

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA

Thalita do Bem Mattos

Modelos Não Lineares e suas Aplicações

Juiz de Fora
2013

Thalita do Bem Mattos

Modelos Não Lineares e suas Aplicações

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito para a obtenção do grau de Bacharel em Estatística.

Orientador: Clécio da Silva Ferreira
Doutor em Estatística - Universidade de São Paulo

Juiz de Fora
2013

Ficha catalográfica elaborada através do Programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Mattos, Thalita do Bem .
Modelos Não Lineares e suas Aplicações / Thalita do Bem
Mattos. -- 2013.
58 p. : il.

Orientador: Clécio da Silva Ferreira
Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas, 2013.

1. Modelos Não Lineares. 2. Métodos Iterativos. 3. Análise de Diagnóstico. I. Ferreira, Clécio da Silva, orient. II. Título.

Thalita do Bem Mattos

Modelos Não Lineares e suas Aplicações

Monografia apresentada ao Curso de Estatística da Universidade Federal de Juiz de Fora, como requisito para a obtenção do grau de Bacharel em Estatística.

Aprovado em 28 de agosto de 2013

BANCA EXAMINADORA

Clécio da Silva Ferreira

Doutor em Estatística - Universidade de São Paulo

Ângela Mello Coelho

Doutora em Estatística e Experimentação Agronômica -
ESALQ/USP

Ronaldo Rocha Bastos

Ph.D. in Urban and Regional Planning - Liverpool University

*Aos meus pais, Marilene e Milton,
à minha irmã, Areta,
e ao meu sobrinho, Luccas.
Dedico com muito amor e carinho.*

Agradecimentos

A Deus por nunca me desamparar, nos momentos mais difíceis da minha vida me deu força, coragem e companhia para seguir na minha caminhada.

Aos meus pais, Marilene e Milton, que sempre me incentivaram para os estudos, e nunca mediram esforços para me ajudar ao longo destes anos. Agradeço por tanto amor, carinho, dedicação, paciência, e por sempre acreditar em mim, abdicando dos seus próprios sonhos em favor dos meus.

À minha irmã, Areta, pelo braço esquerdo e direito sempre que preciso, pelos conselhos de vida, pelas broncas. Um orgulho ter uma irmã tão batalhadora e firme em seus objetivos. Meu exemplo!

Ao meu sobrinho, Lucas, pelas brigas, surros e murros, pelos rabiscos nas minhas matérias, por querer jogar no computador sempre quando estou utilizando. Obrigada por tornar meus dias mais divertidos e estressantes também.

Ao meu cunhado, Carlos, um irmão mais velho que Deus me deu. Obrigada pelo apoio, pelas risadas, pelas palavras de incentivo e por me aturar.

Ao Renan, por entender minhas ausências, por compartilhar minhas lágrimas e sorrisos. Agradeço pela paciência, companheirismo, carinho e amor. Obrigada por sempre me apoiar e incentivar nas horas difíceis.

Às minhas inesquecíveis amigas de faculdade. À Jéssica, minha companheira de ensino médio, faculdade e profissão, por ser uma amiga incrível que sempre esteve ao meu lado em todos os momentos. À Virginia, por ser uma mãe, dedicada e vaidosa. À Bell, por ser a amiga mais chata do mundo, mas sempre preocupada com todos. À Laila, pelo carinho e atenção, longe mais ao mesmo tempo tão perto. À Karla, pelas músicas sertanejas compartilhadas, pelas caronas e pela amizade. À Gisele, pelas matérias ensinadas, pelas palavras de força nos momentos difíceis e por ser uma botafoguense. À Monique por ser uma sementinha do mal.

Aos meus amigos da minha pequena grande turma do BI. Ao Daniel, por dizer “Calma, que vai dar certo”. Ao Guilherme, por dizer “A matéria tá muito difícil, estamos ferra-

dos!”. Ao Leandro, por sempre dizer “E aí? Já estudou?”. Muito obrigada meninos, pelo companheirismo e amizade durante esta jornada, pelas palavras de apoio e incentivo.

Aos meu companheiros de turma agregada. À Jéssica de Bicas, por estar sempre com um sorriso no rosto, sempre alegre e amiga. À Vanessa e Franciele, pela companhia para João Pessoa, viagem inesquecível. Ao Manoel, pelos ensinamentos e exemplo de força de vontade. À Camila, Anna Cláudia e Amanda pelas conversas nos corredores.

Ao professor Clécio, pelos ensinamentos passados durante os 2 anos de Iniciação Científica, pela paciência e dedicação na orientação deste trabalho.

Aos professores Ângela e Ronaldo por participarem da banca e todos os professores do Departamento de Estatística pelas aulas ministradas e seus conhecimentos repassados.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

“To infinity ... and beyond!”

(Toy Story)

*“O correr da vida embrulha tudo,
a vida é assim: esquenta e esfria,
aperta e daí afrouxa, sossega depois desorienta,
o que ela quer da gente é coragem.”*

(Guimarães Rosa)

*“Uma resposta aproximada para o problema certo
vale muito mais do que uma resposta exata
para um problema aproximado.”*

(John Wilder Tukey)

Resumo

Este trabalho tem como objetivo principal revisar os modelos normais não lineares, contemplando suas classes, o método de mínimos quadrados para estimação dos seus parâmetros, ressaltando os métodos iterativos de Gauss-Newton e Newton-Raphson. Ainda, apresenta alguns resultados assintóticos, medidas de não linearidade e técnicas de diagnóstico. Uma característica dos modelos não lineares é que técnicas de diagnósticos utilizadas em modelos lineares são inadequadas caso o grau de não linearidade seja alto. Neste trabalho, estudamos as medidas de não linearidade existentes para a verificação desses pressupostos e as utilizamos em nossas aplicações. Por fim, analisamos o conjunto de dados *Coelhos Europeus* para demonstrar a aplicabilidade do modelo não linear proposto na literatura. Ainda, aos mesmos dados ajustamos outros modelos não lineares, além de métodos de estimação de curva não-paramétricas, para efeitos de comparação. Concluimos que em modelos não lineares existe uma forte ligação entre o funcional não linear e o fenômeno sob estudo.

Palavras-chaves: Modelos Não Lineares; Métodos Iterativos; Análise de Diagnóstico; Coelhos Europeus

Abstract

The main objective in this work is to present a review of the normal nonlinear models, contemplating their classes, the least squares method to estimate its parameters, emphasizing the iterative method of Gauss-Newton and Newton-Raphson. Still, some asymptotic results, measures of nonlinearity and diagnostic techniques are presented. A characteristic of the nonlinear models is that diagnostic techniques used in linear models are inadequate if the degree of nonlinearity is high. In this work, we studied the existing nonlinearity measures for the verification of these assumptions and we use them in our applications. Finally, we analyze the data set *European Rabbits* to demonstrate the applicability of a nonlinear model proposed in the literature. Also, to the same data set we adjust other nonlinear models, including estimation methods for nonparametric curves, for comparison purposes. We conclude that in nonlinear models there is a strong link between the nonlinear functional and the phenomenon under study.

Keywords: Nonlinear Models; Iterative Methods; Diagnostic Analysis; European Rabbits

Sumário

Lista de Figuras

Lista de Tabelas

1	Introdução	p. 14
2	Modelo Normal Não Linear	p. 16
2.1	Definição	p. 16
2.2	Classes de Modelos	p. 17
2.2.1	Modelos de Crescimento	p. 17
2.2.2	Modelos de Rendimento	p. 18
2.2.3	Modelos Compartimentados	p. 18
2.3	Métodos de Estimação	p. 19
2.3.1	Método dos Mínimos Quadrados	p. 20
2.3.2	Método dos Mínimos Quadrados Generalizados	p. 20
2.3.3	Método da Máxima Verossimilhança	p. 21
2.3.4	Métodos Iterativos	p. 21
2.3.5	A importância de bons valores iniciais	p. 23
2.4	Resultados Assintóticos	p. 24
2.5	Medidas de não linearidade	p. 26
2.5.1	Medida de Vício de Box	p. 26
2.5.2	Medidas de curvatura de Bates e Watts	p. 27
2.6	Técnicas de Diagnóstico	p. 29

2.6.1	Matriz de Projeção	p. 29
2.6.2	Resíduo Projetado	p. 29
2.6.3	Medidas de Influência	p. 31
3	Aplicação	p. 33
3.1	Estimação e Ajuste	p. 34
3.2	Modelos Alternativos	p. 38
3.2.1	Smoothing Splines	p. 38
3.2.2	Modelo Logístico e de Gompertz	p. 39
4	Conclusão	p. 42
	Referências	p. 43
	Apêndice A Programação no <i>software</i> R	p. 46
	Anexo A Banco de Dados	p. 58

Lista de Figuras

2.1	Um diagrama do compartimento ou sistema para o modelo de tetraciclina. . .	p. 19
3.1	Gráfico do modelo normal ajustado aos dados sobre Coelhos Europeus. . . .	p. 35
3.2	Gráfico dos elementos da diagonal da matriz de projeção, dos resíduos ordinários, estudentizados e dos resíduos projetados estudentizados <i>versus</i> a ordem das observações.	p. 36
3.3	Gráfico da Distância de Cook <i>versus</i> ordem das observações.	p. 36
3.4	Modelo ajustado via Smoothing Splines comparado com o modelo ajustado como em (3.1). O Intervalo de Confiança foi obtido como em (2.9).	p. 39
3.5	Gráfico de dispersão das observações com as curvas ajustadas dos modelos Logístico, Gompertz e (3.1).	p. 40
3.6	Gráfico dos elementos da diagonal da matriz de projeção e dos resíduos estudentizados <i>versus</i> a ordem das observações do modelo Logístico.	p. 41
3.7	Gráfico dos elementos da diagonal da matriz de projeção e dos resíduos estudentizados <i>versus</i> a ordem das observações do modelo Gompertz.	p. 41

Lista de Tabelas

2.1	Modelos de regressão não linear utilizados para descrever curvas de crescimento	p. 17
2.2	Modelos de regressão não linear utilizados para descrever curvas de rendimento	p. 18
3.1	Estimativas para o modelo (3.1) com os respectivos erros padrão (assintóticos).	p. 35
3.2	Variação das estimativas do modelo não linear normal ao excluir as observações citadas na análise de diagnóstico. Os erros-padrão assintóticos estão entre parênteses.	p. 37
3.3	Estimativas para os modelos Logístico e Gompertz com os respectivos erros padrão (assintóticos).	p. 40
A.1	Pesos das lentes dos olhos de coelhos europeus, y (em mg) e idade, x (em dias) numa amostra de 71 observações	p. 58

1 Introdução

A análise de dados através do modelo clássico de regressão, também, denominado modelo normal linear, é uma das técnicas mais usadas de estimação. Porém, em muitas situações práticas, algumas de suas suposições, como a normalidade e a linearidade nos parâmetros, não são satisfeitas. Este fato, alavancou o desenvolvimento de novas técnicas estatísticas para os modelos de regressão, surgindo, então, novas classes de modelos, os modelos de regressão não linear e os modelos lineares generalizados.

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. Nelder e Wedderburn (1972) ampliaram a distribuição da variável de resposta para a família exponencial de distribuições, definindo os Modelos Lineares Generalizados. Mesmo assim, os modelos normais não lineares continuaram recebendo um tratamento especial, surgindo diversos artigos científicos na década de 70 e em décadas posteriores. Particularmente, destaca-se o livro de Ratkowsky (1983), que descreve vários modelos normais não lineares segundo diversos aspectos.

Dentro deste contexto, podemos citar os seguintes trabalhos: Cordeiro et al. (2000), propuseram uma correção das estimativas de máxima verossimilhança na classe simétrica de modelos de regressão não linear. Cysneiros e Vanegas (2008) fizeram um estudo analítico e empírico das propriedades estatísticas dos resíduos nos modelos simétricos de regressão não linear. Vanegas e Cysneiros (2010), propõem procedimentos de diagnóstico baseados no método de eliminação de casos para modelos de regressão não linear simétricos. Cancho et al. (2010) introduzem o modelo de regressão não linear skew-normal (SN-NLM) e apresentam uma análise inferencial completa.

Um modelo é classificado como não linear se pelo menos uma das derivadas parciais da função não linear em relação aos parâmetros depende de pelo menos um dos parâmetros do modelo. Uma das principais características desses modelos é que os mesmos em geral são deduzidos a partir de suposições teóricas e os parâmetros resultantes são interpretá-

veis. Assim, aproximá-los para os modelos normais lineares, mesmo que sejam alcançados ajustes satisfatórios, prejudicaria bastante a obtenção de estimativas mais realistas dos parâmetros de interesse.

Diferentemente dos modelos de regressão lineares, em que a qualidade e principalmente a validade do ajuste são simplesmente avaliadas por meio de diagnósticos de regressão, no caso não linear, além de diagnósticos usuais, medidas de não linearidade devem ser calculadas. Estas medidas, particulares dos modelos de regressão não lineares, são úteis na análise da extensão do comportamento não linear. Modelos não lineares com comportamento distante do comportamento linear podem ter seus resultados assintóticos invalidados, principalmente em situações em que pequenas amostras são disponíveis.

Neste trabalho temos como objetivo fazer uma revisão dos modelos não lineares, com foco no modelo normal não linear. Métodos de estimação, resultados assintóticos, medidas de não linearidade e técnicas de diagnósticos serão discutidos. Posteriormente, uma aplicação do modelo normal não linear a um conjunto de dados reais é realizada.

Neste primeiro capítulo apresentamos toda a estrutura e objetivos do trabalho. No Capítulo 2 definimos o modelo normal não linear. Adicionalmente, apresentamos as classes dos modelos, os métodos para a obtenção da estimativa de mínimos quadrados dos parâmetros que, nesse caso, coincide com o método de máxima verossimilhança, e alguns resultados assintóticos. Apresenta-se, ainda, medidas de não linearidade e algumas técnicas de diagnóstico. No Capítulo 3 apresentamos uma aplicação da metodologia discutida a um conjunto de dados reais bem como os seus resultados. Por fim, conclusões com base nos resultados obtidos são apresentados no Capítulo 4. No Apêndice encontram-se os códigos computacionais que foram implementados no desenvolvimento deste trabalho e no Anexo o banco de dados utilizado na análise.

As análises e gráficos desenvolvidos em todo o trabalho foram realizadas no ambiente de programação R em sua versão 2.15.3 para o sistema operacional Windows. Esta linguagem é distribuída gratuitamente e está disponível em <http://www.r-project.org/>.

2 Modelo Normal Não Linear

Os modelos lineares são bastante utilizados para um grande número de aplicações em regressão, com o objetivo de relacionar uma variável resposta a uma ou mais variáveis explicativas. Porém, em algumas situações um modelo não linear pode ser mais apropriado, uma vez que muitos fenômenos geralmente não são lineares.

O modelo de regressão não linear é frequentemente utilizado por pesquisadores em diversas áreas do conhecimento, tais como, agricultura, biologia, econometria, engenharia, química, etc. Na maioria das vezes, as formulações de possíveis modelos são deduzidos a partir de suposições teóricas inerentes ao fenômeno que se tem interesse modelar, e os parâmetros resultantes são interpretáveis. O modelo não linear com erros normais é um dos métodos mais utilizados quando a resposta é contínua, porque é de aplicação fácil e tem um grande número de teorias desenvolvidas para ele.

2.1 Definição

O modelo de regressão não linear pode ser escrito como

$$y_i = f(\mathbf{x}_i, \boldsymbol{\theta}) + \epsilon_i, \quad i = 1, \dots, n \quad (2.1)$$

em que y_i é a variável resposta, $f(\cdot)$ é uma função não linear contínua, com forma conhecida, do vetor de variáveis explicativas \mathbf{x}_i , e dos parâmetros desconhecidos $\boldsymbol{\theta}$. Neste trabalho os erros aleatórios, ϵ_i , são considerados independentes e identicamente distribuídos, normais com média zero e variância constante σ^2 , isto é, $\epsilon_i \stackrel{iid}{\sim} N(\mathbf{0}, \sigma^2 \mathbf{I})$.

A função esperança ou função de regressão é dada por $E(y_i|x_i) = f(\mathbf{x}_i, \boldsymbol{\theta})$. Segundo Bates e Watts (1988) para os modelos de regressão não linear, *pelo menos uma das derivadas parciais da função f em relação aos parâmetros depende de pelo menos um dos parâmetros do modelo*, ou seja, pode-se dizer que um modelo é não linear quando ele é não linear nos parâmetros.

2.2 Classes de Modelos

Nesta seção iremos apresentar algumas classes de modelos não lineares que, historicamente, têm proporcionado uma boa motivação para o desenvolvimento estatístico em modelos de regressão não linear.

2.2.1 Modelos de Crescimento

Segundo Seber e Wild (2003), a análise dos dados sobre crescimento é importante em muitas áreas de pesquisa. Na biologia, o interesse está na descrição do crescimento de animais e plantas na tentativa de compreender seus mecanismos essenciais. Os químicos estão interessados na formulação do produto de uma reação química ao longo do tempo. Na agricultura, são óbvias as vantagens econômicas e administrativas em conhecer como a produção cresce, o quão rápido ela cresce, e como esses fatores respondem a condições ou tratamentos ambientais. O crescimento infantil é de interesse na medicina, assim como o crescimento de tumores e os efeitos de tratamentos sobre tais crescimentos. Cientistas sociais têm interesse em crescimento da população, no fornecimento de alimentos e na demanda de energia.

O termo curva de crescimento usualmente evoca a imagem de curvas sigmoidais que representam o tempo de vida de medidas de dimensão, geralmente de altura e peso. Na modelagem de dados de crescimento, desejamos obter informações sobre a interpretação física dos parâmetros, a fim de construir um modelo padrão para as observações em estudo.

Para modelar esse tipo de comportamento, a literatura apresenta vários modelos de crescimento, ver Seber e Wild (2003). Na Tabela 2.1, são apresentados alguns dos principais modelos de crescimento.

Tabela 2.1: Modelos de regressão não linear utilizados para descrever curvas de crescimento

Modelo	Função
Logístico	$f(x) = \alpha / [1 + e^{(\beta - \gamma x)}]$
Gompertz	$f(x) = \alpha e^{(-e^{(\beta - \gamma x)})}$
Von Bertalanffy	$f(x) = \alpha [1 - e^{(-\gamma(x - \delta))}]$
Richards	$f(x) = \alpha / (1 + e^{(\beta - \gamma x)})^{1/\delta}$
Weibull	$f(x) = \alpha - \beta e^{-\gamma x^\delta}$
Morgan-Mercer-Flodin (MMF)	$f(x) = (\beta\gamma + \alpha x^\delta) / (\gamma + x^\delta)$

De acordo com Prudente (2009), nesses modelos o parâmetro α é o valor máximo esperado para a resposta, ou assíntota. O parâmetro β está relacionado com o intercepto, isto é, o valor de $\mu = E(y)$ correspondente a $x = 0$. O parâmetro γ está relacionado com a taxa média de crescimento da curva e, finalmente, o parâmetro δ , que aparece em alguns modelos, é utilizado para aumentar a flexibilidade dos mesmos no ajuste dos dados.

2.2.2 Modelos de Rendimento

Vários modelos, que estão fortemente relacionados com as curvas de crescimento encontram-se na agricultura, onde eles são utilizados para quantificar a relação entre o rendimento total de uma cultura (y) e a densidade da cultura (x).

Por causa da competição entre plantas por recursos, o rendimento por planta (y/x) tende a diminuir com o aumento da densidade de plantio. O agrônomo tende a interessar-se em modelos de previsão de rendimento, para encontrar a densidade de plantio que deu origem ao rendimento máximo e para a comparação das relações em diferentes condições. Na Tabela 2.2 são apresentados alguns modelos de rendimento.

Tabela 2.2: Modelos de regressão não linear utilizados para descrever curvas de rendimento

Modelo	Função
Holliday et al. (1960)	$f(x) = x/(\alpha + \beta x + \gamma x^2)$
Bleasdale e Nelder (1960)	$f(x) = x(\alpha + \beta x^\phi)^{-1/\theta}$
Ricker (1954)	$f(x) = \alpha x e^{-x/k}$
Hiperbólico	$f(x) = 1/(\alpha + \beta x)$

2.2.3 Modelos Compartmentados

Jacquez (1972) definiu um sistema compartimentado como “um sistema que é composto por um número finito de subsistemas macroscópicos, chamado compartimentos ou categorias, cada um dos quais é homogêneo e bem misturado, e os compartimentos interagem trocando materiais. Pode haver entradas do ambiente em um ou mais dos compartimentos, e pode haver saídas (excreção) de um ou mais compartimentos para o ambiente”.

Modelos compartimentados são uma ferramenta importante em vários ramos da medicina, tais como fisiologia, (JACQUEZ, 1972); (RESCIGNO et al., 1983), e farmacocinética (o estudo de como os medicamentos circulam através do corpo ao longo do tempo - Wag-

ner (1975)) e em muitos outros ramos da biologia, Atkins (1969). Eles também são úteis em química e bioquímica, em áreas especializadas da agricultura como fertilizantes, no estudo dos movimentos de população e da propagação de epidemias.

Exemplo 2.1. *Como um exemplo de modelos compartimentados, consideramos os dados sobre a concentração de cloridrato de tetraciclina no soro. Um composto de tetraciclina foi administrado por via oral a um sujeito, e a concentração de cloridrato de tetraciclina no soro foi medida durante um período de 16 horas, (WAGNER, 1967).*

O sistema biológico pode ser modelado por um compartimento de intestino em que o produto químico é introduzido, um compartimento de sangue, que absorve a substância química a partir do intestino, e um caminho de eliminação. Assumindo cinética de primeira ordem, as concentrações $(\gamma_1(t), \gamma_2(t))'$ de cloridrato de tetraciclina em dois compartimentos pode ser descrito pelo seguinte par de equações diferenciais

$$\begin{aligned}\frac{\partial \gamma_1(t)}{\partial t} &= \dot{\gamma}_1 = -\theta_1 \gamma_1(t) \\ \frac{\partial \gamma_2(t)}{\partial t} &= \dot{\gamma}_2 = \theta_1 \gamma_1(t) - \theta_2 \gamma_2(t)\end{aligned}$$

onde o ponto indica a diferenciação com respeito ao tempo. O sistema pode ser representado graficamente como um compartimento ou diagrama de sistema como mostrado na Figura 2.1.

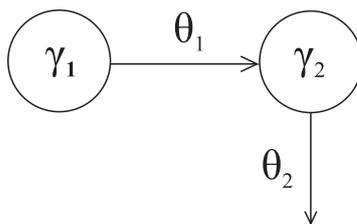


Figura 2.1: Um diagrama do compartimento ou sistema para o modelo de tetraciclina.

2.3 Métodos de Estimação

A obtenção das estimativas dos parâmetros pode ser realizada de várias maneiras dentre elas por meio do método dos mínimos quadrados ou pelo método da máxima verossimilhança. Para os modelos de regressão não lineares, o sistema de equações normais não é resolvido facilmente, sendo necessários métodos iterativos na obtenção dessas

estimativas. A seguir são apresentados alguns métodos de estimação.

2.3.1 Método dos Mínimos Quadrados

Assumindo o modelo descrito em (2.1), a estimativa de mínimos quadrados de $\boldsymbol{\theta}$, denotado por $\hat{\boldsymbol{\theta}}$, minimiza a soma de quadrados dos erros

$$SQRes(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^2 \quad (2.2)$$

sobre $\boldsymbol{\theta} \in \Theta$.

Para encontrar o estimador de mínimos quadrados, é necessário fazer a diferenciação de (2.2) com relação a cada um dos parâmetros e igualar as equações a zero, isto é,

$$\left. \frac{\partial SQRes(\boldsymbol{\theta})}{\partial \theta_r} \right|_{\hat{\boldsymbol{\theta}}} = 0 \quad (r = 1, 2, \dots, p)$$

$$\sum_{i=1}^n [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})] \left. \frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_r} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \quad (r = 1, 2, \dots, p) \quad (2.3)$$

Procedendo desta forma, determina-se um sistema de p equações normais (2.3) para o modelo não linear. Para a maioria dos modelos não lineares o sistema de equações normais não podem ser resolvidos analiticamente, de modo que são necessários métodos iterativos. (ver Seção 2.3.4)

2.3.2 Método dos Mínimos Quadrados Generalizados

Segundo Gallant (1987) este método é utilizado quando os erros são autocorrelacionados. A soma de quadrados dos resíduos a ser minimizada é dada por

$$SQRes(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]' \mathbf{V}^{-1} [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})] \quad (2.4)$$

onde \mathbf{V} é uma matriz conhecida, positiva definida.

Denote por $\hat{\boldsymbol{\theta}}_G$ a estimativa de mínimos quadrados generalizados que minimiza (2.4). Seja $\mathbf{V} = \mathbf{U}'\mathbf{U}$ uma decomposição de Cholesky de \mathbf{V} , onde \mathbf{U} é uma matriz triangular superior. Multiplicando o modelo não linear (2.1) por $\mathbf{R} = (\mathbf{U}')^{-1}$, obtemos

$$\mathbf{z} = \mathbf{k}(\boldsymbol{\theta}) + \boldsymbol{\eta},$$

onde $\mathbf{z} = \mathbf{R}\mathbf{y}$, $\mathbf{k}(\boldsymbol{\theta}) = \mathbf{R}\mathbf{f}(\mathbf{x}, \boldsymbol{\theta})$ e $\boldsymbol{\eta} = \mathbf{R}\boldsymbol{\epsilon}$. Logo, a soma de quadrados descrita por (2.4) pode ser rescrita como

$$SQRes(\boldsymbol{\theta}) = [\mathbf{z} - \mathbf{k}(\boldsymbol{\theta})]'[\mathbf{z} - \mathbf{k}(\boldsymbol{\theta})]$$

A partir deste momento, o vetor de estimativas $\hat{\boldsymbol{\theta}}_G$ passa a ser um vetor de estimativas de mínimos quadrados para os dados transformados, pois esta transformação utilizando a decomposição de Cholesky, permite tratar os dados como de um caso de variâncias homogêneas com observações independentes. Maiores detalhes sobre esse procedimento podem ser encontrados em Seber e Wild (2003) e Gallant (1987).

2.3.3 Método da Máxima Verossimilhança

Se os ϵ_i são independentes e identicamente distribuídos (i.i.d) com distribuição $N(\mathbf{0}, \sigma^2\mathbf{I})$, então a função de verossimilhança L é dada por

$$L(\boldsymbol{\theta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^2 \right\} \quad (2.5)$$

Ignorando as constantes, a função log-verossimilhança de (2.5) é dada por

$$\begin{aligned} \ell(\boldsymbol{\theta}, \sigma^2) &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - f(\mathbf{x}_i, \boldsymbol{\theta})]^2 \\ &= -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} SQRes(\boldsymbol{\theta}) \end{aligned} \quad (2.6)$$

Dado σ^2 (2.6) é maximizada com respeito a $\boldsymbol{\theta}$ quando $SQRes(\boldsymbol{\theta})$ é minimizado, isto é, quando $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ (estimativa de mínimos quadrados), (SEBER; WILD, 2003). Além disso, $\partial\ell/\partial\sigma^2 = 0$ tem solução $\sigma^2 = SQRes(\boldsymbol{\theta})/n$, o que dá um valor máximo (para dado $\boldsymbol{\theta}$) quando a segunda derivada é negativa. Isto sugere que $\hat{\boldsymbol{\theta}}$ e $\hat{\sigma}^2 = SQRes(\hat{\boldsymbol{\theta}})/n$ são estimativas de máxima verossimilhança. Podemos utilizar o algoritmo EM como uma solução numérica das estimativas de máxima verossimilhança dos parâmetros dos modelos não lineares, Monahan (2011).

2.3.4 Métodos Iterativos

Para minimizar $SQRes(\boldsymbol{\theta})$ em (2.2) ou (2.4), quanto para maximizar $L(\boldsymbol{\theta}, \sigma^2)$ é necessário o uso de algum método numérico iterativo. Na literatura vários métodos numéricos iterativos são propostos, como por exemplo, o método Newton-Raphson, método Gauss-Newton, método *Steepest-Descent*, método de Marquardt, entre outros.

O algoritmo de estimativa de regressão não linear mais comum é o método de Gauss-Newton, que se baseia em aproximações lineares para a função esperança $f(\mathbf{x}_i, \boldsymbol{\theta})$ a cada passo.

- **Método Gauss-Newton**

Suponha que $\boldsymbol{\theta}^{(a)}$ é uma aproximação da estimativa de mínimos quadrados $\hat{\boldsymbol{\theta}}$ de um modelo não linear. Para $\boldsymbol{\theta}$ próximo de $\boldsymbol{\theta}^{(a)}$, considera-se uma expansão em série de Taylor de primeira ordem como

$$\mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \approx \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}^{(a)}) + \mathbf{F}(\boldsymbol{\theta}^{(a)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})$$

onde $\mathbf{F}(\boldsymbol{\theta}) = \frac{\partial \mathbf{f}(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \left[\left(\frac{\partial f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_j} \right) \right]$. Definindo $\mathbf{r}(\boldsymbol{\theta})$ como sendo um vetor de resíduos, pode-se escrever

$$\begin{aligned} \mathbf{r}(\boldsymbol{\theta}) &= \mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) \\ &\approx \mathbf{r}(\boldsymbol{\theta}^{(a)}) - \mathbf{F}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}), \end{aligned}$$

onde $\mathbf{F}^{(a)} = \mathbf{F}(\boldsymbol{\theta}^{(a)})$. Substituindo $\mathbf{r}'(\boldsymbol{\theta})\mathbf{r}(\boldsymbol{\theta})$ em $SQRes(\boldsymbol{\theta})$ obtém-se

$$SQRes(\boldsymbol{\theta}) \approx \mathbf{r}'(\boldsymbol{\theta}^{(a)})\mathbf{r}(\boldsymbol{\theta}^{(a)}) - 2\mathbf{r}'(\boldsymbol{\theta}^{(a)})\mathbf{F}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})'\mathbf{F}'^{(a)}\mathbf{F}^{(a)}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}).$$

Portanto, $SQRes(\boldsymbol{\theta})$ será minimizada quando

$$\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)} = [\mathbf{F}'^{(a)}\mathbf{F}^{(a)}]^{-1}\mathbf{F}'^{(a)}\mathbf{r}(\boldsymbol{\theta}^{(a)}).$$

Assim, devido à aproximação $\boldsymbol{\theta}^{(a)}$, a próxima aproximação é dada por

$$\boldsymbol{\theta}^{(a+1)} = \boldsymbol{\theta}^{(a)} + [\mathbf{F}'^{(a)}\mathbf{F}^{(a)}]^{-1}\mathbf{F}'^{(a)}\mathbf{r}(\boldsymbol{\theta}^{(a)}), \quad (2.7)$$

resultando no processo iterativo conhecido como Método de Gauss-Newton. O processo (2.7) é repetido até obter a convergência.

Um outro método iterativo bastante utilizado é o de Newton-Raphson, que usa uma aproximação quadrática local para a função objetivo (função a ser minimizada), Bates e Watts (1988).

- **Método Newton-Raphson**

Perto de $\boldsymbol{\theta}^{(a)}$, aproximamos

$$SQRes(\boldsymbol{\theta}) \approx SQRes(\boldsymbol{\theta}^{(a)}) + \mathbf{U}'(\boldsymbol{\theta}^{(a)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})'\frac{\mathbf{H}(\boldsymbol{\theta}^{(a)})}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}),$$

onde $\mathbf{U}(\boldsymbol{\theta}) = \frac{\partial SQRes(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ é o vetor gradiente e $\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2 SQRes(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ é a matriz Hessiana de $SQRes(\boldsymbol{\theta})$. A aproximação da função de soma de quadrados terá um ponto estacionário quando seu gradiente é zero, isto é, quando

$$\mathbf{U}(\boldsymbol{\theta}^{(a)}) + \mathbf{H}(\boldsymbol{\theta}^{(a)})(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)}) = 0$$

e este ponto estacionário será um mínimo se $\mathbf{H}(\boldsymbol{\theta}^{(a)})$ é positiva definida. Se $\mathbf{H}(\boldsymbol{\theta}^{(a)})$ é positiva definida, o passo de Newton-Raphson é

$$\boldsymbol{\theta}^{(a+1)} = \boldsymbol{\theta}^{(a)} - [\mathbf{H}(\boldsymbol{\theta}^{(a)})]^{-1} \mathbf{U}(\boldsymbol{\theta}^{(a)}). \quad (2.8)$$

O processo (2.8) é repetido até obter a convergência.

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

$$|SQRes(\boldsymbol{\theta}^{(a+1)}) - SQRes(\boldsymbol{\theta}^{(a)})| < 0,000001.$$

Em Draper e Smith (1998) podem ser encontrados mais detalhes sobre o método *Steepest-Descent* e o método de Marquardt.

Vale notar que mesmo estabelecendo critérios para o término do processo iterativo, a convergência não é garantida. Valores iniciais mal escolhidos, muitas vezes, podem necessitar de um número muito grande de iterações para convergir. Além disso, pode também ocorrer convergência para um mínimo local, ou mesmo, a não convergência. Entretanto, quando os valores iniciais dos parâmetros são suficientemente próximos dos valores verdadeiros dos parâmetros, o método iterativo irá convergir em poucos passos e pode levar a um mínimo global, mesmo quando existir vários mínimos locais. Assim, a escolha dos valores iniciais para o método iterativo no contexto não linear tem um grande papel na estimativa dos parâmetros. Draper e Smith (1998) e Gallant (1987) apresentam algumas técnicas para a obtenção de bons valores iniciais para que viabilizem uma convergência rápida do processo de estimação.

2.3.5 A importância de bons valores iniciais

A escolha dos valores iniciais é muito importante em todos os procedimentos iterativos para a estimativa dos parâmetros, porque uma escolha pobre pode resultar em convergência lenta, a convergência para um mínimo local, ou mesmo divergência. Bons valores iniciais geralmente permitem uma técnica iterativa a convergir para uma solução muito mais rápida do que seria de outro modo possível. Além disso, se existirem múltiplos

mínimos ou se existem vários mínimos locais, além de um mínimo absoluto, os valores iniciais podem resultar em má convergência para um ponto estacionário indesejado da soma dos quadrados de superfície. Este ponto indesejado pode ter valores de parâmetros que são fisicamente impossíveis ou que não proporcionam o verdadeiro valor mínimo de $SQRes(\boldsymbol{\theta})$. No entanto, não existe qualquer procedimento padrão para obter estimativas iniciais. Uma variedade de métodos estão disponíveis para a obtenção de valores iniciais, o método mais frequente é através da utilização de uma informação prévia. Alguns outros métodos são os seguintes:

1. *Linearização*

Depois de ignorar o erro aleatório, veja a forma do modelo para ver se poderia ser transformada em uma forma linear por meio de uma transformação. Em tais casos, a regressão linear pode ser utilizada para obter os valores iniciais.

2. *Resolver um sistema de equações*

Se houver p parâmetros, substitui p conjuntos de observações (x_i, y_i) dentro do modelo postulado, *ignorando o erro*. Resolva as p equações resultantes para os parâmetros, se possível. Pontos amplamente separados, muitas vezes funcionam melhor.

3. *Usando as propriedades do modelo*

Considere o comportamento da função de resposta com o x_i indo para zero ou infinito, e substituir por observações que mais de perto representam as condições na escala e contexto do problema. Resolver, se possível, as equações resultantes.

4. *Método gráfico*

Algumas vezes, uma estimativa visual pode ser obtida através da representação gráfica dos dados.

2.4 Resultados Assintóticos

Segundo Seber e Wild (2003), assumindo que os erros, ϵ_i , são independentes e identicamente distribuídos com variância σ^2 , sob certas condições de regularidade, $\hat{\boldsymbol{\theta}}$ e $s^2 = SQRes(\hat{\boldsymbol{\theta}})/(n - p)$ são estimadores consistentes de $\boldsymbol{\theta}$ e σ^2 respectivamente; $\hat{\boldsymbol{\theta}}$ é também assintoticamente normalmente distribuído quando $n \rightarrow \infty$. Além disso, assumindo que ϵ_i são normalmente distribuídos, então $\hat{\boldsymbol{\theta}}$ é também um estimador de máxima verossimilhança.

Teorema 2.1. *Dado que $\boldsymbol{\epsilon} \stackrel{iid}{\sim} N(\mathbf{0}, \sigma^2 \mathbf{I})$, sob certas condições de regularidade, e n suficientemente grande, tem-se aproximadamente os seguintes resultados:*

- (i) $\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, \sigma^2 \mathbf{C})$, onde $\mathbf{C} = [\mathbf{F}'(\boldsymbol{\theta})\mathbf{F}(\boldsymbol{\theta})]^{-1} = [\mathbf{F}'\mathbf{F}]^{-1}$;
- (ii) $\frac{(n-p)s^2}{\sigma^2} \approx \frac{\boldsymbol{\epsilon}'(\mathbf{I}_n - \mathbf{P})\boldsymbol{\epsilon}}{\sigma^2} \sim \chi_{n-p}^2$, onde $\mathbf{P} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$;
- (iii) $\hat{\boldsymbol{\theta}}$ é estatisticamente independente de σ^2 ;
- (iv) $\frac{[SQRes(\boldsymbol{\theta}) - SQRes(\hat{\boldsymbol{\theta}})]/p}{SQRes(\hat{\boldsymbol{\theta}})/(n-p)} \approx \frac{\boldsymbol{\epsilon}'\mathbf{P}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}'(\mathbf{I}_n - \mathbf{P})\boldsymbol{\epsilon}} \cdot \frac{n-p}{p} \sim F_{p, n-p}$.

Uma vez obtida a estimativa dos parâmetros, é possível estabelecer um intervalo de confiança $100(1 - \alpha)\%$ para a resposta esperada,

$$f(x_i, \hat{\boldsymbol{\theta}}) \pm t_{\alpha/2} \sqrt{s^2 \hat{c}_{jj}}, \quad (2.9)$$

em que $t_{\alpha/2}$ é o quantil $(1 - \alpha/2)$ de uma distribuição t de Student com $n - p$ graus de liberdade e \hat{c}_{jj} indica o j -ésimo elemento da diagonal da matriz $\hat{\mathbf{C}}$.

Uma região de, aproximadamente, $100(1 - \alpha)\%$ de confiança para $\boldsymbol{\theta}$ foi proposta por Beale (1960), sendo formada pelos contornos de $SQRes(\boldsymbol{\theta})$ tais que

$$SQRes(\boldsymbol{\theta}) = SQRes(\hat{\boldsymbol{\theta}}) \left[1 + \frac{p}{n-p} F_{p, n-p}^\alpha \right],$$

em que $F_{p, n-p}^\alpha$ é o quantil $(1 - \alpha)$ de uma distribuição F .

Para testar uma hipótese não linear composta por $q (< p)$ restrições não lineares independentes, isto é, $H_0 : \mathbf{a}(\boldsymbol{\theta}) = [a_1(\boldsymbol{\theta}), a_2(\boldsymbol{\theta}), \dots, a_q(\boldsymbol{\theta})]' = \mathbf{0}$, três métodos de testes assintoticamente equivalentes estão acessíveis, o teste Razão-Verossimilhança (RV), teste de Wald (W) e o Multiplicador de Lagrange (ML). As três estatísticas de teste são expressas por

$$\begin{aligned} RV &= n[\log SQRes(\tilde{\boldsymbol{\theta}}) - \log SQRes(\hat{\boldsymbol{\theta}})], \\ W &= \frac{\mathbf{a}'(\hat{\boldsymbol{\theta}})[\hat{\mathbf{A}}(\hat{\mathbf{F}}'\hat{\mathbf{F}})^{-1}\hat{\mathbf{A}}']^{-1}\mathbf{a}(\hat{\boldsymbol{\theta}})}{SQRes(\hat{\boldsymbol{\theta}})/n}, \\ ML &= \frac{[\mathbf{y} - \mathbf{f}(\mathbf{x}, \tilde{\boldsymbol{\theta}})]'\tilde{\mathbf{F}}(\tilde{\mathbf{F}}'\tilde{\mathbf{F}})^{-1}\tilde{\mathbf{F}}'[\mathbf{y} - \mathbf{f}(\mathbf{x}, \tilde{\boldsymbol{\theta}})]}{SQRes(\tilde{\boldsymbol{\theta}})/n}, \end{aligned}$$

onde $\mathbf{A} = [(\frac{\partial a_i(\boldsymbol{\theta})}{\partial \theta_j})]$ e $\hat{\boldsymbol{\theta}}$ e $\tilde{\boldsymbol{\theta}}$ são as estimativas de máxima verossimilhança do modelo irrestrito e restrito, respectivamente. As três estatísticas são assintoticamente equivalentes e assintoticamente distribuídas como χ_q^2 quando H_0 é verdadeira.

2.5 Medidas de não linearidade

Medidas de não linearidade tem como objetivo verificar se o grau de não linearidade de um problema de estimação não linear é suficientemente pequeno para que as técnicas usuais de estimação, desenvolvidas para a regressão linear, sejam utilizadas como uma boa aproximação para o modelo não linear.

Diversos autores se preocuparam em estudar a não linearidade e propuseram medidas para a sua quantificação. A primeira tentativa relevante no sentido de desenvolver uma medida de não linearidade foi de Beale (1960), que propôs quatro medidas. De acordo com Guttman e Meeter (1965), essas medidas não devem ser usadas na prática, uma vez que elas tendem a subestimar o verdadeiro grau de não linearidade do modelo. Uma contribuição importante foi a de Box (1971) que calculou o viés de ordem n^{-1} do estimador de máxima verossimilhança (EMV) do vetor $\boldsymbol{\theta}$ de um modelo normal não linear. Gillis e Ratkowsky (1978), via simulação, concluíram que a medida de vício de Box não só estima o vício de maneira muito aproximada como também fornece uma boa indicação do comportamento não linear do modelo. Bates e Watts (1980), utilizando alguns conceitos de geometria diferencial, desenvolveram duas medidas de curvatura para os modelos normais não lineares. Essas medidas indicam, respectivamente, o grau de não linearidade intrínseca de um modelo e o grau de não linearidade aparente ou devida à parametrização utilizada.

Na prática, as medidas de vício de Box e as medidas de curvatura de Bates e Watts são as ferramentas mais empregadas na avaliação da não linearidade de um modelo, podendo também ser usadas em problemas de discriminação, uma vez que o melhor modelo possível, dentre todos os propostos, seria aquele que apresenta um comportamento mais próximo do linear, Ratkowsky (1983).

2.5.1 Medida de Vício de Box

Os estimadores de máxima verossimilhança em modelos não lineares tendem a ser viesados em função da não linearidade do modelo, (COOK et al., 1986). Cox e Snell (1968) deduziram o viés de ordem n^{-1} do estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\theta}$ em uma classe geral de modelos que inclui o modelo normal não linear como um caso especial. Box (1971), chegou à seguinte fórmula para cálculo de viés em

casos univariados homocedásticos

$$\mathbf{B}(\hat{\boldsymbol{\theta}}) = -\frac{\sigma^2}{2} \left(\sum_{i=1}^n \mathbf{F}(\boldsymbol{\theta}) \mathbf{F}'(\boldsymbol{\theta}) \right)^{-1} \sum_{i=1}^n \mathbf{F}(\boldsymbol{\theta}) \text{tr} \left[\left(\sum_{i=1}^n \mathbf{F}(\boldsymbol{\theta}) \mathbf{F}'(\boldsymbol{\theta}) \right)^{-1} \mathbf{H}(\boldsymbol{\theta}) \right], \quad (2.10)$$

em que $\mathbf{F}(\boldsymbol{\theta})$ é o vetor $(p \times 1)$ da primeira derivada do modelo e $\mathbf{H}(\boldsymbol{\theta})$ é a matriz $(p \times p)$ de segunda derivada, ambos com respeito aos elementos de $\boldsymbol{\theta}$, avaliados em x_i , $i = 1, \dots, n$. Na prática, são usados $\hat{\boldsymbol{\theta}}$ 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). Portanto, escrevendo a equação (2.10) na forma matricial, obtém-se

$$\mathbf{B}(\hat{\boldsymbol{\theta}}) = (\mathbf{F}'\mathbf{F})^{-1} \mathbf{F}'\mathbf{b}, \quad (2.11)$$

em que \mathbf{b} é um vetor $(n \times 1)$ com elementos $-\frac{\sigma^2}{2} \text{tr} [(\mathbf{F}'\mathbf{F})^{-1} \mathbf{H}_i]$. Por essa exposição pode-se dizer que $\mathbf{B}(\hat{\boldsymbol{\theta}})$ é um conjunto de coeficientes obtidos do ajuste de \mathbf{b} às colunas de \mathbf{F} por mínimos quadrados ordinários.

Cook et al. (1986) mostraram que \mathbf{b} é igual à diferença entre os valores esperados das aproximações linear e quadrática ao redor de $\hat{\boldsymbol{\theta}}$. Logo, o viés será pequeno se todos os elementos de \mathbf{b} forem suficientemente próximos de zero, o que indica que o modelo é, essencialmente, linear ou se \mathbf{b} é ortogonal às colunas de \mathbf{F} .

A porcentagem de viés é uma medida mais interessante, pois é livre de escala e é dada por

$$\%B(\hat{\boldsymbol{\theta}}) = \frac{100 \cdot \mathbf{B}(\hat{\boldsymbol{\theta}})}{\hat{\boldsymbol{\theta}}}. \quad (2.12)$$

Segundo Box (1971), um vício aceitável seria aquele em que $\mathbf{B}(\hat{\boldsymbol{\theta}}_i) < 1\%$ e que, em situações mais severas, deve-se fazer a correção das estimativas com translação dos intervalos de confiança na mesma quantidade. O autor afirma ainda que o viés é proporcional à matriz de covariância dos erros, conseqüentemente proporcional à matriz de covariância dos parâmetros e concluiu que o erro padrão das estimativas é mais importante que o vício e que um critério experimental que minimize a variância generalizada tenderá a reduzir o vício nas estimativas.

2.5.2 Medidas de curvatura de Bates e Watts

Usando conceitos de geometria diferencial, Bates e Watts (1980) estenderam as idéias de Beale (1960) e desenvolveram medidas de não linearidade baseadas no conceito geométrico de curvatura e propuseram a quantificação em duas componentes. Da teoria de

modelos lineares, um intervalo de confiança $1 - \alpha$ para $\boldsymbol{\theta}$ é obtido obedecendo a desigualdade

$$(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})' \mathbf{X}' \mathbf{X} (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \leq ps^2 F_{p,n-p}^\alpha, \quad (2.13)$$

em que \mathbf{X} é matriz do modelo e $\boldsymbol{\theta}$ o vetor de parâmetros. Estes mesmos autores, então estipularam a quantidade $\rho = s\sqrt{p}$ como raio padrão da esfera cujos valores internos $\boldsymbol{\theta}$ pertencem ao intervalo de confiança $1 - \alpha$.

Para esses autores, a validade da aproximação linear depende da magnitude do segundo termo da expansão de série de Taylor relativo ao primeiro termo e, para isso, dividiram o componente $\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})' \mathbf{H}(\boldsymbol{\theta}^{(a)}) (\boldsymbol{\theta} - \boldsymbol{\theta}^{(a)})$ em dois termos, de forma que um pertence ao plano da aproximação e o outro ao espaço complementar por um processo de projeção ortogonal. Essa decomposição pode ser obtida definindo-se uma matriz de projeção

$$\widehat{\mathbf{P}} = \widehat{\mathbf{F}} (\widehat{\mathbf{F}}' \widehat{\mathbf{F}})^{-1} \widehat{\mathbf{F}}',$$

e, dessa maneira, as duas componentes são dadas por

$$\widehat{\mathbf{H}} = \widehat{\mathbf{H}}^T + \widehat{\mathbf{H}}^N$$

em que T e N indicam as componentes tangencial e normal ao plano definido pela aproximação linear, respectivamente.

Esses autores, então, definiram as medidas comparando o termo quadrático com o linear da seguinte forma

$$K^T = \frac{\|(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})' \mathbf{H}^T (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})\|}{\|\mathbf{F}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})\|^2} \quad e \quad K^N = \frac{\|(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})' \mathbf{H}^N (\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})\|}{\|\mathbf{F}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})\|^2}, \quad (2.14)$$

em que K^T e K^N representam as curvaturas dos círculos que melhor aproximam do formato da superfície esperada em um modelo $\eta(\widehat{\boldsymbol{\theta}})$. Esses círculos têm raio $R = 1/K$ e, para obter uma medida invariante à escala da variável resposta, adotaram padronizar essas medidas pelo raio padrão ρ . Assim

$$\Gamma^T = K^T \rho \quad e \quad \Gamma^N = K^N \rho$$

são medidas adotadas para julgar a não linearidade das componentes do modelo em que Γ^T está associada a desvios de uniformidade e paralelismo do plano tangente correspondente à parametrização do modelo e, por isso, denominada de curvatura devido ao efeito de parâmetros, e Γ^N está associada a desvios da planicidade da superfície esperada e invariante à reparametrização, denominada de curvatura intrínseca do modelo.

Bates e Watts (1980) sugerem o critério $\Gamma > 2(F_{p,n-p}^\alpha)^{-1/2}$ como guia para indicar se o modelo ajustado tem uma curvatura aparentemente acentuada. Vale ressaltar que essas medidas apresentadas são obtidas para uma direção definida no espaço em que um processo iterativo é necessário para que se obtenha a maior curvatura. Bates e Watts (1980) apresentam a relação que existe entre a medida proposta e os trabalhos de Beale (1960) e Box (1971), enquanto Ratkowsky (1983) apresentou um algoritmo em FORTRAN para a obtenção da máxima curvatura.

2.6 Técnicas de Diagnóstico

Muitos fenômenos podem ser representados por meio de modelos estatísticos. Para validar tais modelos é necessário verificar se as suposições envolvidas estão satisfeitas e avaliar a qualidade e performance dos ajustes dos modelos em questão. A seguir apresentamos e discutimos alguns métodos de diagnóstico.

2.6.1 Matriz de Projeção

No modelo normal não linear, para a detecção de pontos mais afastados dos demais, possivelmente, pontos influentes, utiliza-se a matriz de projeção definida como

$$\hat{\mathbf{P}} = \mathbf{F}(\hat{\boldsymbol{\theta}}) [\mathbf{F}'(\hat{\boldsymbol{\theta}})\mathbf{F}(\hat{\boldsymbol{\theta}})]^{-1} \mathbf{F}'(\hat{\boldsymbol{\theta}}).$$

Esses pontos afastados, além de serem alavancas e influentes, em geral exercem grande influência sobre a matriz de variâncias e covariâncias. Ao contrário da regressão linear, essa é uma matriz de projeção local, pois depende de $\boldsymbol{\theta}$ e deve ser estimada em $\hat{\boldsymbol{\theta}}$. Mesmo assim, o critério $\hat{p}_{ii} \geq 2p/n$ continua sendo adotado como guia para detectar pontos suspeitos de serem influentes.

2.6.2 Resíduo Projetado

Os resíduos ordinários no modelo não linear são definidos por $r_i = y_i - f(\mathbf{x}_i, \boldsymbol{\theta})$, $i = 1, \dots, n$. Em alguns casos, esses resíduos podem produzir resultados enganosos quando usados em métodos de diagnósticos análogos aos da regressão linear, não refletindo de forma correta a distribuição dos erros. Cook e Tsai (1985) apresentaram um novo tipo de resíduo que pode superar muitas deficiências dos resíduos usuais.

Seja $C(\mathbf{F})$ o subespaço gerado pelas colunas de \mathbf{F} . De agora em diante, usa-se a

notação $\overline{C}(\mathbf{F})$ para representar o subespaço ortogonal a $C(\mathbf{F})$, ou seja, o subespaço gerado pelas colunas ortogonais àquelas de \mathbf{F} .

Definem-se as matrizes $p \times p$, $\mathbf{W}_i = \frac{\partial^2 f(\mathbf{x}_i, \boldsymbol{\theta})}{\partial \theta_r \partial \theta_s}$, $r, s = 1, \dots, p$, de derivadas de segunda ordem para $i = 1, \dots, n$. Define-se ainda \mathbf{W} como uma matriz $n \times (p \times p)$ com n faces cuja i -ésima face é simplesmente igual a \mathbf{W}_i .

É apresentado por Cook e Tsai (1985) uma aproximação para o resíduo ordinário \mathbf{r} , dado da seguinte maneira

$$\mathbf{r} \cong (\mathbf{I} - \mathbf{P})\boldsymbol{\epsilon} - \mathbf{F} \sum_{i=1}^n r_i \mathbf{W}_i \boldsymbol{\delta} - \frac{1}{2}(\mathbf{I} - \mathbf{P})\boldsymbol{\delta}'\mathbf{W}\boldsymbol{\delta}, \quad (2.15)$$

tal que $\boldsymbol{\delta} = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, $\boldsymbol{\delta}'\mathbf{W}\boldsymbol{\delta}$ representa $\boldsymbol{\delta}'\mathbf{W}_i\boldsymbol{\delta}$ para produzir a i -ésima componente de \mathbf{r} .

A matriz \mathbf{W} apresenta p faces, e cada uma de suas faces é uma matriz $n \times p$, em que, n é o número de observações e p o de parâmetros. Projeta-se cada coluna de derivadas de segunda ordem de \mathbf{W} sobre $\overline{C}(\mathbf{F})$, tem-se $\mathbf{P}_1 = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'$ que é o operador de projeção ortogonal do espaço gerado por sobre essa projeção. $\mathbf{S} = (\mathbf{I} - \mathbf{P})\mathbf{T}$, em que \mathbf{T} é uma matriz $n \times q$ obtida pelos vetores não nulos, de derivadas de segunda ordem e $\mathbf{P} = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$.

Apresenta-se \mathbf{V} uma matriz $n \times (p+q)$ definida como $\mathbf{V} = (\mathbf{F}, \mathbf{S})$ e $\mathbf{P}_2 = \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'$ o operador de projeção ortogonal em $C(\mathbf{V})$.

Utilizando as matrizes descritas anteriormente, Cook e Tsai (1985) definem o resíduo projetado dado por

$$(\mathbf{I} - \mathbf{P}_2)\mathbf{r} = (\mathbf{I} - \mathbf{P})\boldsymbol{\epsilon} - (\mathbf{I} - \mathbf{P}_1)\boldsymbol{\epsilon}. \quad (2.16)$$

O primeiro termo de (2.16) é a aproximação linear para o resíduo ordinário \mathbf{r} , enquanto o segundo termo reflete a perda de informação necessária para remover os componentes não lineares de (2.15). Se $q = \text{posto}(\mathbf{P}_1)$ for pequeno comparado a $n - p$, essa perda será pequena. Independente disso, se as medidas de não linearidade intrínseca e de efeito do parâmetro forem significativas, ou seja, $\Gamma^N > 2(F_{p,n-p}^\alpha)^{-1/2}$, o ganho será considerável. Além disso, de (2.16), mostra-se que

$$E\{(\mathbf{I} - \mathbf{P}_2)\mathbf{r}\} = \mathbf{0}, \quad \text{Var}\{(\mathbf{I} - \mathbf{P}_2)\mathbf{r}\} = \sigma^2(\mathbf{I} - \mathbf{P}_2)$$

e

$$E\{\mathbf{r}'(\mathbf{I} - \mathbf{P}_2)\mathbf{r}\} = \sigma^2 \text{tr}(\mathbf{I} - \mathbf{P}_2).$$

Logo, uma estimativa alternativa para σ^2 é expressa por

$$\tilde{\sigma}^2 = \frac{\mathbf{r}'(\mathbf{I} - \hat{\mathbf{P}}_2)\mathbf{r}}{\text{tr}(\mathbf{I} - \hat{\mathbf{P}}_2)}.$$

Em geral, os resíduos projetados apresentam maior qualidade de diagnóstico do que os ordinários, e, algumas das técnicas de diagnóstico utilizadas na regressão linear, são válidas para esses resíduos.

Na regressão linear, mesmo para erros não-correlacionados e de variância constante, os resíduos são correlacionados e têm variâncias diferentes. Definem-se, então, os resíduos estudentizados que mesmo correlacionados, apresentam média zero e variância constante e igual a um. O i -ésimo resíduo ordinário estudentizado é expresso por

$$t_i = \frac{r_i}{s(1 - p_{ii})^{1/2}}, \quad i = 1, \dots, n.$$

Os resíduos projetados estudentizados têm esperança nula e variância $\sigma^2(\mathbf{I} - \mathbf{P}_2)$ e podem ser, então, definidos, como

$$s_i = \frac{\{(\mathbf{I} - \hat{\mathbf{P}}_2)\mathbf{r}\}_i}{\tilde{\sigma}\{(\mathbf{I} - \hat{\mathbf{P}}_2)\}_{ii}^{1/2}}, \quad i = 1, \dots, n. \quad (2.17)$$

De acordo com Prudente (2009) para avaliar se os erros ϵ'_i s têm distribuição, aproximadamente, normal, assim como para detectar se há pontos aberrantes e/ou influentes, o gráfico de probabilidades dos resíduos projetados ordenados $s_{(i)}$ versus $\Phi^{-1}\left(\frac{i-3/8}{n+1/4}\right)$ pode ser útil, sendo $\Phi(\cdot)$ a função de distribuição acumulada da normal padrão. A análise dos resíduos em (2.17) procede-se, similarmente, ao modelo normal linear.

2.6.3 Medidas de Influência

Um tópico de grande importância na análise de diagnóstico é a detecção de observações influentes, isto é, pontos que exercem um peso desproporcional nas estimativas dos parâmetros ou até mesmo na significância dos parâmetros. As técnicas mais conhecidas para detectar esse tipo de influência são baseadas na eliminação de um único ponto, as quais procuram avaliar o impacto da retirada de uma observação particular nas estimativas da regressão.

As medidas de influência para o modelo normal não linear são baseadas na regressão linear. A única diferença, que pode ser relevante, é a substituição da estimativa $\hat{\theta}_{(i)}$ pela

estimativa correspondente $\hat{\boldsymbol{\theta}}_{(i)}^1$, que é obtida inicializando o processo iterativo em $\hat{\boldsymbol{\theta}}$ sem a i -ésima observação e tomando a estimativa de um passo.

Cook e Weisberg (1982) mostraram que essa estimativa de um passo é dada por

$$\hat{\boldsymbol{\theta}}_{(i)}^1 = \hat{\boldsymbol{\theta}} - \frac{(\hat{\mathbf{F}}' \hat{\mathbf{F}})^{-1}}{(1 - \hat{p}_{ii})} \mathbf{f}_i r_i ,$$

em que \mathbf{f}_i é a i -ésima linha de \mathbf{F} . Para calcular o valor numérico de $\hat{\boldsymbol{\theta}}_{(i)}^1$, a matriz \mathbf{F} deve ser avaliada em $\hat{\boldsymbol{\theta}}$.

Quando tivermos pequenas amostras, a aproximação $\hat{\boldsymbol{\theta}}_{(i)}^1$ poderá ficar muito afastada de $\hat{\boldsymbol{\theta}}_{(i)}$ se $L(\boldsymbol{\theta})$ não for localmente quadrática, vide, por exemplo, Cook e Weisberg (1982). Entretanto, vários estudos de simulação mostraram que essa aproximação é suficiente para detectar os pontos mais influentes, Pregibon (1981).

Uma outra medida de influência, conhecida como distância de Cook, foi proposta por Cook (1977) para modelos normais lineares e rapidamente assimilada e estendida para diversas classes de modelos, sendo definida na classe de modelos normais não lineares por

$$D_i = (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)})' (\hat{\mathbf{F}}' \hat{\mathbf{F}}) (\hat{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}_{(i)}) / p \hat{\sigma}^2 . \quad (2.18)$$

Usando as fórmulas anteriores de t_i e $\hat{\boldsymbol{\theta}}_{(i)}^1$ na expressão (2.18), obtém-se a forma aproximada

$$D_i^1 = \frac{t_i^2}{p} \frac{\hat{p}_{ii}}{(1 - \hat{p}_{ii})}, \quad i = 1, \dots, n.$$

O gráfico de D_i^1 versus a ordem das observações tem sido útil para avaliar a influência das observações nas estimativas dos parâmetros. Se o interesse é detectar pontos influentes nas estimativas individuais $\hat{\theta}_j$, $j = 1, \dots, p$, sugere-se o gráfico de

$$\Delta_i \hat{\theta}_j = \frac{\hat{\theta}_j - \hat{\theta}_{(i)j}}{DP(\hat{\theta}_j)}$$

versus a ordem das observações, $DP(\cdot)$ denota o desvio padrão.

3 Aplicação

Neste Capítulo apresentamos uma aplicação da metodologia discutida no Capítulo 2 a um conjunto de dados reais. A aplicação será para a classe de modelos de crescimento, discutida anteriormente. Toda parte computacional foi resolvida utilizando o software R e os comandos encontram-se no Apêndice A.

Os dados dessa aplicação foram originalmente apresentados em Dudzinski e Mykutowycz (1961) e estudados posteriormente por Ratkowsky (1983) baseados em um modelo de regressão normal não linear. Os dados consistem em um conjunto de 71 observações em que a variável resposta representa o peso das lentes (em mg) dos olhos de coelhos europeus (*Oryctolagus Cuniculus*) na Austrália e a variável explicativa corresponde à idade (em dias) dos coelhos. Este animal é largamente distribuído na população selvagem da Austrália. Um aspecto interessante para este conjunto de dados, que suporta o uso de erros com distribuição com caudas mais pesadas do que a normal, é a suspeita de outliers sob estimação de mínimos quadrados, ver Wei (1998). Para reanalisar os dados, usamos o modelo sugerido por Galea et al. (2005)

$$\log(Y_i) = \alpha - \frac{\beta}{x_i + \gamma} + \epsilon_i, \quad i = 1, \dots, 71, \quad (3.1)$$

em que $\epsilon_i \sim N(0, \sigma^2)$ são erros mutuamente independentes. Os dados dessa aplicação encontram-se na Tabela A.1, apresentada no Anexo.

Galea et al. (2005) indicam que o modelo de regressão não linear em (3.1) com erros t-Student com 4 graus de liberdade parece ajustar-se melhor aos dados do que o modelo de regressão não linear normal, no sentido de robustez contra observações discrepantes e influentes.

3.1 Estimação e Ajuste

Os valores iniciais foram obtidos utilizando o método de *Resolver um sistema de equações* com as observações (50; 61, 13), (82; 92, 5), (535; 209, 7). As observações foram escolhidas aleatoriamente sendo uma observação no início, uma no meio e a outra no fim do banco de dados.

Consideramos 4 métodos iterativos para a busca da melhor estimativa dos parâmetros: Newton-Raphson (NR), BFGS, L-BFGS-B e `nls()`. BFGS e L-BFGS-B são métodos de otimização disponíveis na função `optim()` do R. O método BFGS é um método de quase-Newton, necessitando apenas que o gradiente da função objetivo esteja disponível em cada iteração. O método L-BFGS-B (Byrd et al. (1995)), permite definir intervalos de limite inferior e/ou superior do espaço paramétrico. Este método utiliza uma modificação de memória limitada do método quase-Newton. O último método utilizado foi a função `nls()` disponível no R.

Notamos pela Tabela 3.1, que três métodos (NR, BFGS e `nls`) apresentaram a mesma qualidade de ajuste. Dessa forma, o método L-BFGS-B foi desconsiderado por apresentar o maior *SQRes*. O método BFGS estima os parâmetros sob a distribuição dos erros enquanto NR e `nls` são obtidos por aproximação linear (sem suposição da distribuição dos erros). Nesta situação, uma opção de escolha dos parâmetros estimados poderia ser pelo método que fornecesse menores erros padrão dos coeficientes. Como não tivemos uma diferença importante dos erros padrão dos coeficientes e como as curvas estimadas foram muito semelhantes (Figura 3.1), optamos pelo método BFGS por trabalhar com a distribuição dos erros.

Na Figura 3.1 apresentamos os dados observados juntamente com o modelo ajustado para cada método de estimação utilizado. Nota-se que o modelo é adequado para ajustar esse conjunto de dados.

Ao calcular as medidas de curvatura de efeito do parâmetro e intrínseca do modelo, obteve-se, respectivamente $\Gamma^T = 0,1313$ e $\Gamma^N = 0,0268$ e o valor crítico para um nível de significância de 0,05 é igual a $2F^{-1/2} = 1,208$ calculado com 3 e 68 graus de liberdade. Portanto, como ambos valores, Γ^T e Γ^N , não foram significativos, pode-se concluir que uma aproximação linear dos resíduos é aceitável.

Tabela 3.1: Estimativas para o modelo (3.1) com os respectivos erros padrão (assintóticos).

Parâmetro	Newton-Raphson	BFGS	L-BFGS-B	nls()
α	5,640 (0,020)	5,640 (0,020)	5,598 (0,021)	5,640 (0,020)
β	130,584 (5,725)	130,627 (5,608)	112,722 (5,255)	130,584 (5,725)
γ	37,603 (2,323)	37,646 (2,277)	28,682 (2,081)	37,603 (2,323)
σ^2	0,004	0,004	0,005	0,004
<i>SQRes</i>	0,269	0,269	0,366	0,269

¹A estimativa do σ^2 é s^2 (Erro Quadrático Médio).

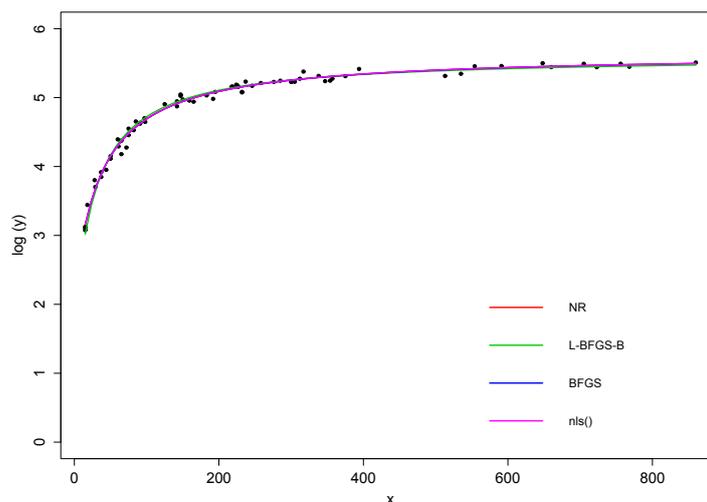


Figura 3.1: Gráfico do modelo normal ajustado aos dados sobre Coelhos Europeus.

A Figura 3.2 exibe os gráficos dos elementos da diagonal da matriz de projeção $\hat{\mathbf{P}}$, dos resíduos ordinários r_i , dos resíduos estudantizados t_i e dos resíduos projetados estudantizados s_i versus a ordem das observações.

Pode-se verificar possíveis pontos influentes através do gráfico dos elementos da diagonal da matriz de projeção. As observações 1, 2, 3 e 4 aparecem como pontos de alavanca, mostrando a dificuldade em prever a resposta para animais mais jovens. As observações 4, 5, 16 e 17 aparecem como pontos discrepantes e influentes.

Os gráficos dos resíduos (Figura 3.2) e o gráfico da distância de Cook (Figura 3.3) indicam que as observações 1, 3, 4, 5, 16 e 17 são pontos influentes.

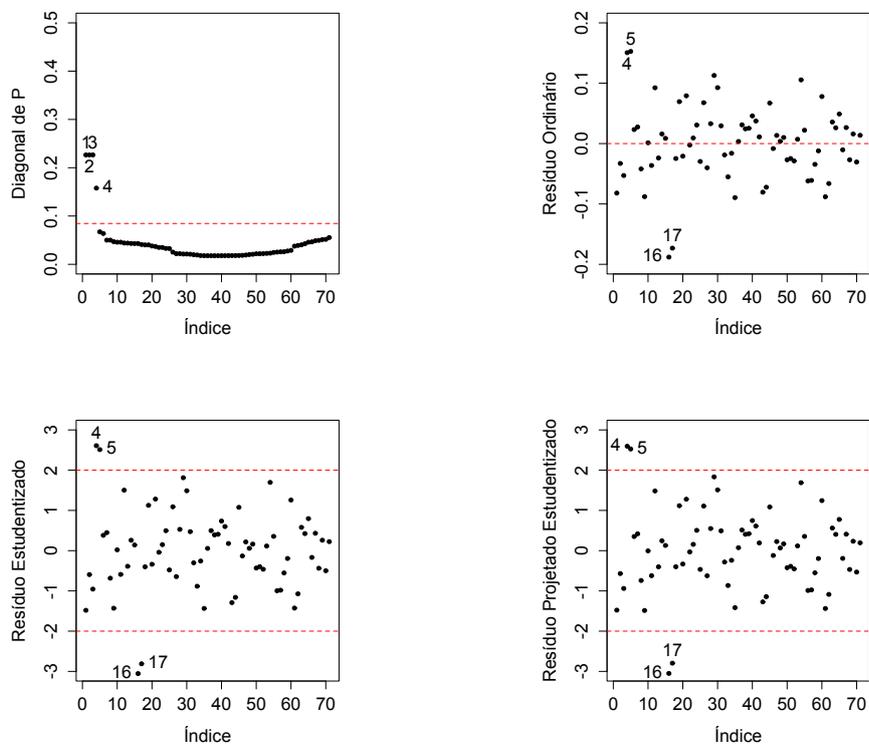


Figura 3.2: Gráfico dos elementos da diagonal da matriz de projeção, dos resíduos ordinários, estudentizados e dos resíduos projetados estudentizados *versus* a ordem das observações.

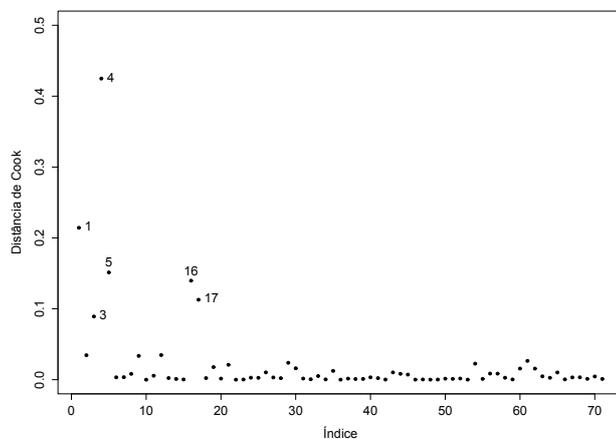


Figura 3.3: Gráfico da Distância de Cook *versus* ordem das observações.

Com o objetivo de analisar o impacto das observações detectadas nas Figuras 3.2 e 3.3 sobre as estimativas dos parâmetros, realizamos uma análise confirmatória. Para isso, consideramos a variação percentual das estimativas definidas por

$$VP = \frac{\hat{\theta}_j - \hat{\theta}_{j(m)}}{\hat{\theta}_j} \times 100\%,$$

em que $\hat{\theta}_j$ é a estimativa de θ com todas as observações e $\hat{\theta}_{j(m)}$ denota a estimativa de θ retirando a m -ésima observação.

Primeiramente, reajustamos o modelo eliminando individualmente cada observação (1, 4, 5, 16 e 17). Em seguida, reajustamos o modelo eliminando conjuntamente as observações detectadas na análise de diagnóstico. As estimativas dos parâmetros dos modelos podem ser vistas na Tabela 3.2. Todos os parâmetros continuam sendo individualmente significativos ao nível de significância 1% mesmo excluindo todas as observações do grupo acima. A maior variação percentual (VP) ocorreu em γ quando excluímos a observação 4.

Como não houve mudanças inferenciais em nenhum dos casos optou-se pelo modelo inicial.

Tabela 3.2: Variação das estimativas do modelo não linear normal ao excluir as observações citadas na análise de diagnóstico. Os erros-padrão assintóticos estão entre parênteses.

Obs	Estimativas			VP(%)		
	α	β	γ	α	β	γ
1	5,645 (0,020)	132,811 (6,037)	38,912 (2,545)	-0,089	-1,705	-3,481
4	5,635 (0,019)	128,272 (5,346)	36,076 (2,192)	0,089	1,771	4,061
5	5,642 (0,019)	130,998 (5,469)	37,384 (2,204)	-0,035	-0,317	0,582
16	5,633 (0,019)	127,669 (5,303)	36,553 (2,155)	0,124	2,232	2,792
17	5,634 (0,019)	127,945 (5,374)	36,626 (2,183)	0,106	2,021	2,598
Todas	5,648 (0,020)	133,731 (6,000)	38,709 (2,504)	-0,142	-2,410	-2,941

3.2 Modelos Alternativos

Nesta seção são apresentados 3 modelos distintos ao proposto na literatura (3.1) para o conjunto de dados *Coelhos Europeus*.

Existem vários métodos de se ajustar curvas aos dados. A técnica de regressão não paramétrica, por exemplo, permite maior flexibilidade na possível forma de f , pois com esta técnica, assume-se que a curva de regressão pertence a uma coleção infinito dimensional de funções. Utilizaremos a técnica de *Smoothing Splines* para verificar a validade do modelo paramétrico proposto em (3.1).

Para fins de comparação, ajustaremos aos dados os modelos Logístico e de Gompertz, pertencentes a classe de modelos de crescimento, para verificar se os comportamentos das curvas se assemelham ao modelo (3.1).

3.2.1 Smoothing Splines

Considere o modelo de regressão dado por

$$y_i = g(x_i) + \epsilon_i, \quad i = 1, \dots, n,$$

onde ϵ 's são variáveis não correlacionadas com média zero e independentes de Y_i .

O método de regressão que resolve o problema de mínimos quadrados penalizados é denominado *Smoothing Splines*. Que define a estimativa é a solução do problema de otimização dado por

$$L_\lambda(g) = \|\mathbf{y} - \mathbf{g}\|^2 + \lambda \int (g'')^2,$$

onde $\mathbf{y} = (y_1, \dots, y_n)'$ e $\mathbf{g} = (g(x_1), \dots, g(x_n))'$.

O parâmetro λ é de grande importância neste critério de otimização, pois é ele que controla o balanço entre suavidade e bondade de ajuste do estimador. Para valores grandes de λ , a suavidade tem maior influência no estimador, enquanto que para valores pequenos de λ , a bondade de ajuste predomina no critério de soma de quadrados penalizada, daí λ ser chamado de parâmetro de suavização.

A Figura 3.4 mostra estimativas usando o método de *Smoothing Splines* com parâmetro de suavização pequeno, predominando a bondade do ajuste. O ajuste do modelo não linear foi obtido da mesma maneira que na Seção 3.1 e o intervalo de confiança para resposta esperada foi obtido utilizando (2.9).

Mais detalhes sobre *Smoothing Splines*, ver Seber e Wild (2003), Dias (2001).

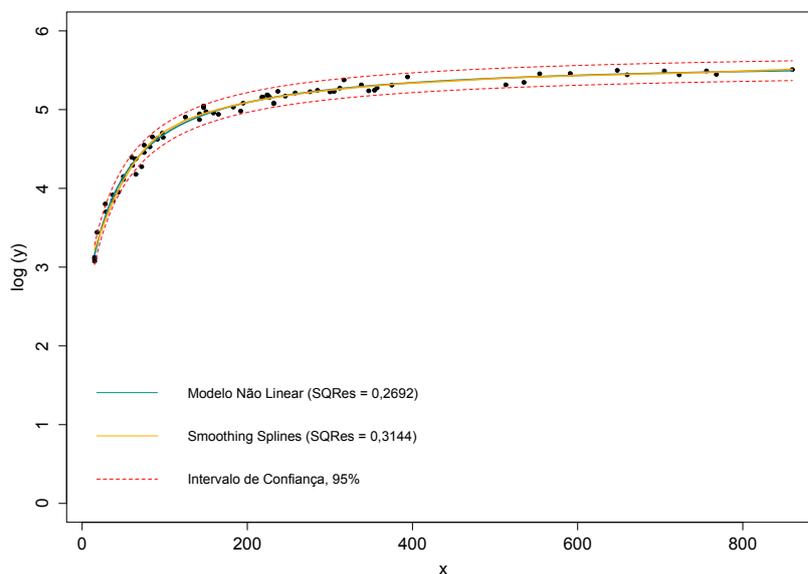


Figura 3.4: Modelo ajustado via Smoothing Splines comparado com o modelo ajustado como em (3.1). O Intervalo de Confiança foi obtido como em (2.9).

3.2.2 Modelo Logístico e de Gompertz

Para os dados do Coelho Europeu ajustamos os modelos Logístico e Gompertz, apresentados na Tabela 2.1. Como a variabilidade da resposta cresce quando a idade do animal cresce, consideramos $\log(y)$ para y .

Os valores iniciais para o modelo Logístico foram obtidos fazendo-se uma regressão linear de $z_0 = \log((\alpha_0/y) - 1) = \beta - \gamma x$ em relação a x , sendo $\alpha_0 = y_{m\acute{a}x}$, $\beta_0 = coef_1$ e $\gamma_0 = -coef_2$.

Os valores iniciais para o modelo Gompertz foram obtidos fazendo-se uma regressão linear de $z_0 = \log(-\log(y/\alpha_0)) = \beta - \gamma x$ em relação a x , sendo $\alpha_0 = y_{m\acute{a}x}$, $\beta_0 = coef_1$ e $\gamma_0 = -coef_2$.

O valor de $coef_1$ representa a constante de regressão (intercepto da reta com o eixo vertical), e o valor de $coef_2$ representa o coeficiente de regressão (declive da reta).

Na Tabela 3.3 estão as estimativas dos parâmetros dos modelos. A representação gráfica das curvas ajustadas e dos valores observados são mostradas na Figura 3.5, notamos que a curva ajustada do modelo de Gompertz sobrepõe a curva do modelo Logístico.

Pelos *SQRes* dos modelos na Tabela 3.3, podemos perceber que o modelo Gompertz ajusta melhor aos dados do que o modelo Logístico. Entretanto, Gompertz não é melhor que o modelo em (3.1).

Tabela 3.3: Estimativas para os modelos Logístico e Gompertz com os respectivos erros padrão (assintóticos).

Parâmetro	Logístico	Gompertz
α	5,309 (0,021)	5,323 (0,020)
β	-0,271 (0,044)	-0,52 (0,033)
γ	0,018 (0,001)	0,016 (0,001)
σ^2	0,013	0,011
<i>SQRes</i>	0,877	0,757

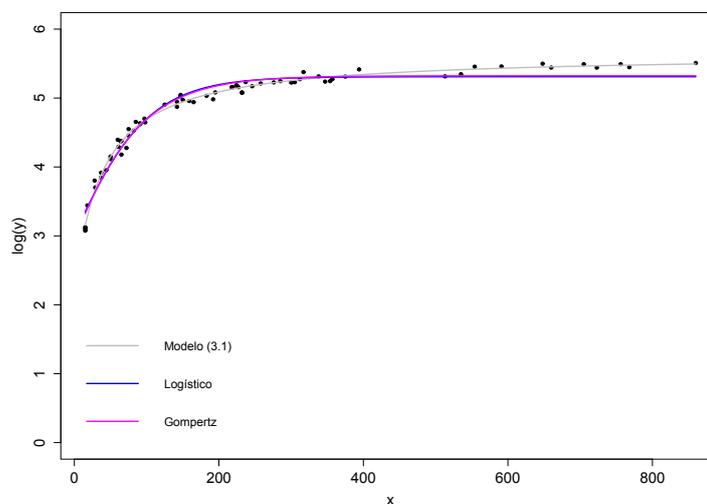


Figura 3.5: Gráfico de dispersão das observações com as curvas ajustadas dos modelos Logístico, Gompertz e (3.1).

A Figura 3.6 exibe os gráficos dos elementos da diagonal da matriz de projeção $\hat{\mathbf{P}}$ e dos resíduos estudentizados t_i versus a ordem das observações do modelo Logístico. A Figura 3.7 exibe os gráficos dos elementos da diagonal da matriz de projeção $\hat{\mathbf{P}}$ e dos resíduos estudentizados t_i versus a ordem das observações do modelo Gompertz.

Notamos que nas Figuras 3.6 e 3.7, os pontos 1, 2, 3 e 4 aparecem como pontos de

alavanca e influentes. Nos gráfico dos resíduos estudentizados, em ambas as figuras, os resíduos não estão aleatórios, indicando uma possível heterocedasticidade de variâncias.

Assim, certificamos que o modelo proposto em (3.1) se ajusta melhor aos dados do que os modelos Logístico e Gompertz.

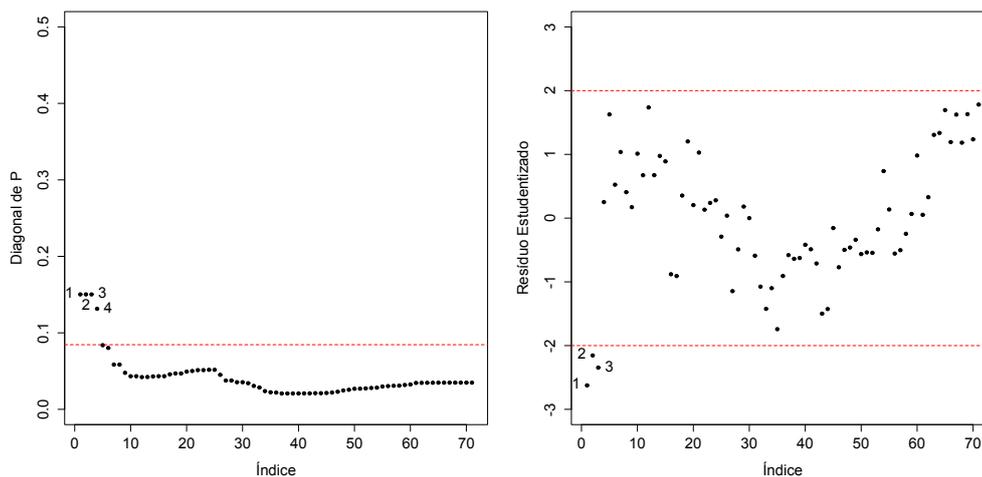


Figura 3.6: Gráfico dos elementos da diagonal da matriz de projeção e dos resíduos estudentizados *versus* a ordem das observações do modelo Logístico.

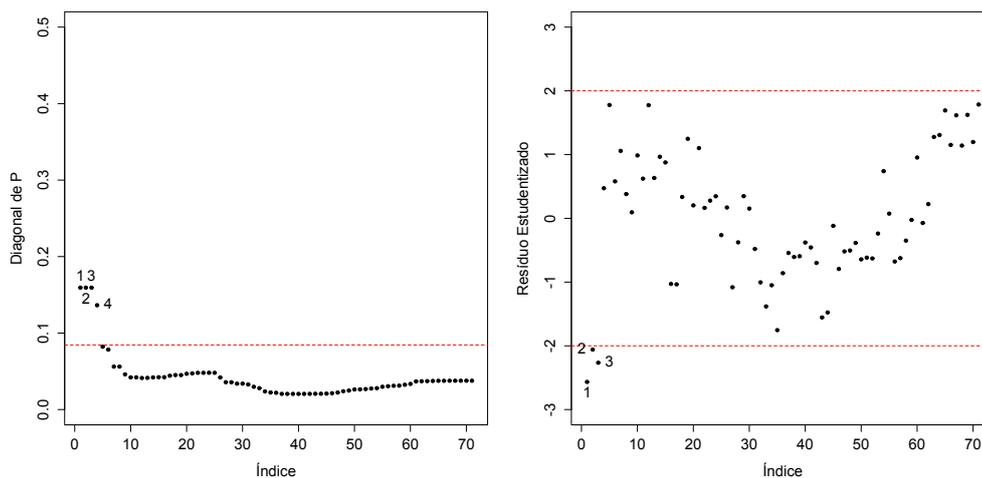


Figura 3.7: Gráfico dos elementos da diagonal da matriz de projeção e dos resíduos estudentizados *versus* a ordem das observações do modelo Gompertz.

4 Conclusão

Neste trabalho, o principal objetivo foi apresentar uma revisão dos modelos normais não lineares de uma maneira geral. Primeiramente introduziu-se a definição de modelos não lineares, suas classes, os métodos de estimação dos parâmetros e alguns resultados assintóticos. Logo após, foram apresentados as medidas de não linearidade e técnicas de diagnóstico para este modelo. Na aplicação, foram apresentados alguns resultados para o conjunto de dados reais, comprovando que o modelos não lineares são deduzidos a partir de suposições teóricas.

Uma dificuldade encontrada durante o desenvolvimento da aplicação foi na escolha dos valores iniciais para o modelo proposto da literatura. Tentativas foram feitas antes de um estudo mais detalhado e alguns desses valores tornavam a matriz $\mathbf{F}(\boldsymbol{\theta})\mathbf{F}'(\boldsymbol{\theta})$ singular ou levavam para um mínimo local e não um mínimo global. Para o conjunto de dados analisados (Coelhos Europeus), o modelo proposto na literatura teve dificuldade em ajustar a resposta para animais mais jovens. Além disso, ajustamos 3 modelos alternativos aos dados, não resultando melhores ajustes.

Em resumo, a aplicação ao conjunto de dados dos Coelhos Europeus afirma que os modelos não lineares são utilizados com o objetivo de explicar ou descrever o fenômeno sob estudo. Como trabalho futuro, podemos ajustar modelos não lineares com erros heterocedásticos para o conjunto de dados dos Coelhos Europeus. Como a maioria dos dados de modelos não lineares envolve tempo, pode-se incorporar estruturas de erros autoregressivos (XIE et al., 2009).

Referências

- ATKINS, G. L. *Multicompartment models for biological systems*. [S.l.]: Methuen London, 1969.
- BATES, D. M.; WATTS, D. G. Relative curvature measures of nonlinearity. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 1–25, 1980.
- BATES, D. M.; WATTS, D. G. *Nonlinear regression: iterative estimation and linear approximations*. [S.l.]: Wiley Online Library, 1988.
- BEALE, E. Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 41–88, 1960.
- BLEASDALE, J.; NELDER, J. *Plant population and crop yield*. Nature Publishing Group, 1960.
- BOX, M. Bias in nonlinear estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 171–201, 1971.
- BYRD, R. H.; LU, P.; NOCEDAL, J.; ZHU, C. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, SIAM, v. 16, n. 5, p. 1190–1208, 1995.
- CANCHO, V.; LACHOS, V.; ORTEGA, E. A nonlinear regression model with skew-normal errors. *Statistical Papers*, Springer-Verlag, v. 51, n. 3, p. 547–558, 2010. ISSN 0932-5026.
- COOK, R.; TSAI, C.; WEI, B. Bias in nonlinear regression. *Biometrika*, v. 73, n. 3, p. 615–623, 1986.
- COOK, R.; TSAI, C.-L. Residuals in nonlinear regression. *Biometrika*, Biometrika Trust, v. 72, n. 1, p. 23–29, 1985.
- COOK, R. D. Detection of influential observation in linear regression. *Technometrics*, JSTOR, p. 15–18, 1977.
- COOK, R. D.; WEISBERG, S. *Residuals and influence in regression*. [S.l.]: Chapman and Hall New York, 1982.
- CORDEIRO, G. M.; FERRARI, S. L.; URIBE-OPAZO, M. A.; VASCONCELLOS, K. L. Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics & probability letters*, Elsevier, v. 46, n. 4, p. 317–328, 2000.
- COX, D. R.; SNELL, E. J. A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, JSTOR, p. 248–275, 1968.

- CYSNEIROS, F. J. A.; VANEGAS, L. H. Residuals and their statistical properties in symmetrical nonlinear models. *Statistics & Probability Letters*, Elsevier, v. 78, n. 18, p. 3269–3273, 2008.
- DIAS, R. A review of non-parametric curve estimation methods with application to econometrics. Citeseer, 2001.
- DRAPER, N. R.; SMITH, H. *Applied Regression Analysis (Wiley Series in Probability and Statistics)*. Third. [S.l.]: Wiley-Interscience, 1998. Hardcover. ISBN 0471170828.
- DUDZINSKI, M.; MYKYTOWYCZ, R. The eye lens as an indicator of age in the wild rabbit in australia. *Wildlife Research*, CSIRO, v. 6, n. 2, p. 156–159, 1961.
- GALEA, M.; PAULA, G. A.; CYSNEIROS, F. J. A. On diagnostics in symmetrical nonlinear models. *Statistics & Probability Letters*, v. 73, n. 4, p. 459 – 467, 2005.
- GALLANT, A. R. *Nonlinear statistical models*. [S.l.]: John Wiley & Sons, 1987.
- GILLIS, P.; RATKOWSKY, D. The behaviour of estimators of the parameters of various yield-density relationships. *Biometrics*, JSTOR, p. 191–198, 1978.
- GUTTMAN, I.; MEETER, D. A. On beale’s measures of non-linearity. *Technometrics*, v. 7, n. 4, p. 623–637, 1965.
- HOLLIDAY, R. et al. Plant population and crop yield. *Nature*, Lond., v. 186, n. 4718, p. 22–4, 1960.
- JACQUEZ, J. A. Compartmental analysis in biology and medicine. Elsevier: New York, 1972.
- MONAHAN, J. F. *Numerical methods of statistics*. [S.l.]: Cambridge University Press, 2011.
- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, JSTOR, p. 370–384, 1972.
- PREGIBON, D. Logistic regression diagnostics. *The Annals of Statistics*, JSTOR, p. 705–724, 1981.
- PRUDENTE, A. A. *Modelos não-lineares de regressão: Alguns aspectos de teoria assintótica*. 2009.
- RATKOWSKY, D. A. Nonlinear regression modeling. Dekker, New York., 1983.
- RESCIGNO, A.; LAMBRECHT, R. M.; DUNCAN, C. C. Mathematical methods in the formulation of pharmacokinetic models. In: *Tracer Kinetics and Physiologic Modeling*. [S.l.]: Springer, 1983. p. 59–119.
- RICKER, W. E. Stock and recruitment. *Journal of the Fisheries Board of Canada*, NRC Research Press, v. 11, n. 5, p. 559–623, 1954.
- SEBER, G.; WILD, C. *Nonlinear Regression*. [S.l.]: Wiley, 2003. (Wiley Series in Probability and Statistics). ISBN 9780471471356.

VANEGAS, L. H.; CYSNEIROS, F. J. A. Assessment of diagnostic procedures in symmetrical nonlinear regression models. *Computational Statistics & Data Analysis*, Elsevier, v. 54, n. 4, p. 1002–1016, 2010.

WAGNER, J. Use of computers in pharmacokinetics. *Clinical pharmacology and therapeutics*, v. 8, n. 1, p. 201, 1967.

WAGNER, J. G. Fundamentals of clinical pharmacokinetics. Drug Intelligence Publications Hamilton, Illinois, 1975.

WEI, B.-C. *Exponential family nonlinear models*. [S.l.]: Springer Singapore, 1998.

XIE, F.-C.; LIN, J.-G.; WEI, B.-C. Diagnostics for skew-normal nonlinear regression models with ar (1) errors. *Computational Statistics & Data Analysis*, Elsevier, v. 53, n. 12, p. 4403–4416, 2009.

APÊNDICE A

Programação no *software* R

```

### ===== ###
### FUNÇÕES NECESSÁRIAS PARA A ESTIMAÇÃO DOS PARÂMETROS ###
### ===== ###

## Newton-Raphson:

# Derivadas do SQRes:
derivadas <- deriv3(~(y-(a-(b/(x+c))))^2, c("a","b","c"), function(y,x,a,b,c){})

# Gradiente e Hessiana da função SQRes:
GH <- function(y,x,a,b,c)
{
  m <- derivadas(y,x,a,b,c)
  g <- attr(m,"gradient")
  G <- apply(g,2,sum)
  h <- attr(m, "hessian")
  h <- matrix(h, ncol=3*3)
  H <- apply(h, 2, sum)
  H <- matrix(H, 3, 3)
  return(list(G,H))
}

# Algoritmo Newton-Raphson:
thetaNR <- function(y,x,theta0,maxIter,tolerancia)
{
  criterio <- 1
  count <- 0

```

```

while(criterio > tolerancia)
{
  count<-count+1
  gra.hess <- GH(y,x,theta0[1], theta0[2], theta0[3])
  gradiente <- gra.hess[[1]]
  hessiana <- gra.hess[[2]]
  inv.hessiana <- solve(hessiana)
  theta1 <- theta0 - inv.hessiana%*%gradiente
  #criterio <- sum(((theta1 - theta0)/(theta0))^2)
  criterio <- sum((theta0-theta1)^2)
  theta0 <- theta1
  if (count==maxIter)
  {
    break
  }
}
return(list(beta = theta1, iter=count))
}

## Função da log-verossimilhança:
logvero <- function(y,x,theta)
{
  n <- length(y)
  b1 <- theta[1]
  b2 <- theta[2]
  b3 <- theta[3]
  sigma2 <- theta[4]
  vero <- -(n/2)*log(2*pi) - (n/2)*log(sigma2) -
          (1/(2*sigma2))*sum((y - (b1 - b2/(x+b3)))^2)
  return(-vero)
}

#####
## Aplicação ##
#####

# Lendo os dados:
coelhos <- read.table("coelhos.txt", header=TRUE, dec=",")

```

```
# Variável explicativa: Idade dos coelhos (em dias)
x <- coelhos[,1]

# Variável resposta: Peso lentes dos olhos (em mg)
y <- coelhos[,2]
y <- log(y)
n <- length(y)
p <- 3

# Gradiente do modelo proposto:
fun.derivada <- deriv3(~ a-(b/(x + c)), c("a","b","c"), function(x,a,b,c){})

# Valor inicial:
x1 <- x[11]
x2 <- x[20]
x3 <- x[62]
y1 <- y[11]
y2 <- y[20]
y3 <- y[62]

gamma <- (y1*(x1*x2 - x1*x3) + y2*(x2*x3-x1*x2) +
          y3*(x1*x3-x2*x3))/(y1*(x3-x2)+y2*(x1-x3)+y3*(x2-x1))
beta <- ((y2-y1)*(x2+gamma)*(x1+gamma))/(x2-x1)
alfa <- y1 + (beta/(x1+gamma))

theta0 <- c(alfa, beta, gamma)

#####
## Estimação ##
#####

# Newton-Raphson:
tolerancia <- 1e-20
maxIter <- 200

NR <- thetaNR(y,x,theta0,maxIter,tolerancia)
theta_nr <- round(NR$beta,3)
```

```

sqres_nr <- sum((y - (theta_nr[1] - (theta_nr[2]/(x + theta_nr[3]))))^2)
sigma2_nr <- sqres_nr/(n-p)
deriv.f <- fun.derivada(x,theta_nr[1],theta_nr[2],theta_nr[3])
grad <- attr(deriv.f,"gradient")
C_nr <- solve(t(grad)%*%grad)
ep <- sqrt(diag(C_nr)*sigma2_nr)

# optim ()
sqres0 <- sum((y- (theta0[1] - theta0[2]/(x + theta0[3]))))^2)
sigma20 <- sqres0/(n)
theta0 <- c(theta0,sigma20)

# BFGS:
thetaM <- optim(theta0,logvero, method = "BFGS",control=list(maxit=20000),
               y=y,x=x,hessian=TRUE)

theta_bf <- round(thetaM$par,3)
sqres_bf <- sum((y- (theta_bf[1] - theta_bf[2]/(x + theta_bf[3]))))^2)
sigma2_bf <- sqres_bf/n
deriv.f <- fun.derivada(x,theta_bf[1],theta_bf[2],theta_bf[3])
grad <- attr(deriv.f,"gradient")
C_bf <- solve(t(grad)%*%grad)
ep <- sqrt(diag(C_bf)*sigma2_bf)

# L-BFGS-B:
L=c(0,0,0,0.000000000001)
U=c(Inf,Inf,Inf,Inf)
thetaF <- optim(theta0,logvero,method="L-BFGS-B",lower=L,upper=U,y=y,x=x)
theta_lb <- round(thetaF$par,3)
sqres_lb <- sum((y- (theta_lb[1] - theta_lb[2]/(x + theta_lb[3]))))^2)
sigma2_lb <- sqres_lb/n
deriv.f <- fun.derivada(x,theta_lb[1],theta_lb[2],theta_lb[3])
grad <- attr(deriv.f,"gradient")
C_lb <- solve(t(grad)%*%grad)
ep <- sqrt(diag(C_lb)*sigma2_lb)

# nls ():
n0 <- nls(y~fun.derivada(x,A,B,C), start=c(A=theta0[1],B=theta0[2],C=theta0[3]),
        trace=TRUE)

```

```

theta_nls <- round(coef(n0),3)
sqres_nls <- sum((y- (theta_nls[1] - theta_nls[2]/(x + theta_nls[3]))))^2)
sigma2_nls <- (summary(n0)$sigma)^2
ep <- summary(n0)$coef[,2]

# Gráfico:
plot(y~x,pch=20, ylim=c(0,6), ylab="log (y)", ,mgp=c(2.6, 1, 0))
with(list(theta=theta_nr), curve(theta[1] - (theta[2]/(x + theta[3])),
                                col=2,add=TRUE,lty=1,lwd=2))
with(list(theta=theta_lb), curve(theta[1] - (theta[2]/(x + theta[3])),
                                col=3,add=TRUE,lty=1,lwd=2))
with(list(theta=theta_bf), curve(theta[1] - (theta[2]/(x + theta[3])),
                                col=4,add=TRUE,lty=1,lwd=2))
with(list(theta=theta_nls), curve(theta[1] - (theta[2]/(x + theta[3])),
                                col=6,add=TRUE,lty=1,lwd=2))
legend("bottomright", c("NR","L-BFGS-B","BFGS","nls()"),lty=1,col=c(2,3,4,6),
      cex=0.8,lwd=2, bty="n")

## Métodos BFGS escolhido :
theta <- round(thetaM$par,3)
sqres <- sum((y- (theta[1] - theta[2]/(x + theta[3]))))^2)
sigma2 <- sqres/n
sigma <- sqrt(sigma)
deriv.f <- fun.derivada(x,theta[1],theta[2],theta[3])
grad <- attr(deriv.f,"gradient") #gradiente
C <- solve(t(grad)%*%grad)
sd.theta <- sqrt(diag(C)*sigma2)
H <- attr(deriv.f, "hessian") # hessiana

FF <- t(grad)%*%grad

#####
## Medidas de não-linearidade ##
#####

# Medida de Vício de Box
b <- - (sigma2/2)*sapply(1:n,function(x){sum(diag(solve(FF)%*%H[x, ,]))})
B <- as.vector(solve(FF)%*%t(grad)%*%b)

```

```

B.theta <- 100*B/theta

# Curvatura de Battes e Watts
rms.curv(n0) # Curvaturas
f <- qf(0.95,p,(n-p))
valor <- 2*(f)^(-1/2)

#####
## Técnicas de Diagnóstico ##
#####

# Matriz de Projeção
P <- grad%%solve(FF)%*%t(grad)
hii <- diag(P)

#Resíduo ordinário
y_hat <- c(deriv.f)
ri <- y - y_hat

#Resíduo Estudentizado
ti <- ri/(sigma*sqrt(1-hii))

#Resíduo Projetado
h <- matrix(H,n,p*p)
Th <- cbind(h[,9])          #vetores não-nulos
I <- diag(n)
S <- (I-P)%*%Th
St <- t(S)
StS <- t(S)%*%S
P1 <- S%%solve(StS)%*%St
V <- cbind(grad,S)
Vt <- t(V)
VtV <- t(V)%*%V
P2 <- V%%solve(VtV)%*%Vt
trP2 <- sum(diag(I-P2))
sigma2.p <- (t(ri)%*%(I-P2)%*%ri)/(trP2)
sigma.p <- sqrt(sigma2.p)
si <- ((I-P2)%*%ri)/(sigma.p*diag(sqrt(I-P2)))

```

```
# Gráficos:
par(mfrow=c(2,2))
plot(hii,xlab="Índice",ylab="Diagonal de P", pch=20,main="",
      ylim=c(0,0.5))
abline(h=(2*p)/n,lty=2,col="red")
identify(hii)
plot(ri, xlab="Índice", ylab="Resíduo Ordinário", pch=20,main="",
      ylim=c(-0.20,0.20))
abline(h=0,lty=2,col="red")
identify(ri)
plot(ti,xlab="Índice", ylab="Resíduo Estudentizado", pch=20,main="",
      ylim=c(-3,3))
abline(h=-2,lty=2,col="red")
abline(h=2,lty=2,col="red")
identify(ti)
plot(si,xlab="Índice", ylab="Resíduo Projetado Estudentizado", pch=20,main="",
      ylim=c(-3,3))
abline(h=-2,lty=2,col="red")
abline(h=2,lty=2,col="red")
identify(si)

# Distância de Cook:
Di <- ((ti^2)/p)*(hii/(1-hii))
plot(Di,xlab="Índice", ylab="Distância de Cook", pch=20,main="", ylim=c(0,0.5))
identify(Di)

#####
## Variação Percentual ##
#####

# eliminação:
# 1
x1 <- x[-1]
y1 <- y[-1]
n1 <- nls(y1~fun.derivada(x1,A,B,C), start=c(A=theta0[1],B=theta0[2],
      C=theta0[3]), trace=TRUE)

theta1 <- round(coef(n1),3)
```

```
VP1 <- ((theta - theta1)/theta)*100

# 4
x4 <- x[-4]
y4 <- y[-4]
n4 <- nls(y4~fun.derivada(x4,A,B,C), start=c(A=theta0[1],B=theta0[2],
                                           C=theta0[3]), trace=TRUE)

theta4 <- round(coef(n4),3)
summary(n4)
VP4 <- ((theta - theta4)/theta)*100

# 5
x5 <- x[-5]
y5 <- y[-5]
n5 <- nls(y5~fun.derivada(x5,A,B,C), start=c(A=theta0[1],B=theta0[2],
                                           C=theta0[3]), trace=TRUE)

theta5 <- round(coef(n5),3)
summary(n5)
VP5 <- ((theta - theta5)/theta)*100

# 16
x16 <- x[-16]
y16 <- y[-16]
n16 <- nls(y16~fun.derivada(x16,A,B,C), start=c(A=theta0[1],B=theta0[2],
                                               C=theta0[3]), trace=TRUE)

theta16 <- round(coef(n16),3)
summary(n16)
VP16 <- ((theta - theta16)/theta)*100

# 17
x17 <- x[-17]
y17 <- y[-17]
n17 <- nls(y17~fun.derivada(x17,A,B,C), start=c(A=theta0[1],B=theta0[2],
                                               C=theta0[3]), trace=TRUE)

theta17 <- round(coef(n17),3)
summary(n17)
VP17 <- ((theta - theta17)/theta)*100
```

```

# 1,4,5,16 e17
xt <- x[-1]
xt <- xt[-4:-5]
xt <- xt[-16:-17]
yt <- y[-1]
yt <- yt[-4:-5]
yt <- yt[-16:-17]
nt <- nls(yt~fun.derivada(xt,A,B,C), start=c(A=theta0[1],B=theta0[2],
                                           C=theta0[3]), trace=TRUE)

thetat <- round(coef(nt),3)
summary(nt)
VPt <- ((theta - thetat)/theta)*100

#####
## Smoothing Splines ##
#####

fit.spline <- smooth.spline(x,y, nknots=4)
ynest <- predict(fit.spline,x)$y
sqres.spline <- sum((y - ynest)^2)

# Gráfico:
xnovo <- seq(min(x),max(x),by=0.1)
ypred <- predict(n0,newdata=list(x=xnovo))
up <- ypred + qt(0.975,summary(n0)$df[2])*summary(n0)$sigma
down <- ypred - qt(0.975,summary(n0)$df[2])*summary(n0)$sigma
plot(y~x,pch=20,ylim=c(0,6), ylab="log (y)", ,mgp=c(2.6, 1, 0))
with(list(theta=theta_nls), curve(theta[1] - (theta[2]/(x + theta[3])),
                                  col="orange",add=TRUE,lwd=2))

lines(fit.spline, col="blue", lwd=2)
lines(up~xnovo, col="red", lty=2, lwd=1)
lines(down~xnovo, col="red", lty=2, lwd=1)
legend("bottomleft", c("Modelo Não Linear (SQRes = 0,2692)",
                       "Smoothing Splines (SQRes = 0,3144)", "Intervalo de Confiança, 95%"),
      lty=c(1,1,2),col=c("orange", "blue", "red"),cex=0.8,lwd=1, bty="n")

```

```
#####  
## Modelo Logístico e Gompertz ##  
#####  
  
# Logístico:  
  
alpha0 <- max(y) + 1  
z0 <- log((alpha0/y) - 1)  
reg <- lm(z0~x)  
beta0 <- coef(reg)[1]  
gamma0 <- -coef(reg)[2]  
  
log.der <- deriv3(~ a/(1+exp(b - c*x)), c("a","b","c"), function(x,a,b,c){})  
reg.nls <- nls(y ~ log.der(x,a,b,c), start=list(a=alpha0, b=beta0, c=gamma0))  
ajuste <- summary(reg.nls)  
theta.log <- ajuste$coefficients[,1]  
sd <- ajuste$parameters[,2]  
  
## Diagnósticos  
  
n <- length(y)  
p <- length(theta.log)  
sigma <- ajuste$sigma  
sigma2 <- (sigma)^2  
der.log <- log.der(x,theta.log[1],theta.log[2],theta.log[3])  
grad.log <- attr(der.log,"gradient") # gradiente  
C <- solve(t(grad.log)%*%grad.log)  
sd.theta <- sqrt(diag(C)*sigma2)  
H.log <- attr(der.log, "hessian") # hessiana  
FtF <- t(grad.log)%*%grad.log  
  
## Matriz de Projeção  
P <- grad.log%*%solve(FtF)%*%t(grad.log)  
hii <- diag(P)  
  
## Resíduo ordinário  
y_hat <- c(der.log)  
ri <- y - y_hat
```

```
## Resíduo Estudentizado
ti <- ri/(sigma*sqrt(1-hii))

# Gráficos
par(mfrow=c(1,2))
plot(hii,xlab="Índice",ylab="Diagonal de P", pch=20,main="", ylim=c(0,0.5))
abline(h=(2*p)/n,lty=2,col="red")
identify(hii)
plot(ti,xlab="Índice", ylab="Resíduo Estudentizado", pch=20,main="", ylim=c(-3,3))
abline(h=-2,lty=2,col="red")
abline(h=2,lty=2,col="red")
identify(ti)

# Gompertz

alpha0 <- max(y) + 1
z0 <- log(-log(y/alpha0))
fit <- lm(z0~x)
beta0 <- coef(fit)[1]
gamma0 <- - coef(fit)[2]

gomp.der <- deriv3(~ a*exp(-exp(b - c*x)), c("a","b","c"), function(x,a,b,c){})
fit.nls <- nls(y ~ gomp.der(x,a,b,c), start=list(a=alpha0, b=beta0, c=gamma0))
ajuste <- summary(fit.nls)
theta.g <- ajuste$coefficients[,1]
sd <- ajuste$parameters[,2]

## Diagnósticos

n <- length(y)
p <- length(theta.g)
sigma <- ajuste$sigma
sigma2 <- (sigma)^2
der.gomp <- gomp.der(x,theta.g[1],theta.g[2],theta.g[3])
grad.gomp <- attr(der.gomp,"gradient") # gradiente
C <- solve(t(grad.gomp)%*%grad.gomp)
sd.theta <- sqrt(diag(C)*sigma2)
```

```
H.gomp <- attr(der.gomp, "hessian") # hessiana
FtF <- t(grad.gomp)%*%grad.gomp

## Matriz de Projeção
P <- grad.gomp%*%solve(FtF)%*%t(grad.gomp)
hii <- diag(P)

## Resíduo ordinário
y_hat <- c(der.gomp)
ri <- y - y_hat

## Resíduo Estudentizado
ti <- ri/(sigma*sqrt(1-hii))

# Gráficos
par(mfrow=c(1,2))
plot(hii,xlab="Índice",ylab="Diagonal de P", pch=20,main="", ylim=c(0,0.5))
abline(h=(2*p)/n,lty=2,col="red")
identify(hii)
plot(ti,xlab="Índice",ylab="Resíduo Estudentizado",pch=20,main="",ylim=c(-3,3))
abline(h=-2,lty=2,col="red")
abline(h=2,lty=2,col="red")
identify(ti)

# Gráficos (dados com curvas ajustadas):
plot(y~x,pch=20, ylab="log(y)", ,mgp=c(2.6, 1, 0), ylim=c(0,6))
with(list(theta=theta_nls), curve(theta[1] - (theta[2]/(x + theta[3])),col=8,
                                add=TRUE,lwd=2))
with(list(theta=theta.log), curve(theta[1]/(1 + exp(theta[2] - (theta[3]*x))),
                                col=4,add=TRUE,lty=1,lwd=2))
with(list(theta=theta.g), curve(theta[1]*exp(-exp(theta[2] - theta[3]*x)),
                                col=6,add=TRUE,lty=1,lwd=2))
legend("bottomleft", c("Modelo (3.1)","Logístico", "Gompertz"),lty=1,
      col=c(8,4,6),cex=0.8,lwd=2, bty="n")
```

ANEXO A

Banco de Dados

Tabela A.1: Pesos das lentes dos olhos de coelhos europeus, y (em mg) e idade, x (em dias) numa amostra de 71 observações

x	y	x	y	x	y	x	y	x	y
15	21,66	65	65,31	150	144,5	246	176,13	513	203,3
15	22,75	72	71,9	159	142,15	258	183,4	535	209,7
15	22,3	75	86,1	165	139,81	276	186,26	554	233,9
18	31,25	75	94,6	183	153,22	285	189,66	591	234,7
28	44,79	82	92,5	192	145,72	300	186,09	648	244,3
29	40,55	85	105	195	161,1	301	186,7	660	231
37	50,25	91	101,7	218	174,18	305	186,8	705	242,4
37	46,88	91	102,9	218	173,03	312	195,1	723	230,77
44	52,03	97	110	219	173,54	317	216,41	756	242,57
50	63,47	98	104,3	224	178,86	338	203,23	768	232,12
50	61,13	125	134,9	225	177,68	347	188,38	860	246,7
60	81	142	130,68	227	173,73	354	189,7		
61	73,09	142	140,58	232	159,98	357	195,31		
64	79,09	147	155,3	232	161,29	375	202,63		
65	79,51	147	152,2	237	187,07	394	224,82		