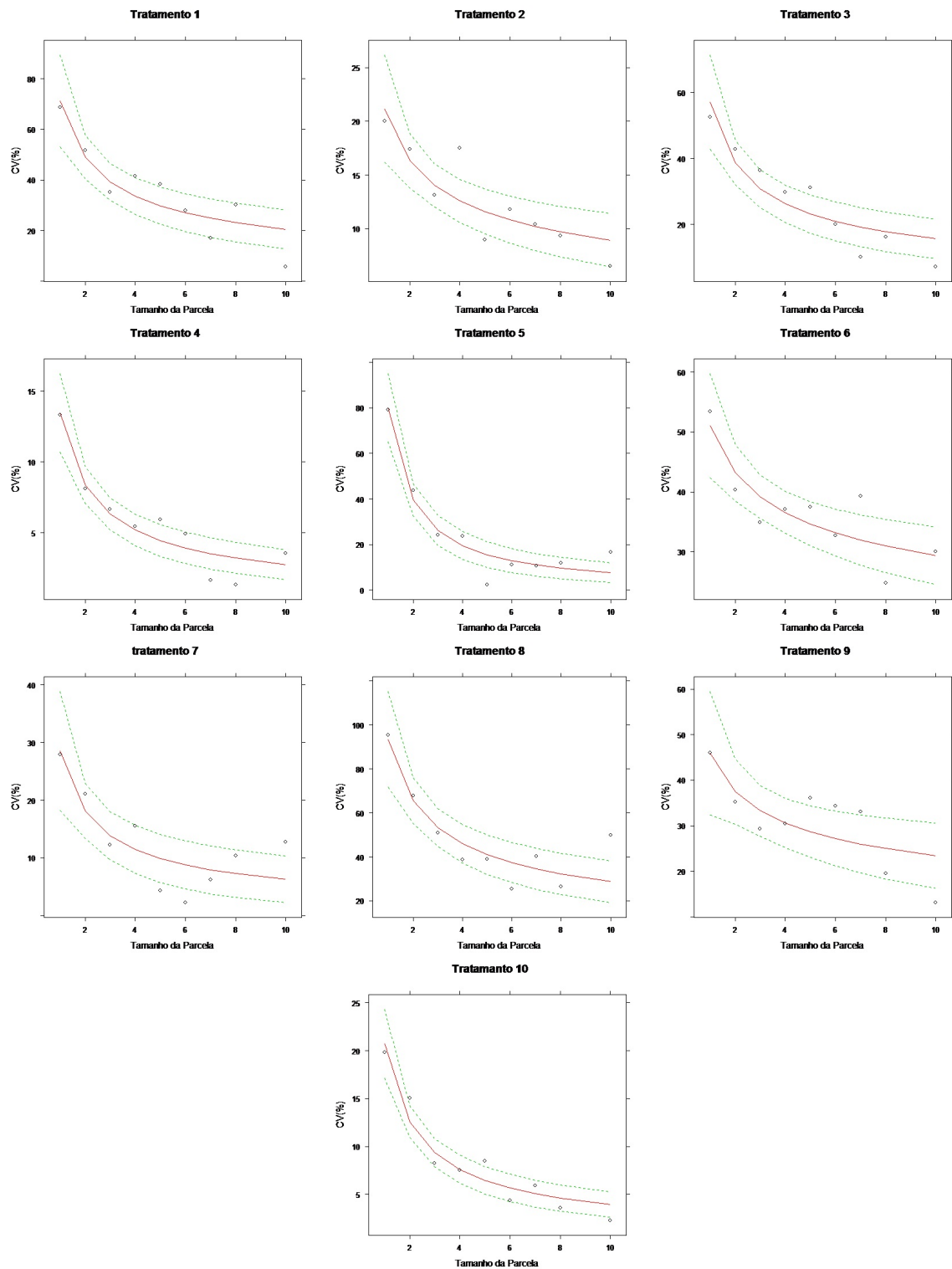


Figura 4: Bandas de confiança para a resposta esperada e para estimativas do modelo exponencial em experimentos de conservação *in vitro* de *Passiflora giberti* N.E.Brown

Para verificar se os resultados obtidos na Tabela 3 podem ser considerados válidos pela aproximação linear realizada no processo de estimação, foram calculadas as medidas de não linearidade propostas por Bates e Watts (1980) a curvatura de efeito parâmetro e

a curvatura intrínseca.

Tabela 4: Medidas de curvatura quadrática média (RMS) para o modelo $CV = aX^{-b}$.

Tratamentos	Curvatura intrínseca	Curvatura de efeito parâmetro
T1	0,1098	0,2385
T2	0,0930	0,2332
T3	0,1093	0,2348
T4	0,0992	0,2056
T5	0,1210	0,2564
T6	0,0632	0,1883
T7	0,1703	0,3545
T8	0,0989	0,2186
T9	0,1118	0,3103
T10	0,0860	0,1776

A curvatura intrínseca foi menor do que 0,5 para todos os tratamentos, indicando-se que o local da solução na vizinhança das estimativas é relativamente plano, já que a curvatura intrínseca define o formato da distância do local da solução, tornando-se, assim, verdadeira a aproximação linear por um plano. Observa-se também que a curvatura de efeito parâmetro apresentou resultado entre 0,1 e 0,3, para os tratamentos, indicando-se que a forma como parâmetros aparecem no modelo não teve muita influência sobre as inferências pela aproximação linear, mesmo que o tamanho da amostra seja pequeno (Tabela 4).

Não houve necessidade de reajuste do modelo com uma reparametrização, pois, todos os tratamentos apresentaram curvaturas intrínsecas baixas indicando-se que a forma com que os parâmetros aparecem no modelo são adequadas para este conjunto de dados, tendo em vista que as inferências obtidas via aproximação linear e teoria assintótica podem ser válidas, pois as propriedades dos estimadores referente a este modelo estão próximas em relação às propriedades dos estimadores lineares.

Esses resultados reforçam que os estimadores de mínimos quadrados em modelos não lineares têm comportamento amostral de lineares, à medida que a superfície esperada se torna plana e uniforme. Pelo estudo da não linearidade do modelo para o conjunto de dados foi possível avaliar o quão distante estão as propriedades dos estimadores não lineares de suas propriedades assintóticas, ou seja, o quão próximo está um modelo não linear de um modelo linear.

Os ajustes e as inferências realizados até aqui foram todos realizados por meio da

aproximação linear. Para uma validação ou não desses resultados foi realizado também uma simulação *bootstrap*, comparando-se os resultados da Tabela 5 com os da Tabela 3, pode-se constatar que os intervalos de confiança *bootstrap* são muito semelhantes aos intervalos de confiança obtidos pela aproximação linear, validando-se ainda mais os resultados obtidos anteriormente.

Tabela 5: Estimativas (E) dos parâmetros (P) a e b baseadas na inferência por aproximação linear, erro padrão da estimativa (E.P. E), valor de p para o teste t , intervalo de confiança (IC) de 95% para os parâmetros, vício de Box e coeficiente de determinação (R^2) para o modelo exponencial, nos 10 tratamentos (T) analisados

T	P	E.A.L	Média	W	Viés %	IC (95%)	T. A.	T. C.
T1	a	72,4265	71,5994	0,9688 ^{ns}	0,3487	[7,9729; 81,4070]	-0,4523*	2,4161*
	b	1,1366	0,5419	0,9972*	0,0017	[0,3838;0,6915]	-0,0787 ^{ns}	2,6721 ^{ns}
T2	a	20,9037	21,1775	0,9558 ^{ns}	0,0237	[18,1868;25,6577]	0,7522*	3.2525*
	b	0,3700	0,3742	0,9958 ^{ns}	0,0001	[0,2442;0,5203]	0,7522*	3.1376*
T3	a	57,5795	57,2294	0,9780 ^{ns}	0,0888	[47,0748;66,0066]	-0,1690*	2,0976*
	b	0,5602	0,5608	0,9975*	-0,0005	[0,3955;0,7133]	0,0455 ^{ns}	2,7910 ^{ns}
T4	a	13,6280	13,4239	0,9579 ^{ns}	-0,0220	[11,2873;15,1358]	-0,4579*	2,3227*
	b	0,6819	0,6841	0,9956 ^{ns}	-0,0023	[0,3955;0,7133]	0,0455*	2,8693*
T5	a	80,2585	80,0671	0,9232 ^{ns}	0,0502	[66,5495;89,0377]	-0,9161*	3.8565*
	b	1,0243	1,0300	0,9946 ^{ns}	0,0120	[0,5309;0,8504]	0,2675*	3.1550*
T6	a	51,0779	51,0640	0,9945 ^{ns}	0,0314	[44,7311;57,8132]	0,1149*	2,5814*
	b	0,2392	0,2390	0,9977*	-0,6985	[0,1503;0,3284]	-0,0498 ^{ns}	2,7212 ^{ns}
T7	a	28,8048	28,5883	0,9771 ^{ns}	0,2279	[21,2038;35,2115]	-0,1889*	2,1057*
	b	0,6686	0,6677	0,9960 ^{ns}	1,6165	[0,1503;0,3284]	-0,1111*	2,6404*
T8	a	92,0936	93,3060	0,9443 ^{ns}	-0,1232	[80,7086;112,8233]	0,8645*	3.5380*
	b	0,5087	0,5123	0,9969*	0,1532	[0,3711;0,6623]	0,1844 ^{ns}	3.0434 ^{ns}
T9	a	45,6784	46,0642	0,9856 ^{ns}	0,3333	[36,5649;54,8805]	-0,0160*	2,2364*
	b	0,2919	0,2896	0,9970*	-0,8926	[0,1326;0,4413]	-0,0649 ^{ns}	2,6106 ^{ns}
T10	a	20,3991	20,7457	0,9335 ^{ns}	0,1518	[18,9486;23,4061]	0,5943*	2,2405*
	b	0,7206	0,7233	0,9966 ^{ns}	0,0950	[0,6026;0,8575]	0,1456*	2,7359*

* significativo ao nível de 5% de probabilidade

O teste de Shapiro-Wilks W também foi utilizado para verificar se os parâmetros do modelos possuem distribuição aproximadamente normal. O valor calculado de W é estatisticamente significativo ($p = 0,05$), rejeita-se a hipótese que a distribuição estudada é normal, ou seja, para a distribuição ser considerada normal o valor de p deve ser maior que 0,05, como foi observado em alguns tratamentos (Tabela 5).

Nesses casos em que não se observou normalidade, observou-se simetria e curtose para a maior parte dos das estimativas dos parâmetros para os tratamentos do modelo exponencial, pois o valor de p foi menor que o nível de significância α , ou seja a hipótese de nulidade foi rejeitada.

Comparando-se a Tabela 3 com a Tabela 5, observou-se que o vício assintótico foi maior que 1% (vício significativo) para todas as estimativas. O vício *bootstrap*, de maneira geral, é superior ao vício de Box, haja vista que o vício assintótico é menos aproximado do real, pois é obtido por operações que envolvem a expansão do modelo até o termo de segunda ordem e que expansões de maior ordem tenderão a fornecer estimativas de vício mais próximas daquelas obtidas por simulação. O autor ainda considera que, mesmo o vício assintótico não sendo tão exato.

O modelo ajustado aos dados do tratamento 6 foi o que teve melhor ajuste, devido ter apresentado as menores curvaturas. Os testes para desvios de assimetria e de curtose, aplicados às estimativas *bootstrap*, apontaram a falta de normalidade, a maior parte dos tratamentos tiveram valores de curtose significativos o que evidencia a falta de normalidade dos dados, evidenciando-se o comportamento mais próximo do linear.

4 Conclusão

A estimativa dos parâmetros obtidos pelo modelo exponencial foram significativas e é possível afirmar que o modelo apresenta o comportamento próximo do linear de acordo com as medidas de não linearidade encontradas para todos os experimentos de conservação *in vitro* de *Passiflora giberti* N.E. *Brown* analisados. Pode-se concluir em termos estatísticos que o modelo teve um ajuste razoavelmente bom, já que as medidas de curvaturas, intrínseca e paramétrica foram baixas, tornando-se assim, verdadeira a aproximação linear por um plano, e que a forma como parâmetros aparecem no modelo não teve muita influência sobre as inferências pela aproximação linear, mesmo que o tamanho da amostra seja pequeno, sendo assim não houve necessidade de reajuste do modelo por reparametrizações.

Houve alguns pontos que não se ajustaram, porém são efeitos que não pode-se controlar, pois não temos acesso ao material experimental para saber qual o motivo dos pontos discrepantes, contudo o modelo pode ser utilizado em estudos futuros para estimação de tamanhos de parcelas experimentais devido a suas propriedades serem próximas de modelos lineares.

5 Referências Bibliográficas

- BATES, D. M.; WATTS, D. G. Relative curvature measures of nonlinearity (With discussion). **Journal of the Royal Statistical Society**, Ser. B, v.42, n. 1, p. 1-25, 1980.
- BATES, D. M.; WATTS, D. G. **Nonlinear regression analysis and its applications**. New York: Wiley series in probability e mathematical statistics, 1988. 365p.
- BOX, M. J.; Bias in nonlinear estimation. *Journal of royal statistical society. Serie B. Methodological*, London, v. 33, n. 2, p. 171-201, Apr. 1971.
- COOK, R. D.; TSAI, C. L.; WEI, B. C. Bias in nonlinear regression. *Biometrika*, London, v. 73, n. 3, p. 615-623, Dec. 1986.
- CORDEIRO, G. M.; VASCONCELLOS, K. L. P. **Bias correction for a class of multivariate nonlinear regression models**. *Statistics and probability letters*,
- DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3. ed. New York: J. Wiley, 1998. 706p.
- GUJARATI, D. **Econometria básica**. 4. ed. São Paulo: Campus, 2006. 806p.
- GILLS, P. R.; RATKOWSKY, D. A. The behavior of estimators of the parameters of various yield-density relationships. *Biometrics*, Washington, CHARNET, R. et al. *Análise de modelos de regressão linear com aplicações*. São Paulo: Unicamp, 1999. 354p.
- RATKOWSKY, D. A. **Nonlinear regression modelling**. a unified practical approach, New York: Marcel Dekker, 1983. 276p.
- HOFFMAN, Rodolfo; VIEIRA, Sônia. *Análise de Regressão: Uma Introdução à Econometria*. São Paulo: Hucitec, 1998.
- SOUZA, E. M. **Modelagem não linear da extração de zinco em um solo tratado com lodo de esgoto**. 2008, 95p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária)-Universidade Federal de Lavras, Lavras, MG.

APÊNDICE A – Primeiro anexo

PROGRAMA 1C: Rotina R

```
# Ajuste do modelo exponencial Tratamento 1

x=c(1,2,3,4,5,6,7,8,10)
y=c(68.67,51.82,35.17,41.61,38.31,27.97,17.09,30.30,5.75)

#Pacote Necessários

library(MASS)

library(lattice)

library(nlstools)

library(moments)

library(car)

library(NRAIA)

#Rotina para obter os valores iniciais

chute=lm(log(y)~1+log(x)) vetor=as.vector(chute$coefficients),
chutea=exp(vetor[1]);chutea chuteb=-vetor[2],

liber <- data.frame(x, y) xyplot(y~x, type=c("p", "smooth"),
col=1,xlab="Tamanho da parcela (Xc)",ylab="CV (

lmedio <- data.frame(x=unique(liber$x), kmedio=tapply(liber$y,liber$x, mean)),
```

```

quoc.der <- deriv3(~theta1*x^(-theta2) c(theta1", "theta2"), function(theta1,
theta2, x) NULL),"

#ajuste do modelo

m0q <- nls(y~quoc.der(theta1, theta2,x), data=liber,
start=c(theta1= 90.01137,theta2= 0.7792647))

summary(m0q),

plotfit(m0q,xlab=Tamanho da Parcela", ylab="CV(

m0q1 <- lm(y~factor(x), data=liber)

m0qm <- nls(kmedio~(theta1*x^(-theta2)), data=lmedio,
start=c(theta1= 23,theta2=0.8))

SQR=sum(res^2) SQTcorrigida=sum((y-mean(y))^2) R2=1-SQR/SQTcorrigida

#Encontrando o R2

R2 <- 1-deviance(m0qm)/deviance(lm(kmedio~1, data=lmedio));R2

shapiro.test(residuals(m0q))

#with(lib, bartlett.test(klib, tempo)),
xyplot(residuals(m0q)~fitted(m0q), col=1,type=c("p","smooth"),xlab="Valores
ajustados",ylab="Resíduos", main="Tratamento 1")"

qqmath(residuals(m0q), col=1, prepanel=prepanel.qqmathline,
panel=function(x, ...){, panel.qqmathline(x, lty=2, ...),
panel.qqmath(x, ...)})

# intervalo de confiança assintótico

sm <- summary(m0q)$coef
cbind(sm[,1]-sm[,2]*qt(0.975, df=df.residual(m0q)),
sm[,1]+sm[,2]*qt(0.975, df=df.residual(m0q)))

# curvaturas médias de Bates e Watts

rms.curv(m0q)

#Função para cálculo do vício de Box (1971).

```

```

biasbox <- function(nls.obj){, theta <- summary(nls.obj)$coef[,1],
sd.theta <- summary(nls.obj)$coef[,2],
F <- attr(nls.obj$m$fitted(), gradient"),
H <- attr(nls.obj$m$fitted(), hessian"), sig <- summary(nls.obj)$sigma,
n <- dim(F)[1], FlFi <- t(F)%*%F,
d <- -(sig^2/2)*sapply(1:n, function(x){,
sum(diag(solve(FlFi)%*%H[x, , ]))}),
bias <- as.vector(solve(FlFi)%*%t(F)%*%d),
names(bias) <- names(coef(nls.obj)), bias.sd <- 100*bias/sd.theta,
bias.th <- 100*bias/theta, return(list(viés bruto"=bias,"

# vício de Box

biasbox(m0q)

# bandas de confiança para a média estimada

tc <- qt(0.975, df=df.residual(m0q)), liber$fit <- fitted(m0q),
fit=c(71.25072,48.99637,39.35810,
33.69290,29.86653,27.06504,24.90245,23.16931,20.53806),
liber$lwr <- fit-tc*diag(sqrt(m0qg%*%vc%*%t(m0qg))),
liber$upr <- fit+tc*diag(sqrt(m0qg%*%vc%*%t(m0qg)))

xyplot(fit+lwr+upr+y~x, data=liber,distribute.type=TRUE, lty=c(1,2,2),
type=c("l","l","l","p"), col=c(2,3,3,1), xlab="Tamanho da Parcela", ylab="CV(

# simulação bootstrap

m0q.sim <- nlsBoot(m0q, niter=1000)

summary(m0q.sim)$coef[,1], dim(m0q.sim$coefboot),
apply(m0q.sim$coefboot, 2, mean), var(m0q.sim$coefboot),
sd(m0q.sim$coefboot), cor(m0q.sim$coefboot)

# vício bootstrap bruto

apply(m0q.sim$coefboot, 2, mean)-c(sm[,1])
vício bootstrap relativo à estimativa

100*(apply(m0q.sim$coefboot, 2, mean)-c(sm[,1]))/c(sm[,1])

```

```

# vício bootstrap relativo ao erro padrão da estimativa
100*(apply(m0q.sim$coefboot, 2, mean)-c(sm[,1]))/c(sm[,2])

# intervalo de confiança bootstrap
apply(m0q.sim$coefboot, 2, sort)
[as.integer(c(0.025,0.975)*dim(m0q.sim$coefboot)[1]),]

# teste de normalidade
apply(m0q.sim$coefboot, 2,
function(x){data.frame(W=shapiro.test(x)$statistic,
pval=shapiro.test(x)$p.value)})

# teste de assimetria
apply(m0q.sim$coefboot, 2,
function(x){data.frame(A=agostino.test(x)$statistic[1],
pval=agostino.test(x)$p.value)})

# teste de Curtose
apply(m0q.sim$coefboot, 2,
function(x){data.frame(C=anscombe.test(x)$statistic[1],
pval=anscombe.test(x)$p.value)})

scatterplot.matrix(m0q.sim$coefboot, diagonal="histogram")

```