

ALESSANDRO HENRIQUE DA SILVA SANTOS

**MODELOS NÃO-LINEARES DA FAMÍLIA
EXPONENCIAL**

RECIFE-PE - FEV/2009



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

MODELOS NÃO-LINEARES DA FAMÍLIA EXPONENCIAL

Dissertação apresentada ao Programa de Pós-Graduação em Biometria e Estatística Aplicada como exigência parcial à obtenção do título de Mestre.

Área de Concentração: Modelagem Estatística e Computacional

Orientador: Eufrázio de S. Santos

Co-orientador: Gauss Moutinho Cordeiro

RECIFE-PE - FEV/2009

FICHA CATALOGRAFICA

S237m Santos, Alessandro Henrique da Silva
Modelos não-lineares da família exponencial / Alessandro
Henrique da Silva Santos. -- 2009.
72 f. : il.
Orientador : Eufrázio de S. Santos
Dissertação (Mestrado em Biometria e Estatística Aplica-
da) - Universidade Federal Rural de Pernambuco. Departa-
mento de Estatística e Informática.
Inclui bibliografia.

CDD 574.0182

1. Família exponencial
 2. Função de verossimilhança
 3. Modelos não-lineares
- I. Santos, Eufrázio de S.
 - II. Título

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA

MODELOS NÃO-LINEARES DA FAMÍLIA EXPONENCIAL

ALESSANDRO HENRIQUE DA SILVA SANTOS

Dissertação julgada adequada para obtenção do título de Mestre em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 27/02/2009 pela Comissão Examinadora.

Orientador:

Prof. Dr. Eufrázio de S. Santos
Universidade Federal Rural de Pernambuco

Banca Examinadora:

Profa. Dra. Laélia Pumilla Botelho Campos
dos Santos
Universidade Federal Rural de Pernambuco

Profa. Dra. Jacira Guiro Marino
Universidade Federal de Pernambuco

Prof. Dr. Eufrázio de Andrade Lima Neto
Universidade Federal da Paraíba

Dedico, com carinho, a minha família.

Agradecimentos

Agradeço primeiramente ao Senhor Jesus Cristo o qual me proporcionou a graça de obter sucesso na academia, permitindo finalizar mais esta etapa da minha vida e já providenciou um caminho profissional como professor em uma Universidade Federal.

À minha família que sempre esteve presente nos difíceis momentos da minha vida, ajudando e contribuindo com muita paciência e compreensão nas minhas ausências provocadas pela necessidade de dedicação ao estudo. Gostaria de destacar Rosemary Santos e Manoel Floriano, dos quais sou fruto e devo para sempre o que tenho e o que sou.

À minha avó Albenízia Carreiro, obrigado por todo seu carinho e cuidado, eles tornaram mais brando o árduo caminho percorrido durante estes anos.

À minha querida irmã Alexsandra Santos, pela amizade, carinho e alegria. Em fim, mesmo se estivermos longe por conta da distância, sempre estarás no meu coração.

À minha amada Glauciana Izídio pelo carinho, atenção, amizade e sinceridade nestes maravilhosos anos que estamos juntos, os quais, eu acredito serem apenas o início da eternidade de nossa convivência. Não existem palavras que possam expressar a sua grande contribuição na realização deste trabalho.

Aos amigos Wagner Barreto e Rodrigo Bernardo pela sincera amizade e companheirismo que com certeza irá transgredir muitos anos.

À Vanessa Santos e Juliana Kátia por todos estes anos de companheirismo e amizade. Fico contente em saber que continuarei tendo vocês por perto e agora, também, como companheiras de profissão. Desejo muito sucesso e felicidades para nós.

A todos os professores que contribuíram decisivamente para a minha formação acadêmica e profissional. Principalmente, à professora Dra. Jacira Marinho Guiro a qual devo grande parte da minha vida profissional ao empenho que ela dedicou em solucionar os meus problemas na universidade.

À funcionária D. Zuleide França pelo carinho maternal.

Ao secretário Marco Antônio dos Santos pelo seu profissionalismo e atenção.

A todos os amigos que conquistei no curso de Biometria e Estatística Aplicada da Universidade Federal Rural de Pernambuco.

De forma geral, agradeço a todos que de alguma forma contribuíram direta ou indiretamente para a construção da minha vida profissional.

"No futuro, o pensamento estatístico será tão necessário para a cidadania eficiente como saber ler e escrever"

Herbert George Wells (1866 - 1946)

Resumo

Os modelos não-lineares da família exponencial são uma extensão dos modelos generalizados, abrindo um leque de opções para a distribuição da variável resposta e permitindo maior flexibilidade para a ligação entre a média e a componente sistemática. Estes modelos, por serem menos restritivos, têm sido utilizados para modelar diversos fenômenos na natureza. Para estimar os parâmetros destes modelos, vários procedimentos são propostos. Usualmente, o método de máxima verossimilhança, que tem propriedades assintóticas de ordem $n-1$, onde n é o tamanho da amostra, é o mais utilizado. Neste trabalho faremos uma abordagem geral dos modelos não-lineares da família exponencial. Será introduzida a teoria da família exponencial sendo apresentada a função de densidade de probabilidade, função geratriz de cumulantes, função de verossimilhança, razão de verossimilhança e desvio do modelo; tais resultados apresentados facilitarão e/ou serão necessários na compreensão do que será feito para os modelos não-lineares da família exponencial. Será definido o modelo não-linear da família exponencial sendo apresentadas as suposições do modelo, sua função de verossimilhança e algoritmo da estimação dos parâmetros. Faremos a abordagem da análise de diagnóstico e de influência dos modelos não-lineares da família exponencial. Por fim, faremos aplicações e mostraremos a eficiência e importância na utilização desta classe, uma vez que diversos fenômenos apresentam comportamento não-linear.

Palavras-chave: Família Exponencial, Função de Verossimilhança, Modelos não-lineares da Família Exponencial.

Abstract

The exponential family nonlinear models are an extension of the generalized models, opening various options for the distribution of the variable answer and allowing larger flexibility for the connection between the average and the systematic component. These models, for being less restrictive, having been used to model several phenomena in the nature. To estimate the parameters of these models, several procedures are proposed. Usually, the method of maximum likelihood, that has asymptotic properties of order n^{-1} , where n is the size of the sample, it is the used. In this work we will make a general approach to the no-linear models of the exponential family. The theory of the exponential family will be introduced presenting the function of density of probability, function cumulantes geratriz, likelihood function, likelihood ratio and deviation of the model; such presented results will facilitate and/or they will be necessary in the understanding of what will be done for the nonlinear models of the exponential family. The exponential family nonlinear models will be defined by presenting the suppositions of the model, its likelihood function and the algorithm for the estimate of the parameters. We will make the approach of the diagnosis analysis and of influence of the exponential family nonlinear models. Finally, we will present some applications and we will show the efficiency and importance in the use of this class, once several phenomena present nonlinear behavior.

Keywords: Exponential family, Likelihood function, Exponential family nonlinear models.

Lista de Figuras

| | | |
|-----|--|----|
| 5.1 | Gráfico de dispersão e reta de regressão estimada para o ajuste dos bancos de dados 1, 2, 3 e 4. | 68 |
| 5.2 | Gráfico de dispersão e reta de regressão estimada para o ajuste dos bancos de dados 5, 6 e 7. | 69 |

Lista de Tabelas

| | | |
|-----|--|----|
| 2.1 | Características de alguns membros da família exponencial | 18 |
| 4.1 | Resumo das estatísticas de diagnóstico para os modelos não-lineares da família exponencial | 49 |
| 5.1 | Alguns modelos não-lineares. | 66 |
| 5.2 | Estimativas dos parâmetros dos modelos ajustados para os bancos 1 a 7, respectivamente. | 67 |

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 14 |
| 2 | Família Exponencial | 16 |
| 2.1 | Definição | 16 |
| 2.2 | Função geratriz de momentos e cumulantes | 19 |
| 2.3 | Verossimilhança | 20 |
| 2.4 | Razão de Verossimilhanças e Desvio | 22 |
| 3 | Modelos Não-Lineares da Família Exponencial | 26 |
| 3.1 | Definição do Modelo | 26 |
| 3.2 | Verossimilhança | 29 |
| 3.3 | Estimação dos Parâmetros | 34 |
| 4 | Análise de Diagnóstico dos Modelos Não-Lineares da Família Exponencial | 38 |
| 4.1 | Introdução | 38 |
| 4.2 | Análise de diagnóstico de modelos | 40 |
| 4.2.1 | Diagnóstico do modelo | 40 |
| 4.2.2 | Coefficiente de regressão | 41 |
| 4.2.3 | Desvio e parâmetro de dispersão | 43 |
| 4.2.4 | Estatística score de outlier | 44 |
| 4.3 | Diagnóstico de influência baseado em casos apagados | 45 |
| 4.3.1 | Diagnóstico baseado em mínimos quadrados ponderados | 45 |
| 4.3.2 | Diagnóstico baseado em desvio | 47 |

| | | |
|----------|--|-----------|
| 4.4 | Análise de influência local | 49 |
| 4.4.1 | Modelos perturbados | 49 |
| 4.4.2 | Esquema de perturbação aleatória | 52 |
| 4.5 | Influência generalizada | 56 |
| 4.5.1 | Definição e computação | 56 |
| 4.5.2 | Influência generalizada e influência local | 58 |
| 4.6 | Diagnósticos para dispersão da variância | 59 |
| 4.6.1 | Razão de verossimilhanças e estatística score | 60 |
| 4.6.2 | Razão de verossimilhanças ajustada e estatística score | 62 |
| 5 | Aplicação | 66 |
| 6 | Conclusões | 70 |
| | Referências Bibliográficas | 71 |

1 Introdução

Os modelos não-lineares da família exponencial (MNLFE) generalizam os modelos de regressão normal linear e não-linear, abrindo um leque de opções para a distribuição da variável resposta e permitindo maior flexibilidade para a ligação entre a média e a componente sistemática. Os MNLFE são caracterizados por uma componente aleatória, que é o vetor resposta \mathbf{Y} com elementos Y_1, \dots, Y_n pertencentes à família exponencial, e por uma componente sistemática, que é o preditor não-linear η dos modelos normais não-lineares.

Para estimar os parâmetros vários procedimentos podem ser empregados. Porém, usaremos a teoria da verossimilhança apresentada por Cordeiro (1999). Este método baseia-se nos princípios de suficiência, fraco de verossimilhança e forte de verossimilhança. As definições desses princípios são apresentadas a seguir:

- O princípio de suficiência diz que vetores de dados distintos com os mesmos valores das estatísticas suficientes para um vetor de parâmetros θ fornecem conclusões idênticas para θ ;
- O princípio fraco da verossimilhança estabelece que vetores de dados com verossimilhança proporcionais produzem as mesmas conclusões sobre θ ;
- Por sua vez, o princípio forte da verossimilhança garante que para variáveis aleatórias distintas X e Y , que dependem de um mesmo parâmetro e de um mesmo espaço paramétrico, as conclusões sobre θ obtidas destes dois vetores de dados serão idênticas, uma vez que dois modelos são adequados aos vetores de dados \mathbf{x} e \mathbf{y} em questão e que fornecem verossimilhanças proporcionais.

Na etapa de inferência do modelo deve-se verificar a precisão e a interdependência das estimativas, construir regiões de confiança e testes sobre os parâmetros de interesse, fazer análise dos resíduos e realizar previsões, podendo assim, avaliar a bondade de ajuste dos modelos. Nesta etapa é necessário conhecer a distribuição de probabilidade do estimador, que, em geral, é obtida através de resultados assintóticos, que são necessários em dois casos distintos:

1. Quando não se tem solução exata para o problema estatístico ou ela é complicada.
2. Quando não se tem solução exata sendo, assim, inevitável a obtenção da solução aproximada.

Para utilização da teoria assintótica de primeira ordem considera-se que o tamanho da amostra seja grande ($n \rightarrow \infty$) tendo como base a expansão da série de Taylor e os teoremas centrais do limite.

Portanto, neste trabalho, temos o objetivo de fazer uma abordagem geral dos modelos não-lineares da família exponencial. No Capítulo 2 será introduzida a teoria da família exponencial sendo apresentada a função de densidade de probabilidade, função geratriz de cumulantes, função de verossimilhança, razão de verossimilhança e desvio do modelo; tais resultados apresentados neste capítulo facilitam e/ou são necessários na compreensão do que será feito para os modelos não-lineares da família exponencial. No Capítulo 3 será definido o modelo não-linear da família exponencial apresentando as suposições do modelo, sua função de verossimilhança e algoritmo da estimação dos parâmetros. No Capítulo 4 faremos a abordagem da análise de diagnóstico e de influência dos modelos não-lineares da família exponencial. Por fim, no capítulo 5, faremos aplicações e mostraremos a eficiência e importância da utilização desta classe, uma vez que diversos fenômenos apresentam comportamento não-linear.

2 Família Exponencial

2.1 Definição

Uma família de distribuição $P_{\theta, \phi}$ com vetor da variável aleatória $\mathbf{Y} = (y_1, \dots, y_n)^T$ e função de densidade de probabilidade

$$p(\mathbf{y}; \theta, \phi) = \exp\{\phi[\mathbf{y}^T \theta - b(\theta) - c(\mathbf{y}, \phi)]\} \quad (2.1)$$

a respeito de uma medida finita ξ , é chamada de família exponencial, onde $b(\cdot)$ e $c(\cdot, \cdot)$ são funções específicas, $\theta = (\theta_1, \dots, \theta_n)^T$ é um parâmetro natural definido em um espaço de parâmetros naturais $\Theta \subset R^n$, e T denota a transposta de um vetor ou matriz. O parâmetro de dispersão ϕ , definido em um subconjunto Φ da reta real \mathbb{R} , é denotado por $\phi = \sigma^{-2}$.

A notação para uma variável aleatória que pertence à família exponencial (2.1) é dada por $\mathbf{Y} \sim ED(\theta, \phi)$ ou $\mathbf{Y} \sim ED(\mu, \sigma^2)$, em que $\mu = E(\mathbf{Y})$ é vetor de média (Jorgensen, 1987). Uma forma equivalente para a equação (2.1) é dada por

$$p(\mathbf{y}; \theta, \phi) = \exp\{\phi[\mathbf{y}^T \theta - b(\theta) - c(\mathbf{y})] - \frac{1}{2}s(\mathbf{y}, \phi)\} \quad (2.2)$$

em que $b(\cdot)$, $c(\cdot)$ e $s(\cdot, \cdot)$ são funções específicas (Cordeiro e McCullagh, 1991).

Quando o parâmetro de dispersão ϕ é conhecido, as equações (2.1) e (2.2) definem a família exponencial linear introduzida por Barndorff-Nilsen (1978) e Morris (1982). Neste caso, Θ é um subconjunto convexo do R^n e a função geratriz de cumulante $b(\theta)$ é convexa e analítica no ponto Θ^0 de Θ . Quando ϕ é desconhecido, as equações (2.1) e (2.2) são chamadas de modelos de dispersão exponencial (Jorgensen, 1987).

O modelo (2.2) normalmente ocorre em dois casos especiais:

(a) Sendo $y = y_i$ e $\theta = \theta_i$ escalares, considera-se que a forma (2.2) depende de uma variável

conhecida e independente y_i . Logo, a função de densidade é expressa por

$$p(y_i; \theta_i, \phi) = \exp\{\phi[y_i\theta_i - b(\theta_i) - c(y_i)] - \frac{1}{2}s(y_i, \phi)\}. \quad (2.3)$$

(b) As componentes de \mathbf{Y} em (2.2) são independentes e cada componente y_i possui uma distribuição que pertence à família exponencial da forma (2.3). Logo, a distribuição de $\mathbf{Y} = (y_1, \dots, y_n)^T$ continua sendo da forma (2.2) denotada por

$$p(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp\left\{\sum_{i=1}^n \phi[y_i\theta_i - b(\theta_i) - c(y_i)] - \frac{1}{2}\sum_{i=1}^n s(y_i, \phi)\right\}, \quad (2.4)$$

em que $b(\boldsymbol{\theta}) = \sum_{i=1}^n b(\theta_i)$, $c(\mathbf{y}) = \sum_{i=1}^n c(y_i)$ e $s(\mathbf{y}, \phi) = \sum_{i=1}^n s(y_i, \phi)$, respectivamente.

Concluí-se, então, que a família exponencial (2.1) é equivalente a família exponencial (2.2), enquanto (2.3) e (2.4) são casos especiais de (2.2). Em alguns casos, a função $s(\cdot)$ pode assumir a forma

$$s(\mathbf{y}, \phi) = s(\phi) + t(\mathbf{y}) \quad (2.5)$$

Combinando (2.4) e (2.5), temos que a família exponencial habitualmente usada é dada por

$$p(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp\left\{\sum_{i=1}^n \phi[y_i\theta_i - b(\theta_i) - c(y_i)] - \frac{1}{2}\sum_{i=1}^n s(\phi) - \frac{1}{2}\sum_{i=1}^n t(y_i)\right\}. \quad (2.6)$$

Na Tabela 2.1 são apresentados os componentes da família exponencial para algumas distribuições conhecidas.

Tabela 2.1: Características de alguns membros da família exponencial

| Modelo | Binomial | Poisson | Binomial Negativa | Normal | Gama | Normal Inverso |
|---------------------|---|---|---|---------------------------|--|---------------------------------------|
| Item | $B(n, p_i)$ | $P(\lambda_i)$ | $BN(v, p_i)$ | $N(\mu_i, \sigma^2)$ | $GA(\mu_i, v)$ | $NI(\mu_i, \sigma^2)$ |
| ϕ | 1 | 1 | 1 | σ^{-2} | v | σ^{-2} |
| θ_i | $\log(p_i/(1-p_i))$ | $\log \lambda_i$ | $\log(1-p_i)$ | μ_i | $-\mu_i^{-1}$ | $-(2\mu_i^2)^{-1}$ |
| μ_i | $ne^{\theta_i}/(1+e^{\theta_i})$ | e^{θ_i} | $ve^{\theta_i}/(1-e^{\theta_i})$ | θ_i | $-\theta_i^{-1}$ | $(-1/2\theta_i)^{1/2}$ |
| $b(\theta_i)$ | $n \log(1+e^{\theta_i})$ | e^{θ_i} | $v \log(1-e^{\theta_i})^{-1}$ | $\theta_i^2/2$ | $-\log(-\theta_i)$ | $-(2\theta_i)^{1/2}$ |
| $c(y_i)$ | $-\log\left(\frac{\mu_i}{y_i}\right)$ | $\log(y_i!)$ | $-\log\left(\frac{v+y_i-1}{v-1}\right)$ | $y_i^2/2$ | $-\log(y_i)$ | $(2y_i)^{-1}$ |
| $s(y_i, \phi)$ | 0 | 0 | 0 | $-\log \phi + \log(2\pi)$ | $-2\{\phi \log \phi - \log \Gamma(\phi)\} + 2 \log y_i$ | $-\{\log \phi - \log(2\pi y_i^3)\}$ |
| $\sigma^2 V(\mu_i)$ | $n^{-1} \mu_i(n - \mu_i)$ | μ_i | $\mu_i + v^{-1} \mu_i^2$ | σ^2 | $\sigma^2 \mu_i^2$ | $\sigma^2 \mu_i^3$ |
| $d_i(y_i, \mu_i)$ | $2\left\{y_i \log(y_i/\mu_i) + (n-y_i) \log\left(\frac{n-y_i}{n-\mu_i}\right)\right\}$ | $2\{y_i \log(y_i/\mu_i) - (y_i - \mu_i)\}$ | $2\left[y_i \log\left\{\frac{y_i(v+\mu_i)}{\mu_i(v+y_i)}\right\} + v \log\left(\frac{v+y_i}{v+\mu_i}\right)\right]$ | $(y_i - \mu_i)^2$ | $2\left\{\frac{y_i - \mu_i}{\mu_i} - \log\left(\frac{y_i}{\mu_i}\right)\right\}$ | $(y_i - \mu_i)^2 / \mu_i^2 y_i$ |
| $d'_i(y_i, \mu_i)$ | $-2\left\{y_i \log \frac{\mu_i}{n-\mu_i} + n \log\left(\frac{n-\mu_i}{n}\right) + \log\binom{n}{y_i}\right\}$ | $-2\{y_i \log \mu_i - \mu_i - \log(y_i!)\}$ | $v \log\left(\frac{v}{v+\mu_i}\right) + \log\left(\frac{v+y_i-1}{v-1}\right)$ | $(y_i - \mu_i)^2$ | $2\left\{\frac{y_i}{\mu_i} - \log \frac{y_i}{\mu_i}\right\}$ | $\frac{(y_i - \mu_i)^2}{\mu_i^2 y_i}$ |
| Nota | $\mu_i = np_i$ $\sigma^2 = 1$ | $\mu_i = \lambda_i$ $\sigma^2 = 1$ | $\mu_i = v \frac{-p_i}{p_i}$ $\sigma^2 = 1$ | $d'_i(y_i, y_i) = 0$ | $d'_i(y_i, y_i) = 2$ $\sigma^2 = v^{-1}$ | $d'_i(y_i, y_i) = 0$ |

2.2 Função geratriz de momentos e cumulantes

Na família exponencial da forma (2.1) ou (2.2), a função geratriz de momentos é expressa por

$$M(\tau; \theta, \phi) = E(e^{\tau^T Y}) = \exp\{\phi[b(\theta + \tau/\phi) - b(\theta)]\}. \quad (2.7)$$

Tomando a derivada de $M(\tau; \theta, \phi)$ em $\tau = \mathbf{0}$, pode-se encontrar os cumulantes de \mathbf{Y} de ordem (i_1, \dots, i_n) como segue

$$k_{i_1 \dots i_n}(\theta, \phi) = \phi^{1-i_1-\dots-i_n} \frac{\partial^{i_1+\dots+i_n} b(\theta)}{\partial \theta_1^{i_1} \dots \partial \theta_n^{i_n}}.$$

Em particular, temos

$$E(\mathbf{Y}) = \boldsymbol{\mu} = b'(\theta), \quad \text{Var}(\mathbf{Y}) = \sigma^2 b''(\theta) = \sigma^2 \mathbf{V}(\boldsymbol{\mu}), \quad (2.8)$$

em que $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ é o parâmetro esperado definido em um subconjunto $U \subset \mathbb{R}^n$ quando $\theta \in \Theta^0$; $b'(\theta) = \frac{\partial b(\theta)}{\partial \theta} = \left(\frac{\partial b(\theta)}{\partial \theta_i} \right)$ é um vetor de tamanho n e $b''(\theta) = \frac{\partial^2 b(\theta)}{\partial \theta_i \partial \theta_j}$ é uma matriz $n \times n$. Considerando $\boldsymbol{\mu} = b'(\theta)$, tem-se $\theta = b'^{-1}(\boldsymbol{\mu})$, em que $b'^{-1}(\cdot)$ é a função inversa de $b'(\cdot)$, mostrando uma ligação entre $\boldsymbol{\mu}$ e θ . A função de variância $b''(\theta) = b''(b'^{-1}(\boldsymbol{\mu}))$ é denotada por $\mathbf{V}(\boldsymbol{\mu})$, além disso, nota-se que para os modelos (2.4) e (2.6), \mathbf{V}_s são matrizes diagonais.

De forma similar a (2.7), obtém-se a função geratriz de momentos para os erros aleatórios, $\mathbf{e} = \mathbf{Y} - \boldsymbol{\mu} = (e_1, \dots, e_n)^T$, adquirindo os momentos centrais de \mathbf{Y} . De fato, tem-se

$$E(e^{\tau^T \mathbf{e}}) = \exp\{\phi[b(\theta + \tau/\phi) - b(\theta)] - \tau^T \boldsymbol{\mu}\}.$$

Após alguns cálculos, obtém-se

$$E(e_i) = 0, \quad E(e_i e_j) = \sigma^2 V_{ij}, \quad (2.9)$$

$$E(e_i e_j e_k) = \sigma^4 S_{ijk}, \quad S_{ijk} = \frac{\partial^3 b(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}, \quad (2.10)$$

e

$$E(e_i e_j e_k e_l) = \sigma^4 (V_{ij} V_{kl} + V_{ik} V_{jl} + V_{il} V_{jk}) + \sigma^6 \Delta_{ijkl}, \quad (2.11)$$

em que $\mathbf{V}(\boldsymbol{\mu}) = (V_{ij})$, $\Delta_{ijkl} = \frac{\partial^4 b(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k \partial \theta_l}$ e $i, j, k, l = 1, \dots, n$.

Lema 2.2.1. Para os modelos (2.1) e (2.2), as fórmulas das primeiras e segundas derivadas são

$$\frac{\partial \mu}{\partial \theta^T} = \mathbf{V}(\theta), \quad \frac{\partial^2 \mu}{\partial \theta \partial \theta^T} = b^{(3)}(\theta) = S \quad (2.12)$$

$$\frac{\partial \theta}{\partial \mu^T} = \mathbf{V}^{-1}(\mu), \quad \frac{\partial^2 \theta}{\partial \mu \partial \mu^T} = -[\mathbf{V}^{-1}][\mathbf{V}^{-1}S\mathbf{V}^{-1}] \quad (2.13)$$

em que $b^{(3)}(\theta) = S$, $\frac{\partial^2 \mu}{\partial \theta \partial \theta^T}$ e $\frac{\partial^2 \theta}{\partial \mu \partial \mu^T}$ são vetores $n \times n \times n$ com elementos $\frac{\partial^3 b(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}$, $\frac{\partial^2 \mu_i}{\partial \theta_j \partial \theta_k}$ e $\frac{\partial^2 \theta_i}{\partial \mu_j \partial \mu_k}$, respectivamente. A expressão $[\cdot][\cdot]$ denota o produto de dois vetores (Seber e Wild, 1989).

Lema 2.2.2. Se o modelo (2.2) satisfaz (2.5) e as derivadas de $c(\cdot)$ e $s(\cdot)$ existem, então as derivadas de $b(\cdot)$ e $c(\cdot)$ satisfazem $b'(\cdot) = c'^{-1}(\cdot)$ (McCullagh, 1983; Smyth, 1989).

2.3 Verossimilhança

No modelo (2.2), a log-verossimilhança para o parâmetro θ ou μ é denotada por $l(\mu, \mathbf{y}) = \log p(\mathbf{y}; \theta, \phi)$ para um valor fixo do parâmetro ϕ . Tem-se que

$$l(\mu, \mathbf{y}) = \phi[\mathbf{y}^T \theta - b(\theta) - c(\mathbf{y})] - \frac{1}{2}s(\mathbf{y}, \phi), \quad (2.14)$$

em que $\theta = b'^{-1}(\mu)$ e ϕ estão implícitos em $l(\mu, \mathbf{y})$. Segue do Lema 2.2.1 que a função escore de \mathbf{Y} para θ e μ são, respectivamente,

$$l'_\theta = \frac{\partial l(\mu, \mathbf{y})}{\partial \theta} = \phi(\mathbf{Y} - \mu) = \phi \mathbf{e} \quad (2.15)$$

e

$$l'_\mu = \frac{\partial l(\mu, \mathbf{y})}{\partial \mu} = \phi \mathbf{V}^{-1}(\mu)(\mathbf{Y} - \mu) = \phi \mathbf{V}^{-1} \mathbf{e}, \quad (2.16)$$

em que usa-se ' \mathbf{e} ' para denotar o erro aleatório $(\mathbf{Y} - \mu)$. As características de ' \mathbf{e} ' estão nas equações (2.9) a (2.11). Verifica-se facilmente que os estimadores de máxima verossimilhança (EMV's) de μ e θ são $\tilde{\mu} = \mathbf{y}$ e $\tilde{\theta} = b'^{-1}(\tilde{\mu})$, respectivamente.

Lema 2.3.1. Para o modelo (2.2), as derivadas de segunda ordem do logaritmo da função de

verossimilhança de \mathbf{Y} para θ e μ são, respectivamente, representadas por

$$-l''_{\theta\theta} = -\frac{\partial^2 l(\mu, \mathbf{y})}{\partial \theta \partial \theta^T} = \phi b''(\theta) = \phi \mathbf{V}(\mu) \quad (2.17)$$

e

$$-l''_{\mu\mu} = -\frac{\partial^2 l(\mu, \mathbf{y})}{\partial \mu \partial \mu^T} = \phi \mathbf{V}^{-1} + \phi [\mathbf{e}^T \mathbf{V}^{-1}] [\mathbf{V}^{-1} \mathbf{S} \mathbf{V}^{-1}]. \quad (2.18)$$

A matriz de informação de Fisher para θ e para μ pode ser expressa como

$$\mathbf{J}_\theta(\mathbf{Y}) = \phi \mathbf{V}(\theta), \quad \mathbf{J}_\mu(\mathbf{Y}) = \phi \mathbf{V}^{-1}(\mu). \quad (2.19)$$

Por fim, consideramos o parâmetro de dispersão $\phi = \sigma^{-2}$. Além disso, através de (2.14), verifica-se que

$$l'_\phi = \frac{\partial l(\mu, \mathbf{y})}{\partial \phi} = [\mathbf{y}^T \theta - b(\theta) - c(\mathbf{y})] - \frac{1}{2} s'(\mathbf{y}, \phi),$$

$$-l''_{\phi\phi} = -\frac{\partial^2 l(\mu, \mathbf{y})}{\partial \phi^2} = \frac{1}{2} s''(\mathbf{y}, \phi),$$

e

$$-l''_{\phi\theta} = \mathbf{y} - \mu, \quad -l''_{\phi\mu} = \mathbf{V}^{-1}(\mathbf{y} - \mu),$$

em que $s'(\mathbf{y}, \phi)$ e $s''(\mathbf{y}, \phi)$ são as primeiras e segundas derivadas de $s(\mathbf{y}, \phi)$ com relação a ϕ . Assumimos que $c(\mathbf{Y})$ e $s(\mathbf{Y}, \phi)$ satisfazem $E(l'_\phi) = 0$. É fácil verificar que o parâmetro de dispersão $\phi = \sigma^{-2}$ é ortogonal aos parâmetros θ e μ (Cox e Reid, 1987), isto é, $E(-l''_{\phi\theta}) = 0$ e $E(-l''_{\phi\mu}) = 0$. O EMV $\tilde{\phi}$ de ϕ satisfaz

$$s'(\mathbf{Y}, \tilde{\phi}) = 2[\mathbf{Y}^T \tilde{\theta} - b(\tilde{\theta}) - c(\mathbf{Y})].$$

Considerando que a condição (2.6) é satisfeita, então, $s'(\mathbf{y}, \phi)$ e $s''(\mathbf{y}, \phi)$ será simplificado por $s'(\phi)$ e $s''(\phi)$, respectivamente. Logo, $\tilde{\phi}$ será denotado por

$$\tilde{\phi} = s'^{-1}(2[\mathbf{Y}^T \hat{\theta} - b(\hat{\theta}) - c(\mathbf{Y})]).$$

A informação de Fisher de \mathbf{Y} para ϕ é representado pela forma

$$\mathbf{J}_\phi(\mathbf{Y}) = -l''_{\phi\phi} = \frac{1}{2} s''(\phi) \quad (2.20)$$

assumindo $s''(\phi) > 0$ em Φ .

2.4 Razão de Verossimilhanças e Desvio

O desvio é uma estatística importante e muito conhecida em problemas relacionados com a família exponencial. Inicialmente, introduzimos o desvio através da estatística da razão de verossimilhança. Para o modelo (2.2), com o parâmetro de dispersão ϕ fixo, considere-se o teste de hipótese

$$H_0 : \mu = \mu_0; \quad H_1 : \mu \neq \mu_0.$$

Sendo $\tilde{\mu} = \mathbf{y}$, a estatística da razão de verossimilhanças do teste é

$$\begin{aligned} LR(\mu_0) &= 2[l(\tilde{\mu}, \mathbf{y}) - l(\mu_0, \mathbf{y})] \\ &= 2\phi[\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta})]_{\mu=\mathbf{y}} - 2\phi[\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta})]_{\mu=\mu_0} \\ &= 2\phi[\mathbf{y}^T q(\mathbf{y}) - b(q(\mathbf{y}))] - 2\phi[\mathbf{y}^T q(\mu_0) - b(q(\mu_0))], \end{aligned}$$

em que $\boldsymbol{\theta} = b'^{-1}(\mu) = q(\mu)$. Sabe-se que $LR(\mu_0)$, assintoticamente, tem distribuição $\chi^2(n)$ (Cox e Hinkley, 1974). Logo, pode-se reescrever a equação como

$$LR(\mu) = D^*(\mathbf{y}, \mu) = \phi D(\mathbf{y}, \mu) \sim \chi^2(n),$$

em que

$$\begin{aligned} D(\mathbf{y}, \mu) &= 2[\mathbf{y}^T q(\mathbf{y}) - b(q(\mathbf{y}))] - 2[\mathbf{y}^T q(\mu) - b(q(\mu))] \\ &= 2[\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\mathbf{y})]_{\mu=\mathbf{y}} - 2[\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\mathbf{y})]. \end{aligned} \quad (2.21)$$

A quantidade $D(\mathbf{y}, \mu)$ é chamada desvio do modelo, que é uma estatística importante em modelos lineares generalizados e em MNLFE. De fato o desvio $D(\mathbf{y}, \mu)$ é a parte do núcleo da função log-verossimilhança $l(\mu, \mathbf{y})$ em (2.14)

$$l(\mu, \mathbf{y}) = -\frac{1}{2}\phi D(\mathbf{y}, \mu) - \frac{1}{2}s(\mathbf{y}, \phi) + \phi[\mathbf{y}^T \boldsymbol{\theta} - b(\boldsymbol{\theta}) - c(\mathbf{y})]_{\mu=\mathbf{y}}.$$

Além disso, tem-se a seguinte representação de Hoeffding:

Lema 2.4.1. *A família exponencial (2.1) ou (2.2) e o logaritmo da função de verossimilhança podem ser representadas por*

$$p(\mathbf{y}; \mu, \phi) = p(\mathbf{y}; \mathbf{y}, \phi) \exp\left[-\frac{1}{2}\phi D(\mathbf{y}, \mu)\right] \quad (2.22)$$

e

$$l(\boldsymbol{\mu}, \boldsymbol{\phi}; \mathbf{y}) = l(\mathbf{y}, \boldsymbol{\phi}; \mathbf{y}) - \frac{1}{2} \boldsymbol{\phi} D(\mathbf{y}, \boldsymbol{\mu}), \quad (2.23)$$

respectivamente.

O Lema 2.4.1 mostra que o logaritmo da função de verossimilhança para $\boldsymbol{\mu}$ ou $\boldsymbol{\theta}$ é de fato proporcional ao desvio $D(\mathbf{y}, \boldsymbol{\mu})$. Então, pode-se encontrar o EMV para $\boldsymbol{\mu}$ e $\boldsymbol{\theta}$ através da minimização de $D(\mathbf{y}, \boldsymbol{\mu})$ ao invés de maximizar $l(\boldsymbol{\mu}, \boldsymbol{\phi}; \mathbf{y})$, enquanto o EMV para $\boldsymbol{\phi}$ pode ser obtido através da equação (2.23). Além disso, é fácil verificar que

$$\min_{\boldsymbol{\mu}} D(\mathbf{y}, \boldsymbol{\mu}) = D(\mathbf{y}, \mathbf{y}) = 0.$$

Da equação (2.21), obtém-se as derivadas de $D(\mathbf{y}, \boldsymbol{\mu})$ expressas por

$$D'_{\boldsymbol{\theta}} = \frac{\partial \{D(\mathbf{y}, \boldsymbol{\mu})\}}{\partial \boldsymbol{\theta}} = -2(\mathbf{y} - \boldsymbol{\mu}) = -2\boldsymbol{\phi}^{-1} l'_{\boldsymbol{\theta}},$$

$$D'_{\boldsymbol{\mu}} = \frac{\partial \{D(\mathbf{y}, \boldsymbol{\mu})\}}{\partial \boldsymbol{\mu}} = -2\mathbf{V}^{-1}(\mathbf{y} - \boldsymbol{\mu}) = -2\boldsymbol{\phi}^{-1} l'_{\boldsymbol{\mu}}.$$

Na família exponencial, há uma relação entre o desvio e a distância de Kullback-Leibler, que é definida por

$$K(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = E_{\boldsymbol{\mu}_1} \left\{ \log \frac{p(\mathbf{y}; \boldsymbol{\mu}_1, \boldsymbol{\phi})}{p(\mathbf{y}; \boldsymbol{\mu}_2, \boldsymbol{\phi})} \right\},$$

em que $\boldsymbol{\mu}_1$ e $\boldsymbol{\mu}_2$ são dois valores de $\boldsymbol{\mu}$ e os valores correspondentes de $\boldsymbol{\theta}$ são denotados por $\boldsymbol{\theta}_1$ e $\boldsymbol{\theta}_2$, respectivamente. De (2.1) ou (2.2) temos que

$$K(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \boldsymbol{\phi} [\boldsymbol{\mu}_1^T \boldsymbol{\theta}_1 - b(\boldsymbol{\theta}_1)] - \boldsymbol{\phi} [\boldsymbol{\mu}_2^T \boldsymbol{\theta}_2 - b(\boldsymbol{\theta}_2)].$$

Considerando $\boldsymbol{\mu}_1 = \mathbf{y}$, $\boldsymbol{\mu}_2 = \boldsymbol{\mu}$ e comparando com o resultado com (2.21), temos

$$K(\mathbf{y}, \boldsymbol{\mu}) = \frac{1}{2} \boldsymbol{\phi} D(\mathbf{y}, \boldsymbol{\mu}) = \frac{1}{2} LR(\boldsymbol{\mu}).$$

A partir desta equação, (2.22) pode ser escrito da forma

$$p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\phi}) = p(\mathbf{y}; \mathbf{y}, \boldsymbol{\phi}) [-K(\mathbf{y}, \boldsymbol{\mu})]$$

Algumas propriedades dos modelos (2.4) e (2.6) são importantes. Para estes modelos, o desvio pode ser representado na seguinte forma

$$\begin{aligned} D(\mathbf{y}, \boldsymbol{\mu}) &= \sum_{i=1}^n \{2[y_i \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i) - c(y_i)]_{\boldsymbol{\mu}_i=y_i} - 2[y_i \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i) - c(y_i)]\} \\ &= \sum_{i=1}^n d_i(y_i, \boldsymbol{\mu}_i), \end{aligned}$$

em que

$$\begin{aligned} d_i(y_i, \boldsymbol{\mu}_i) &= 2[y_i \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i) - c(y_i)]_{\boldsymbol{\mu}_i=y_i} - 2[y_i \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i) - c(y_i)] \\ &= 2[y_i q(y_i) - b(q(y_i))] - 2[y_i \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i)] \end{aligned}$$

e $q(y_i) = b'^{-1}(y_i)$. A tabela 2.1 mostra o desvio para alguns elementos comuns da família exponencial.

As equações (2.4) e (2.21) mostram que a parte principal da distribuição da família exponencial é a quantidade $2[y_i \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i) - c(y_i)]$, $i = 1, \dots, n$. Smyth (1989) propôs o uso da notação

$$\begin{aligned} d'_i(y_i, \boldsymbol{\mu}_i) &= -2[y_i \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i) - c(y_i)], \\ D'(\mathbf{y}, \boldsymbol{\mu}) &= \sum_{i=1}^n d'_i(y_i, \boldsymbol{\mu}_i). \end{aligned}$$

Logo, tem-se

$$\begin{aligned} d_i(y_i, \boldsymbol{\mu}_i) &= d'_i(y_i, \boldsymbol{\mu}_i) - d'_i(y_i, y_i), \\ D(\mathbf{y}, \boldsymbol{\mu}) &= D'(\mathbf{y}, \boldsymbol{\mu}) - D'(\mathbf{y}, \mathbf{y}). \end{aligned} \tag{2.24}$$

De (2.22), a família exponencial (2.4) pode ser representada por

$$p(\mathbf{y}; \boldsymbol{\mu}, \phi) = \exp \left[-\frac{1}{2} \phi D'(\mathbf{y}, \boldsymbol{\mu}) - \frac{1}{2} \sum_{i=1}^n s(y_i, \phi) \right]. \tag{2.25}$$

Além disso, para o modelo (2.6), pode-se reduzir (2.25) de acordo com o lema seguinte.

Lema 2.4.2. *Para o modelo (2.6), $d'(y_i, y_i)$ é uma constante, digamos k , independente de y_i e satisfaz*

$$\begin{aligned} d_i(y_i, \boldsymbol{\mu}_i) &= d'_i(y_i, \boldsymbol{\mu}_i) - k, \\ D(\mathbf{y}, \boldsymbol{\mu}) &= D'(\mathbf{y}, \boldsymbol{\mu}) - nk, \end{aligned} \tag{2.26}$$

e

$$p(\mathbf{y}; \boldsymbol{\mu}, \phi) = \exp\left[-\frac{1}{2}\phi D(\mathbf{y}, \boldsymbol{\mu})\right] \cdot \exp\left[-\frac{n}{2}(s(\phi) + k\phi) - \frac{1}{2}\sum_{i=1}^n t(y_i)\right]. \quad (2.27)$$

A Tabela (2.1) mostra que $d'_i(y_i, \mu_i) = k = 0$ para família normal e normal inversa. Neste caso temos

$$\begin{aligned} d_i(y_i, \mu_i) &= d'_i(y_i, \mu_i) = -2[y_i\theta_i - b(\theta_i) - c(y_i)], \\ D(\mathbf{y}, \boldsymbol{\mu}) &= D'(\mathbf{y}, \boldsymbol{\mu}). \end{aligned}$$

Para família Gama, $d'(y_i, \mu_i) = 2$, então

$$\begin{aligned} d_i(y_i, \mu_i) &= -2[y_i\theta_i - b(\theta_i) - c(y_i)] - 2, \\ D(\mathbf{y}, \boldsymbol{\mu}) &= D'(\mathbf{y}, \boldsymbol{\mu}) - 2n. \end{aligned}$$

Em algumas situações práticas, o parâmetro de dispersão $\phi = \sigma^{-2}$ em (2.4) pode ser ponderado por pesos w_i (McCullagh e Nelder, 1989; Smyth, 1989), isto é, $y_i \sim ED(\mu_i, \sigma^2 w_i^{-1})$, $i = 1, \dots, n$. Neste caso, (2.4) é expressa por

$$p(\mathbf{y}; \boldsymbol{\theta}, \phi) = \exp\left\{\sum_{i=1}^n \phi w_i [y_i\theta_i - b(\theta_i) - c(y_i)] - \frac{1}{2}\sum_{i=1}^n s(y_i, \phi w_i)\right\}. \quad (2.28)$$

Obviamente, o comportamento estatístico desta família é idêntico ao da família (2.4) a não ser que cada log-verossimilhança de y_i em (2.4) seja ponderado por w_i em todas equações (2.14) a (2.27). Por exemplo, o desvio $d'_i(y_i, \mu_i)$ definido em (2.24) passa a ser

$$d'_i(y_i, \mu_i) = -2w_i[y_i\theta_i - b(\theta_i) - c(y_i)].$$

As outras equações são tratadas semelhantemente.

3 Modelos Não-Lineares da Família Exponencial

3.1 Definição do Modelo

Suponha que os componentes de $\mathbf{Y} = (y_1, \dots, y_n)^T$ são variáveis aleatórias independentes, em que cada y_i possui distribuição da família exponencial dada por (2.3) e pode depender de uma variável conhecida x_i ($i = 1, \dots, n$). Então, a distribuição de \mathbf{Y} é da forma (2.4).

Suponha agora que o parâmetro de interesse em (2.4), $\beta = (\beta_1, \dots, \beta_p)^T$, definido no subconjunto B de \mathbb{R}^p ($p < n$) e a distribuição de y_i que depende de x_i satisfaz a seguinte condição:

$$\eta_i \equiv g(\mu_i) = f(x_i; \beta) \equiv f_i(\beta), i = 1, \dots, n, \quad (3.1)$$

em que $g(\cdot)$ é uma função de ligação monótona conhecida; $f(\cdot, \cdot)$ é uma função conhecida de um vetor β de parâmetros desconhecidos e uma variável explanatória conhecida " x_i ", respectivamente. Então, os modelos (3.1) com (2.4) são denominados de modelos não-lineares da família exponencial (MNLFE) ou modelos não-lineares generalizados (MNLG).

Desta definição, muitos modelos de regressão são encontrados como casos particulares. Os dois exemplos a seguir são casos especiais de interesse:

(a) Se $f(x_i; \beta) = x_i^T \beta$, então (3.1) representa os modelos lineares generalizados (MLG), os quais foram completamente estudados nas últimas duas décadas (McCullagh e Nelder, 1989; Firth, 1991).

(b) Se $g(\mu_i) = \mu_i$, então (3.1) representa uma classe geral de modelos não-lineares de regressão. Em particular, se $g(\mu_i) = \mu_i$ com $b(\theta_i) = \theta_i^2/2$, então (3.1) representa o modelo normal não-linear (Bates e Watts, 1988; Seber e Wild, 1989; Ratkowsky, 1990).

Para utilização destes modelos é necessário que sejam satisfeitas as seguintes suposições:

Suposições A

(a) Quando $\beta_1 \neq \beta_2$, $\phi_1 \neq \phi_2$, $\xi \{p(\mathbf{Y}; \beta_1, \phi_1) \neq p(\mathbf{Y}; \beta_2, \phi_2)\} > 0$

(b) Seja $l(\beta, \phi; \mathbf{y}) = \log p(\mathbf{y}; \theta(\beta), \phi)$. Para β e ϕ fixados, $\partial l / \partial \beta_a$ ($a = 1, \dots, p$) e $\partial l / \partial \phi$ são linearmente independentes.

(c) Os momentos das variáveis aleatórias $\partial l / \partial \beta_a$ ($a = 1, \dots, p$) e $\partial l / \partial \phi$ existem pelo menos até a terceira ordem.

(d) As derivadas parciais $\partial / \partial \beta_a$ ($a = 1, \dots, p$), $\partial / \partial \phi$ e a integral em relação à medida $\xi(dy)$ sempre pode ser substituído por qualquer função integrável $h(\mathbf{y}; \theta, \phi)$. Em particular, $\xi(dy) = dy$, a medida de Lebesgue no \mathbb{R}^n é suposta na maioria das situações.

(e) A função $g(\cdot)$ é diferenciável, pelo menos, até a terceira ordem. $f(x_i; \beta)$ em (3.1) é uma função de x_i definida em um subconjunto compacto X contido no \mathbb{R}^q , é função de β o qual é definido em um subconjunto aberto B contido em \mathbb{R}^p e é diferenciável pelo menos até a terceira ordem. Todas as derivadas anteriores são contínuas em $X \times B$.

(f) $\inf_{\theta} \mathbf{b}''(\theta) > 0$, $\sup_{\theta} |\mathbf{b}^{(3)}(\theta)| < +\infty$.

As condições de regularidades anteriores, (a) até (f), são requeridas para os modelos (3.1) com (2.4). A definição feita acima mostra que o parâmetro β pode ser conectado com um parâmetro natural θ_i e o parâmetro de média μ_i como segue.

(a) Para $\mu_i = b'(\theta_i)$, $g(\mu_i) = g(b'(\theta_i)) = f(x_i; \beta)$, temos:

$$\theta_i = \theta_i(\beta) = b^{-1} \circ g^{-1} \circ f(x_i; \beta),$$

em que ' \circ ' denota o produto de duas funções. Se a função de ligação $g(\cdot)$ está condicionada tal que $\theta_i = \eta_i = f(x_i; \beta)$, essa medida dada por $b^{-1} \circ g^{-1}$ é a função identidade, isto é $g(\cdot) = b^{-1}$; então ambas funções $g(\cdot)$ e o modelo (3.1) são chamados de ligação canônica, que é mais fácil de se trabalhar do que com as ligações não canônicas (McCullagh e Nelder, 1989).

(b) A equação (3.1) também pode ser denotada pelo parâmetro de médias μ_i , como

$$\mu_i = \mu_i(\beta) = g^{-1} \circ f(x_i; \beta).$$

Normalmente, estuda-se o comportamento estatístico de β em termos do parâmetro de média $\mu = \mu(\beta)$ (McCullagh, 1983; Efron, 1986; Jorgensen, 1987; Cordeiro e Paula, 1991; McCullagh e Nelder, 1989). Neste trabalho, usa-se $\mu = \mu(\beta)$ na maioria das situações. Assim, os

modelos (3.1) com (2.4) podem ser representados por

$$p(\mathbf{y}; \boldsymbol{\mu}(\boldsymbol{\beta}), \phi) = \exp \left\{ \sum_{i=1}^n \phi [y_i; \boldsymbol{\theta}_i - b(\boldsymbol{\theta}_i) - c(y_i)] - \frac{1}{2} \sum_{i=1}^n s(y_i, \phi) \right\},$$

$$\text{em que, } \boldsymbol{\mu}_i = \boldsymbol{\mu}_i(\boldsymbol{\beta}) \text{ ou } \boldsymbol{\theta}_i = \boldsymbol{\theta}_i(\boldsymbol{\beta}); \quad i = 1, \dots, n. \quad (3.2)$$

A notação vetorial é, freqüentemente, usada. Então, a equação (3.1) pode ser escrita na forma

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = f(\boldsymbol{\beta}) \text{ ou } \boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta}), \quad (3.3)$$

em que

$$\begin{aligned} \boldsymbol{\eta} &= (\eta_1, \dots, \eta_n)^T, \quad g(\boldsymbol{\mu}) = (g(\boldsymbol{\mu}_1), \dots, g(\boldsymbol{\mu}_n))^T \\ f(\boldsymbol{\beta}) &= (f_1(\boldsymbol{\beta}), \dots, f_n(\boldsymbol{\beta}))^T, \quad f_i(\boldsymbol{\beta}) = f(x_i; \boldsymbol{\beta}), \end{aligned}$$

e

$$\boldsymbol{\mu}(\boldsymbol{\beta}) = g^{-1} \circ f(\boldsymbol{\beta}) = (\boldsymbol{\mu}_1(\boldsymbol{\beta}), \dots, \boldsymbol{\mu}_n(\boldsymbol{\beta}))^T.$$

Por fim, dada a estrutura dos MNLFE (3.1), as propriedades da família exponencial são válidas para qualquer valor fixado de $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$. Neste caso, tem-se

$$p(\mathbf{y}; \boldsymbol{\mu}(\boldsymbol{\beta}), \phi) = p(\mathbf{y}; \mathbf{y}, \phi) \exp \left[-\frac{1}{2} \phi D(\mathbf{y}, \boldsymbol{\mu}(\boldsymbol{\beta})) \right]. \quad (3.4)$$

Esta equação mostra que podemos encontrar o EMV $\hat{\boldsymbol{\beta}}$ de $\boldsymbol{\beta}$ minimizando o desvio $D(\mathbf{y}, \boldsymbol{\mu}(\boldsymbol{\beta}))$ ao invés de maximizar o logaritmo da função de verossimilhança $l(\boldsymbol{\mu}(\boldsymbol{\beta}), \phi; \mathbf{y})$ e que o estimador de máxima verossimilhança $\hat{\phi}$ de ϕ pode se obtido de (3.3) usando $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}(\hat{\boldsymbol{\beta}})$.

Lema 3.1.1. Para o modelo (3.1) com (3.2), a média e a variância do $D'(\mathbf{Y}, \boldsymbol{\mu}(\boldsymbol{\beta}))$ para qualquer $\boldsymbol{\beta}$ pode ser representado por

$$\begin{aligned} E[D'(\mathbf{Y}, \boldsymbol{\mu}(\boldsymbol{\beta}))] &= -\sum_{i=1}^n E[s'(y_i, \phi)], \\ \text{Var}[D'(\mathbf{Y}, \boldsymbol{\mu}(\boldsymbol{\beta}))] &= 2 \sum_{i=1}^n E[s''(y_i, \phi)], \end{aligned}$$

em que $D'(\mathbf{Y}, \boldsymbol{\mu})$ segue de (2.24).

Aplicando o Lema 3.1.1 para o caso em que $n = 1$, temos o seguinte corolário.

Corolário

Para qualquer β e $d'_i(y_i, \mu_i(\beta)) = -2[y_i\theta_i - b(\theta_i) - c(y_i)]$, temos

$$E[d'_i(y_i, \mu_i(\beta))] = -E[s'(y_i, \phi)],$$

e

$$\text{Var}[d'_i(y_i, \mu_i(\beta))] = 2E[s''(y_i, \phi)].$$

Em particular, para o modelo específico (2.6), podemos conseguir resultados mais diretos.

Lema 3.1.2. Para os MNLFE (3.1) com distribuição do tipo (2.6), a média e a variância de $D(\mathbf{Y}, \mu(\beta))$ e $d_i(y_i, \mu_i(\beta))$ para qualquer β é representado, respectivamente, por

$$\begin{aligned} E[D(\mathbf{Y}, \mu(\beta))] &= -n[s'(\phi) + k], \\ \text{Var}[D(\mathbf{Y}, \mu(\beta))] &= 2ns''(\phi); \\ E[d_i(y_i, \mu_i(\beta))] &= -[s'(\phi) + k], \\ \text{Var}[d_i(y_i, \mu_i(\beta))] &= 2s''(\phi); \end{aligned}$$

em que $d'_i(y_i, y_i)$ é uma constante.

É relevante salientar que para o modelo (3.1) com (2.6), a média e a variância do desvio dependem do parâmetro de dispersão mas não dependem do parâmetro β .

3.2 Verossimilhança

Seja $\hat{\beta}$ e $\hat{\phi}$ os EMV de β e ϕ para os MNLFE (3.1) e as quantidades correspondentes de θ e μ , denotadas, por $\hat{\theta} = \theta(\hat{\beta})$ e $\hat{\mu} = \mu(\hat{\beta})$, respectivamente. A equação (3.4) mostra que podemos encontrar $\hat{\beta}$, $\hat{\phi} = \phi(\hat{\beta})$ separadamente, a partir de um primeiro $\hat{\beta}$. Utilizando um esquema iterativo para obtenção de $\hat{\beta}$, e algumas proposições teóricas, iniciamos com o cálculo da primeira e segunda derivadas da log-verossimilhança para o parâmetro β . Para simplificar, denota-se para ϕ fixo, $l(\mu(\beta), \mathbf{y})$ ou $l(\theta(\beta), \mathbf{y})$ por $l(\beta)$, o qual é expresso como

$$l(\beta) = l(\mu(\beta), \mathbf{y}) = \phi \{ \mathbf{y}^T \theta - b(\theta) - c(\mathbf{y}) \} - \frac{1}{2} s(\mathbf{y}, \phi), \quad \mu = \mu(\beta) \quad (3.5)$$

e

$$\begin{aligned} l(\beta) &= -\frac{1}{2}\phi D'(\mathbf{y}, \mu(\beta)) - \frac{1}{2}s(\mathbf{y}, \phi) \\ &= -\frac{1}{2}\phi D(\mathbf{y}, \mu(\beta)) - \frac{1}{2}s(\mathbf{y}, \phi) - \frac{1}{2}\phi D'(\mathbf{y}, \mathbf{y}). \end{aligned}$$

A primeira e segunda derivada de $l(\beta)$ com relação a β é denotado por $l'(\beta)$ e $l''(\beta)$, respectivamente. Note que no modelo (3.1) com (3.2), a função de variância $b''(\theta) = \mathbf{V}$ é a matriz diagonal. Normalmente, definimos por

$$\mathbf{V} = \text{diag}(V_i), V_i = V_{ii}$$

para $i = 1, \dots, n$. Além disso, definimos

$$\begin{aligned} \mathbf{D}(\beta) &= \frac{\partial \mu(\beta)}{\partial \beta^T}, & \mathbf{W}(\beta) &= \frac{\partial^2 \mu(\beta)}{\partial \beta \partial \beta^T} \\ \mathbf{D}_\theta(\beta) &= \frac{\partial \theta(\beta)}{\partial \beta^T}, & \mathbf{W}_\theta(\beta) &= \frac{\partial^2 \theta(\beta)}{\partial \beta \partial \beta^T}, \end{aligned}$$

em que \mathbf{D} e \mathbf{D}_θ são matrizes $n \times p$, e \mathbf{W} e \mathbf{W}_θ possuem dimensão $n \times p \times p$, respectivamente. Através das **Suposições A**, todas essas derivadas existem. A relação entre essas derivadas é dada por

$$\mathbf{D}_\theta = \mathbf{V}^{-1}\mathbf{D}, \quad \mathbf{W}_\theta = [\mathbf{V}^{-1}][\mathbf{W} - \Gamma], \quad (3.6)$$

em que

$$\Gamma = \mathbf{D}^T \mathbf{V}^{-1} \mathbf{S} \mathbf{V}^{-1} \mathbf{D}. \quad (3.7)$$

Lema 3.2.1. *Para os MNLFE, a função escore e a matriz de informação relacionadas com o logaritmo da função de verossimilhança são dadas, respectivamente, por*

$$l'(\beta) = \phi \mathbf{D}^T(\beta) \mathbf{V}^{-1}(\beta) (\mathbf{y} - \mu(\beta)) = \phi \mathbf{D}_\theta^T(\beta) (\mathbf{y} - \mu(\beta)), \quad (3.8)$$

$$\begin{aligned} -l''(\beta) &= \phi \mathbf{D}^T \mathbf{V}^{-1} \mathbf{D} - \phi [\mathbf{e}^T \mathbf{V}^{-1}] [\mathbf{W} - \Gamma] \\ &= \phi \mathbf{D}_\theta^T \mathbf{V} \mathbf{D}_\theta - \phi [\mathbf{e}^T] [\mathbf{W}_\theta]. \end{aligned} \quad (3.9)$$

Além das suposições já enunciadas, são necessárias mais algumas.

Suposições B

O verdadeiro parâmetro de estimação de β a ser calculado é um ponto interno de B (subconjunto de \mathbb{R}^p , $p < n$). Os EMV $\hat{\beta}$ e $\hat{\phi}$ de β e ϕ existem e são únicos em B e Φ , respectivamente.

Unificando as suposições A e B, temos as seguintes conclusões:

(a) O EMV $\hat{\beta}$ satisfaz

$$\mathbf{D}_{\theta}^T(\hat{\beta})(\mathbf{Y} - \mu(\hat{\beta})) = \mathbf{D}^T(\hat{\beta})\mathbf{V}^{-1}(\hat{\beta})(\mathbf{Y} - \mu(\hat{\beta})) = 0. \quad (3.10)$$

Note que a equação não envolve nenhum $\hat{\phi}$, ou seja, a estimativa $\hat{\beta}$ não depende de $\hat{\phi}$.

(b) A matriz de informação para β em MNLFE é

$$\mathbf{J}_{\beta}(\mathbf{Y}) \equiv \mathbf{J}(\beta) = \phi \mathbf{D}^T \mathbf{V}^{-1} \mathbf{D} = \phi \mathbf{D}_{\theta}^T \mathbf{V} \mathbf{D}_{\theta}. \quad (3.11)$$

(c) O parâmetro de dispersão ϕ é ortogonal a β (Cox e Reid, 1987), isto é $E\left(\frac{-\partial^2 l}{\partial \beta \partial \phi}\right) = 0$.

Considere o EMV do parâmetro de dispersão ϕ , $\hat{\phi}$. Há uma relação entre $\hat{\phi}$ e o desvio. De fato, segue de (2.25) que

$$l'_{\phi} = \frac{\partial l}{\partial \phi} = -\frac{1}{2} D'(\mathbf{y}, \mu) - \frac{1}{2} \sum_{i=1}^n s'(y_i, \phi).$$

Assim, $\hat{\phi}$ satisfaz

$$\sum_{i=1}^n s'(y_i, \hat{\phi}) = -D'(\mathbf{y}, \hat{\mu}) \quad (3.12)$$

e pode-se obter $\hat{\phi}$ como função de $\hat{\beta}$. Em particular, para o modelo (3.1) com (2.6), $\hat{\phi}$ satisfaz

$$\begin{aligned} -s'(\hat{\phi}) &= n^{-1} D'(\mathbf{Y}, \hat{\mu}) = n^{-1} D(\mathbf{Y}, \hat{\mu}) + k \\ &= \frac{1}{n} \sum_{i=1}^n [d_i(y_i, \hat{\mu}_i) + k]. \end{aligned} \quad (3.13)$$

A matriz de informação observada e a matriz de informação de Fisher para ϕ em (3.1) com (2.6), pode ser representada por

$$-l''_{\phi\phi} = \mathbf{J}_{\phi}(\mathbf{Y}) = \frac{1}{2} n s''(\phi). \quad (3.14)$$

Depois de calcular $\hat{\beta}$, pode-se encontrar $\hat{\mu} = \mu(\hat{\beta})$, $\hat{\theta} = \theta(\hat{\beta})$ e o desvio $D(\mathbf{Y}, \hat{\mu})$. Então $\hat{\phi}$

pode ser calculado de (3.12) ou (3.13).

Exemplo (Modelos não lineares Normal, Normal Inversa e Gama).

Através da Tabela 2.1, para o modelo normal e modelo normal inverso, tem-se que $s(\phi) = -\log \phi$, $s'(\phi) = -\phi^{-1}$ e $k = 0$. Conseqüentemente, de (3.13), tem-se

$$\hat{\phi}^{-1} = \hat{\sigma}^2 = n^{-1}D(\mathbf{Y}, \hat{\mu}) = \frac{1}{n} \sum_{i=1}^n d_i(y_i, \hat{\mu}_i).$$

Em particular, para o modelo normal inverso,

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n [(y_i - \hat{\mu}_i)^2 / \hat{\mu}_i^2 y_i].$$

No caso do modelo gama, $s(\phi) = -2[\phi \log \phi - \log \Gamma(\phi)]$, $s'(\phi) = -2[1 + \log \phi - \Psi(\phi)]$ e $k = 2$. Conseqüentemente, de (3.13), temos

$$\log \hat{\phi} - \Psi(\hat{\phi}) = \frac{1}{2n} D(\mathbf{Y}, \hat{\mu}),$$

em que $\Psi(\phi)$ é a função digama. Uma aproximação de $\hat{\phi}$ é dada por

$$\hat{\phi} = \frac{n\{1 + [1 + 2D(\mathbf{Y}, \hat{\mu})/3n]^{1/2}\}}{2D(\mathbf{Y}, \hat{\mu})}$$

(Cordeiro e McCullagh, 1991).

Existem outros métodos para estimar o parâmetro de dispersão $\sigma^2 = \phi^{-1}$. O resumo seguinte é de Jorgensen (1987):

(a) $\hat{\sigma}_1^2 = (n - p)^{-1}D(\mathbf{y}, \hat{\mu})$ que é um estimador assintótico não viciado de σ^2 .

(b) $\hat{\sigma}_2^2 = (n - p)^{-1}\mathbf{X}^2$, em que $\mathbf{X} = (\mathbf{y} - \hat{\mu})^T \mathbf{V}(\hat{\mu})^{-1}(\mathbf{y} - \hat{\mu})$ é a estatística de Pearson generalizada. Este é de fato um estimador de momento.

(c) $\hat{\sigma}_3^2$ que maximiza a seguinte modificação na verossimilhança perfilada para o parâmetro σ^2 (Barndorff-Nilsen, 1983): $L^0(\sigma^2) = \sigma^p p(\mathbf{y}; \hat{\mu}, \sigma^2)$, em que $p(\mathbf{y}; \mu, \sigma^2)$ é a densidade da distribuição $ED(\mu, \sigma^2)$.

Para realizar teste de hipótese nos parâmetros do modelo, tem-se

(a) Seja o teste $H_0 : g(\mu) = f(\beta)$ para um valor fixo de ϕ , a estatística da razão de verossimilhanças é

$$LR(\hat{\mu}) = 2[l(\tilde{\mu}, \mathbf{y}) - l(\hat{\mu}, \mathbf{y})].$$

Sendo $\tilde{\mu} = \mathbf{y}$, $LR(\hat{\mu})$ é expresso por

$$LR(\hat{\mu}) = D^*(\mathbf{y}, \hat{\mu}) = \phi D(\mathbf{y}, \hat{\mu}).$$

Conseqüentemente, $D(\mathbf{y}, \hat{\mu})$ possui, assintoticamente, distribuição $\phi^{-1} \chi^2(n-p)$.

(b) Seja o teste $H_0 : \beta = \beta_0$ para um valor fixo de ϕ , a estatística da razão de verossimilhanças é

$$\begin{aligned} LR(\beta_0) &= 2[l(\mu(\hat{\beta}), \mathbf{y}) - l(\mu(\beta_0), \mathbf{y})] \\ &\equiv 2[l(\hat{\beta}) - l(\beta_0)], \end{aligned} \quad (3.15)$$

que, assintoticamente, tem distribuição $\chi^2(p)$ para valor fixo de ϕ . É fácil verificar que

$$\begin{aligned} LR(\beta_0) &= 2[l(\tilde{\mu}, \mathbf{y}) - l(\mu(\beta_0), \mathbf{y})] - 2[l(\tilde{\mu}, \mathbf{y}) - l(\mu(\hat{\beta}), \mathbf{y})] \\ &= \phi D(\mathbf{y}, \mu(\beta_0)) - \phi D(\mathbf{y}, \mu(\hat{\beta})). \end{aligned}$$

(c) Seja o teste $H_0 : \beta_2 = \beta_{20}$ para um valor fixo de ϕ , em que β_2 são os p_2 últimos parâmetros de β , a estatística da razão de verossimilhanças é

$$LR_s(\beta_{20}) = 2[l(\hat{\beta}) - l(\tilde{\beta}_0)],$$

em que $\beta = (\beta_1^T, \beta_2^T)^T$, $\tilde{\beta}_0 = (\tilde{\beta}_1^T(\beta_{20}), \beta_{20}^T)^T$ e $\tilde{\beta}_1(\beta_{20})$ maximiza $l(\beta_1, \beta_{20})$ para o valor fixo β_{20} . Então, $LR_s(\beta)$ tem, assintoticamente, distribuição $\chi^2(p_1)$ para ϕ fixo, em que $p_1 = p - p_2$ é a dimensão de β_1 (Cox e Hinkley, 1974).

A equação anterior serve para qualquer valor fixo de β_{20} . Em geral, introduzimos a verossimilhança perfilada de \mathbf{y} para o subconjunto de parâmetros β_2 . A verossimilhança perfilada é freqüentemente utilizada sendo definida como

$$l_p(\beta_2) = l(\tilde{\beta}_1(\beta_2), \beta_2),$$

em que $\tilde{\beta}_1(\beta_2)$ maximiza $l(\beta_1, \beta_2)$ para cada valor de β_2 . Seja $\hat{\beta} = (\hat{\beta}_1^T, \hat{\beta}_2^T)^T$, é fácil verificar

que

$$\tilde{\beta}_1(\hat{\beta}_2) = \hat{\beta}_1 \text{ e } l_p(\hat{\beta}_2) = l(\hat{\beta}_1, \hat{\beta}_2) = l(\hat{\beta}).$$

Assim, $LR_s(\beta_{20})$ pode ser expresso por

$$LR_s(\beta_{20}) = 2[l_p(\hat{\beta}_2) - l_p(\beta_{20})].$$

Esta equação serve para qualquer $\beta_2 = \beta_{20}$, assim podemos escrever

$$LR_s(\beta_2) = 2[l_p(\hat{\beta}_2) - l_p(\beta_2)].$$

3.3 Estimação dos Parâmetros

Considere que os parâmetros β são ortogonais ao parâmetro de dispersão ϕ , podemos calcular $\hat{\beta}$ e $\hat{\phi}$ separadamente. Na prática, calculamos $\hat{\beta}$ usando o método iterativo de Newton baseado na $l'(\beta)$ e $l''(\beta)$ ou então em $\mathbf{J}_\beta(\mathbf{Y})$ visto em (3.8), (3.9) e (3.11). Após obter $\hat{\beta}$, $\hat{\phi}$ é encontrado através da equação (3.12) ou (3.13). Vamos começar introduzindo o método computacional usado para encontrar $\hat{\beta}$.

Considerando a aproximação de primeira ordem da Série de Taylor na equação de verossimilhança $l'(\hat{\beta}) = \mathbf{0}$, temos

$$\begin{aligned} l'(\hat{\beta}) &\approx l'(\beta_0) + l''(\beta_0)(\hat{\beta} - \beta_0) \approx \mathbf{0} \\ \hat{\beta} &\approx \beta_0 + \{-l''(\beta_0)\}^{-1}l'(\beta_0) \end{aligned}$$

Então, o processo iterativo do método de Gauss-Newton pode ser expresso por

$$\beta^{(i+1)} = \beta^{(i)} + \{-l''(\beta^{(i)})\}^{-1}l'(\beta^{(i)}), \quad i = 0, 1, 2, \dots$$

em que $\beta^{(i)}$ é o valor de β na i -ésima iteração. Na equação de iteração, a informação observada $-l''(\beta)$ é substituída pela informação esperada $\mathbf{J}_\beta(\mathbf{Y}) = E(-l''(\beta)) = \phi \mathbf{D}^T \mathbf{V}^{-1} \mathbf{D}$. A substituição é conveniente e aceitável (Jennrich, 1969; Seber e Wild, 1989). Então, o processo de iteração do método Gauss-Newton para encontrar β é dado por

$$\beta^{(i+1)} = \beta^{(i)} + [(\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-1} \mathbf{e}]^{(i)}, \quad (3.16)$$

em que $\{\cdot\}^{(i)}$ é a potência da i -ésima iteração, isto é, \mathbf{D} , \mathbf{V} e $\mathbf{e} = \mathbf{y} - \mu(\beta)$ são avaliadores para

$\beta^{(i+1)} = \beta^{(i)}$. Esta equação de iteração pode ser reescrita como

$$\beta^{(i+1)} = [(\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-1} \mathbf{z}]^{(i)}, \quad (3.17)$$

$$\mathbf{z} = \mathbf{D}\beta + \mathbf{e}. \quad (3.18)$$

A equação (3.17) mostra que o processo de iteração pode ser considerado na forma dos estimadores de mínimos quadrados generalizados. Em geral, segundo a regressão não-linear

$$\mathbf{z} = \mu(\beta) + \delta, \quad \delta \sim (\mathbf{0}, \mathbf{V}),$$

em que $E(\delta) = 0$, $Var(\delta) = \mathbf{V}$.

Em geral, a convergência na iteração (3.16) é rapidamente alcançada, mas isso depende fortemente da escolha do valor inicial $\beta^{(0)}$. Jennrich (1969) apresentou um conjunto de condições de regularidade que garante ao processo iterativo de estimação por mínimos quadrados não-linear de Gauss-Newton, estabilidade numérica assintótica (isto é, $\beta^{(i)} \rightarrow \hat{\beta}$ quando $i \rightarrow \infty$), quando o valor inicial $\beta^{(0)}$ está próximo do valor de $\beta_{(0)}$.

Teorema 3.1

Para o modelo (3.2), se as suposições A e B forem satisfeitas, então existe um conjunto $N(\delta_0)$ de β_0 com raio δ_0 e valores inteiros $n_0(\mathbf{Y})$ para quase todo \mathbf{Y} , tal que $\beta^{(i)} \rightarrow \hat{\beta}_n(\mathbf{Y})$ quando $i \rightarrow +\infty$, para qualquer $n \geq n_0(\mathbf{Y})$ e $\beta^{(0)} \in N(\delta_0)$, em que β_0 é o valor verdadeiro do estimador de β , $\hat{\beta}_n(\mathbf{Y})$ é o estimador de máxima verossimilhança de β_0 baseado em $\mathbf{Y} = (y_1, \dots, y_n)^T$ e $\beta^{(0)}$ é o valor inicial do processo de iteração (3.16).

Nestes termos, assumimos que $\beta^{(i)}$ em (3.16) converge para $\hat{\beta}$. Então, segue de (3.17) e (3.18) que

$$\hat{\beta} = [(\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-1} \mathbf{z}]_{\hat{\beta}}, \quad (3.19)$$

$$\mathbf{z} = \mathbf{D}\hat{\beta} + \hat{\mathbf{e}}, \quad (3.20)$$

em que \mathbf{D}, \mathbf{V} e \mathbf{z} são avaliados em $\hat{\beta}$, e $\hat{\mathbf{e}} = \mathbf{e}(\hat{\beta}) = \mathbf{y} - \mu(\hat{\beta})$. Estas equações são bastante utilizadas em aplicações. Uma vez estabelecida a equação (3.19), muitos resultados importantes são conseqüências imediatas (Andersen, 1992). Em (3.19), os elementos de \mathbf{D} são considerados "variáveis explanatórias" e \mathbf{z} pode ser visto como "resposta", então $\hat{\beta}$ pode ser considerado como "estimador de mínimos quadrados generalizados" usando como pesos os elementos de \mathbf{V}^{-1} . Em

outras palavras, $\hat{\beta}$ é considerado o estimador de mínimos quadrados generalizado do seguinte modelo linear

$$\mathbf{z} = \mathbf{D}\beta + \mathbf{e}, \quad \mathbf{e} \sim (\mathbf{0}, \mathbf{V}); \quad (3.21)$$

ou o estimador de mínimos quadrados ordinário do seguinte modelo linear

$$\mathbf{V}^{-\frac{1}{2}}\mathbf{z} = \mathbf{V}^{-\frac{1}{2}}\mathbf{D}\beta + \varepsilon, \quad \varepsilon \sim (0, \mathbf{I}_n), \quad (3.22)$$

em que \mathbf{I}_n é a matriz identidade.

Como exemplo, definimos o resíduo de Pearson da seguinte forma (Andersen, 1992)

$$\mathbf{rp} = \mathbf{V}^{-\frac{1}{2}}\mathbf{z} - \mathbf{V}^{-\frac{1}{2}}\mathbf{D}\hat{\beta} = \mathbf{V}^{-\frac{1}{2}}\hat{\mathbf{e}}.$$

Consequentemente, temos

$$\mathbf{rp} \equiv (rp_i) = \mathbf{V}^{-\frac{1}{2}}(\hat{\beta})(\mathbf{y} - \mu(\hat{\beta})),$$

$$rp_i = \mathbf{V}_i^{-\frac{1}{2}}(\hat{\beta})(\mathbf{y}_i - \mu_i(\hat{\beta})). \quad (3.23)$$

Baseado em (3.21) e (3.22), verifica-se o problema da análise de diagnóstico para os modelos não-linear da família exponencial.

Lema 3.3.1. *Para qualquer β que satisfaça $l'(\beta) \neq 0$, existe um valor $\lambda^* > 0$, tal que $l(\beta + \lambda\mathbf{G}(\beta)) > l(\beta)$ para $0 < \lambda < \lambda^*$, em que $\mathbf{G}(\beta)$ é o segundo termo da equação (3.16), ou seja*

$$\mathbf{G}(\beta) = [\mathbf{D}^T(\beta)\mathbf{V}^{-1}(\beta)\mathbf{D}(\beta)]^{-1}\mathbf{D}^T(\beta)\mathbf{V}^{-1}(\beta)(\mathbf{y} - \mu(\beta)).$$

Através deste lema, resumimos o método de iteração de Gauss-Newton nos seguintes procedimentos.

(a) Escolhe um valor inicial β_0 e calcula $\mathbf{G}_0 = \mathbf{G}(\beta_0)$. Encontra um $0 < \lambda_0 < 1$ tal que

$$l(\beta_0 + \lambda_0\mathbf{G}_0) > l(\beta_0).$$

(b) Seja $\beta_1 = \beta_0 + \lambda_0\mathbf{G}_0$. Calcule $\mathbf{G}_1 = \mathbf{G}(\beta_1)$ e encontre um $0 < \lambda_1 < 1$, tal que

$$l(\beta_1 + \lambda_1\mathbf{G}_1) > l(\beta_1).$$

(c) Seja $\beta_2 = \beta_1 + \lambda_1\mathbf{G}_1, \dots$

É fácil verificar que neste procedimento, temos $l(\beta_{i+1}) > l(\beta_i)$ para $i = 1, 2, \dots$, e ainda temos que $l(\beta_i) \rightarrow l(\hat{\beta})$ sobre determinadas condições de regularidade. Há várias formas de escolher o valor de λ_i a cada iteração. Alguns valores simples sugeridos para os λ s, são $\lambda = 2^{-1}, 2^{-2}, 2^{-3}, \dots$ (Gallant, 1987).

4 Análise de Diagnóstico dos Modelos Não-Lineares da Família Exponencial

4.1 Introdução

O MNLFE apresentado em (2.1) e (2.2) pode ser denotado por

$$g(\mu_i) = f(x_i; \beta) \equiv f_i(\beta), \quad i = 1, 2, \dots, n, \quad (4.1)$$

em que $E(y_i) = \mu_i$ e $y_i \sim ED(\mu_i, \sigma^2)$ é implícito. Uma aproximação fundamental do diagnóstico de influência baseia-se na comparação dos parâmetros estimados $\hat{\beta}$ e $\hat{\sigma}^2$ com os parâmetros estimados $\hat{\beta}_{(i)}$ e $\hat{\sigma}_{(i)}^2$ correspondente ao ajuste do modelo (4.1) com o i -ésimo caso deletado, também chamado de modelo de casos deletados (MCD), expresso por

$$g(\mu_j) = f(x_j; \beta), \quad j = 1, 2, \dots, n, \quad j \neq i. \quad (4.2)$$

Para o cálculo de $\hat{\beta}_{(i)}$ e $\hat{\sigma}_{(i)}^2$, considera-se o método introduzido na seção 3.3 executando o programa com a i -ésima observação deletada, $i = 1, 2, \dots, n$. Para o diagnóstico proposto, usa-se a aproximação de primeira ordem $\hat{\beta}_{(i)}^I$ e $\hat{\sigma}_{(i)}^I$ para $\hat{\beta}_{(i)}$ e $\hat{\sigma}_{(i)}$ apresentada a seguir.

Lema 4.1.1. *A aproximação de primeira ordem $\hat{\beta}_{(i)}^I$ de β no MCD (4.2) é expresso por*

$$\hat{\beta}_{(i)}^I = \hat{\beta} - \left[\frac{(\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{V}_i^{-\frac{1}{2}} \mathbf{d}_i \mathbf{r} \mathbf{p}_i}{1 - h_{ii}} \right]_{\hat{\beta}}, \quad (4.3)$$

em que

$$\mathbf{r} \mathbf{p}_i = [\mathbf{V}_i^{-\frac{1}{2}} (y_i - \mu_i)]_{\hat{\beta}}$$

é o resíduo de Pearson introduzido em (3.23), h_{ii} é o i -ésimo elemento da diagonal da matriz

$\mathbf{H} = \mathbf{V}^{-\frac{1}{2}} \mathbf{D} (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-\frac{1}{2}}$ e \mathbf{d}_i^T é a i -ésima coluna de $\mathbf{D} = \partial \mu / \partial \beta^T$.

Após obter $\hat{\beta}_{(i)}^I$, $\hat{\sigma}_{(i)}^I$ pode ser calculado a partir de (3.12) e (3.13). Para encontrar os pontos influentes, calcula-se uma distância entre $(\hat{\beta}, \hat{\sigma})$ e $(\hat{\beta}_{(i)}, \hat{\sigma}_{(i)})$ ou $\hat{\beta}$ e $\hat{\beta}_{(i)}$. Este último é usado na prática pois se um caso é influenciado por $\hat{\beta}$, então este caso também é influenciado por $(\hat{\beta}, \hat{\sigma})$.

Serão introduzidos dois tipos de distâncias as quais são bastante usadas em diagnóstico de influência e também podem ser usadas para MNLFE.

(a) Distância Generalizada de Cook

Esta distância é uma norma de $\hat{\beta} - \hat{\beta}_{(i)}$ com respeito à matriz peso $M > 0$ e definida como

$$C_i \approx \left\| \hat{\beta} - \hat{\beta}_{(i)} \right\|_M^2 = (\hat{\beta} - \hat{\beta}_{(i)})^T \mathbf{M} (\hat{\beta} - \hat{\beta}_{(i)}).$$

É natural escolher $M = \mathbf{J}(\beta) = \phi \mathbf{D}^T \mathbf{V}^{-1} \mathbf{D}$, a matriz de informação de \mathbf{Y} para β , que resulta em

$$C_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})^T (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})_{\hat{\beta}} (\hat{\beta} - \hat{\beta}_{(i)})}{\hat{\sigma}^2}, \quad (4.4)$$

que pode ser chamado de distância da informação. Através do Lema 4.1.1 verifica-se que a aproximação de primeira ordem de C_i é

$$C_i^I = \frac{h_{ii}}{1 - h_{ii}} r_i^2, \quad \text{com } r_i = \frac{r p_i}{\hat{\sigma} \sqrt{1 - h_{ii}}}, \quad (4.5)$$

em que r_i é chamado de resíduo estudentizado ou resíduo estudentizado padronizado (McCullagh e Nelder, 1989).

(b) Distância de Verossimilhança

Esta é definida como (Cook e Weisberg, 1982, p.183)

$$LD_i(\beta) = 2[l(\hat{\beta}) - l(\hat{\beta}_{(i)})] \quad (4.6)$$

para β e definida por

$$\begin{aligned} LD_i(\beta_1 | \beta_2) &= 2[l_p(\hat{\beta}_1) - l_p(\hat{\beta}_{1(i)})] \\ &= 2[l(\hat{\beta}_1, \hat{\beta}_2) - l(\hat{\beta}_{1(i)}, \tilde{\beta}_2(\hat{\beta}_{1(i})))] \end{aligned} \quad (4.7)$$

para um subconjunto de parâmetros β_1 , em que $\hat{\beta}_{1(i)} = (\hat{\beta}_{1(i)}^T, \hat{\beta}_{2(i)}^T)^T$ e $\tilde{\beta}_2(\hat{\beta}_{1(i)})$ é o estimador de

máxima verossimilhança de β_2 para um valor fixo de β_1 . Calcula-se LD_i diretamente baseando-se em $\hat{\beta}$, $\hat{\beta}_{(i)}$ e na log-verossimilhança $l(\beta)$. A aproximação de primeira ordem para a expressão (4.6) é dada por

$$LD_i = 2 \left[l'^T(\hat{\beta})(\hat{\beta} - \hat{\beta}_{(i)}) + \frac{1}{2}(\hat{\beta} - \hat{\beta}_{(i)})^T (-l''(\hat{\beta}))(\hat{\beta} - \hat{\beta}_{(i)}) \right].$$

Considerando $l'(\hat{\beta}) = \mathbf{0}$, $(-l''(\hat{\beta})) \approx \mathbf{J}(\hat{\beta})$, tem-se que

$$LD_i^l(\beta) = C_i \approx C_i^l. \quad (4.8)$$

Mais adiante será apresentado um resumo de algumas estatísticas de diagnósticos utilizadas em MCD e que podem ser usados para os modelos não-lineares da família exponencial.

4.2 Análise de diagnóstico de modelos

4.2.1 Diagnóstico do modelo

O modelo de diagnóstico é a base para construção de estatísticas de diagnóstico eficientes. O MCD visto em (4.2) é o mais importante na prática, por ser direto e de fácil implementação. Outro método de diagnóstico de modelo bastante usado é o denominado mean-shift outlier model (MSOM) (Cook e Weisberg, 1982). Para o MNLFE, o MSOM pode ser representado por

$$\begin{cases} g(\mu_j) = f_j(\beta), & j = 1, \dots, n; j \neq i, \\ g(\mu_i) = f_i(\beta) + \gamma, \end{cases}$$

ou usando notação vetorial

$$g(\mu) = f(\beta) + \gamma c_i, \quad (4.9)$$

em que $g(\mu) = g(\mu_i)$ e $f(\beta) = f_i(\beta)$ são vetores de dimensão n , c_i é um vetor de dimensão n contendo da primeira até a i -ésima posição o valor 1 e nas demais o valor zero e γ é um parâmetro extra que representa a presença de valores discrepantes. É fácil verificar que γ diferente de zero implica que a i -ésima observação pode ser um valor discrepante, pois a observação (x_i, y_i) não pertence ao modelo (4.1). Este modelo é normalmente mais fácil de formular que o MCD. Para detectar valores discrepantes utilizando (4.9) pode-se estimar o parâmetro γ ou fazer o teste de hipótese $H_0 : \gamma = 0$. O estimador de máxima verossimilhança de β , γ e σ^2 em (4.9) são denotados por $\hat{\beta}_{mi}$, $\hat{\gamma}_{mi}$ e $\hat{\sigma}_{mi}^2$, respectivamente.

O terceiro modelo de diagnóstico bastante utilizado é o modelo casos-pesos (MCP). Para

este modelo, as estimativas $\hat{\beta}_w$ e $\hat{\sigma}_w^2$ maximizam

$$l_w(\beta, \sigma) = \sum_{j=1}^n w_j l_j(\beta, \sigma)$$

em que $l_j(\beta, \sigma) = \log p(y_j; \beta, \sigma)$ e w_j é o peso usado para indicar a importância do efeito de cada observação no ajuste (Belsley, Kuh e Welsch, 1980; Pregibon, 1981). Um caso especial do MCP é quando $w_i = w$ e $w_j = 1 (j \neq i)$, então a expressão passa a ser

$$l_w(\beta, \sigma) = \sum_{j \neq i} l_j(\beta, \sigma) + w l_i(\beta, \sigma). \quad (4.10)$$

Os estimadores para (4.10) são denotados por $\hat{\beta}_{wi}$ e $\hat{\sigma}_{wi}^2$. É fácil verificar que $\hat{\beta}_{wi} = \hat{\beta}$, $\hat{\sigma}_{wi}^2 = \hat{\sigma}^2$ quando $w = 1$ e, $\hat{\beta}_{wi} \rightarrow \hat{\beta}$, $\hat{\sigma}_{wi}^2 \rightarrow \hat{\sigma}^2$ quando $w \rightarrow 0$.

4.2.2 Coeficiente de regressão

Para os MNLFE (4.1) o logaritmo da função de verossimilhança pode ser denotada por (3.5) e representado por

$$l(\beta, \sigma) = -\frac{1}{2} \sigma^{-2} D(\mathbf{y}, \mu(\beta)) - \frac{1}{2} s(\mathbf{y}, \sigma^{-2}) - \frac{1}{2} \sigma^{-2} D'(\mathbf{y}, \mathbf{y}), \text{ em que } g(\mu) = f(\beta).$$

É fácil verificar que o estimador de máxima verossimilhança $\hat{\beta}$ minimiza o desvio $D(\mathbf{y}, \mu(\beta))$ o qual não inclui o parâmetro de dispersão e pode ser expresso por

$$D(\beta) \equiv D(\mathbf{y}, \mu(\beta)) = \sum_{j=1}^n d_j(y_j, \mu_j(\beta)), \text{ com } \mu_j = g^{-1}(f_j(\beta)), \quad (4.11)$$

em que

$$d_j(y_j, \mu_j(\beta)) = -2[y_j \theta_j - b(\theta_j) - c(y_j)] + 2[y_j \theta_j - b(\theta_j) - c(y_j)]_{\mu_j = y_j}.$$

Da equação apresentada acima, tem-se o seguinte teorema.

Teorema 4.1

Para os MNLFE (4.1), se assumimos as **Suposições A e B** (apresentadas na seção 3.1 e 3.2), sempre tem-se que $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$ (Wei e Shi, 1994).

Corolário 1

Para o modelo de regressão não linear normal e GLM, se as suposições **Suposições A e B** forem satisfeitas, sempre tem-se que $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$.

Corolário 2

Se o MSOM é modificado para a forma geral

$$\begin{cases} g(\mu_j) = f_j(\beta), & j = 1, \dots, n; j \neq i, \\ g(\mu_i) = f(x_i; \beta, \gamma) \equiv f_i(\beta, \gamma), \end{cases} \quad (4.12)$$

em que γ é um vetor de parâmetros e $\partial f_i / \partial \gamma \neq 0$ para todo f_i , então tem-se que $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$ (note que se $\partial f_i / \partial \gamma = 0$ está em alguma região, então $f_i(\beta, \gamma)$ não depende de γ).

Corolário 3

Para o MCP (4.10), tem-se que $\hat{\beta}_{mi} = \hat{\beta}_{wi}$ quando $w \rightarrow 0$.

O Teorema 4.1 e os corolários apresentados podem ser estendidos para o caso múltiplo, em que k observações são deletadas, substituídas ou proporcionadas. Estes modelos são, respectivamente, representados por

$$\begin{aligned} MCD: & \quad g(\mu_j) = f_j(\beta), \quad j \in I, \\ MSOM: & \quad \begin{cases} g(\mu_j) = f_j(\beta), & j \in I \\ g(\mu_i) = f_i(\beta) + \gamma, & i \in I \end{cases} \end{aligned}$$

$$MCP: \quad l_w(\beta, \sigma) = \sum_{j \in I} l_j(\beta, \sigma) + \sum_{i \in I} w_i l_i(\beta, \sigma),$$

em que $I = (i_1, \dots, i_k)$ é um subconjunto de índices para $1, \dots, n$. O estimador para estes modelos são denotados, respectivamente, por $\hat{\beta}_{(I)}$, $\hat{\beta}_{mI}$ e $\hat{\beta}_{wI}$. Através de derivação semelhante a apresentada no teorema 4.1, tem-se o seguinte corolário.

Corolário 4

Para o MCD, MSOM e MCP múltiplos, sempre tem-se que $\hat{\beta}_{(I)} = \hat{\beta}_{mI} = \hat{\beta}_{wI}$ (quando $w_i \rightarrow 0, i \in I$).

Corolário 5

Suponha-se que $\hat{\beta}$ e $\hat{\sigma}$ minimizam a função objetivo

$$Q(\beta, \sigma) = \sum_{i=1}^n p_i(f_i(\beta, \sigma)),$$

em que $p_i(\cdot, \cdot)$ e $f_i(\cdot)$ são funções conhecidas. Se $p_i(f_i(\beta, \sigma)) = a(\sigma)\xi_i(f_i(\beta)) + b_i(\sigma)$ e $\xi_i(\cdot)$ é uma função conhecida, então tem-se que $\hat{\beta}_{mi} = \hat{\beta}_{(i)}$, em que $\hat{\beta}_{mi}$ e $\hat{\beta}_{(i)}$ são estimadores de β , respectivamente, correspondente ao MSOM e MCD expressos por:

$$MSOM: Q_{mi}(\beta, \sigma) = \sum_{j \neq i} p_j(f_j(\beta), \sigma) + p_i(f_i(\beta, \gamma), \sigma),$$

$$MCD: Q_{(i)}(\beta, \sigma) = \sum_{j \neq i} p_j(f_j(\beta), \sigma).$$

4.2.3 Desvio e parâmetro de dispersão

Normalmente não se tem $\hat{\sigma}_{mi}^2 = \hat{\sigma}_{(i)}^2$ para o estimador de máxima verossimilhança, até mesmo no caso do modelo de regressão linear normal. De fato, para o modelo de regressão linear normal, temos que $\hat{\sigma}_{mi}^2 = n^{-1}SQR$ e $\hat{\sigma}_{(i)}^2 = (n-1)^{-1}SQR_{(i)}$, em que SQR é a soma de quadrado residual associada ao MCD. Então tem-se que $\hat{\sigma}_{(i)}^2 = (n-1)^{-1}n\hat{\sigma}_{mi}^2 \approx \hat{\sigma}_{mi}^2$. Similarmente, para os MNLFE, não é possível adquirir $\hat{\sigma}_{mi}^2 = \hat{\sigma}_{(i)}^2$, porém, tem-se $\hat{\sigma}_{mi}^2 \approx \hat{\sigma}_{(i)}^2$, pois, os desvios com os quais possuem conexão direta com os estimadores de máxima verossimilhança σ^2 são todos iguais para o MCD e MSOM.

Teorema 4.2

Considerando as condições citadas no Teorema 4.1, sempre tem-se $D_{(i)}(\hat{\beta}_{(i)}) = D_{mi}(\hat{\beta}_{(i)}, \hat{\gamma}_{mi})$ em que $D_{(i)}(\beta)$ e $D_{mi}(\beta, \gamma)$ são, respectivamente, os desvios para MCD e MSOM.

Corolário 6

Os desvios para MCD e MSOM são iguais para o MLG e as somas residuais quadráticas do MCD e MSOM são iguais ao do modelo de regressão não-linear normal.

Teorema 4.3

Considerando as condições estabelecidas no Teorema 4.1 e sendo $s(y_i, \phi) = s(\phi) + t(y_i)$

para todo i , então tem-se

$$s'(\hat{\phi}_{(i)}) = (n-1)^{-1}ns'(\hat{\phi}_{mi}) + (n-1)^{-1}k, \quad (4.13)$$

em que $k = d'_i(y_i, y_i)$ é definida no Lema 2.4.2. Além disso, para o modelo não-linear normal e normal inversa, tem-se

$$\hat{\sigma}_{(i)}^2 = (n-1)^{-1}n\hat{\sigma}_{mi}^2 \approx \hat{\sigma}_{mi}^2. \quad (4.14)$$

4.2.4 Estatística escore de outlier

Nesta seção será mostrada a estatística escore para detectar valores discrepantes baseada no MSOM. De fato, para o modelo (4.9), pode-se fazer um teste de hipótese:

$$H_0 : \gamma = 0; \quad H_1 : \gamma \neq 0.$$

Se H_0 é rejeitada, então o i -ésimo caso pode ser um possível valor discrepante, pois, o caso pode não vir do modelo original (4.1).

Teorema 4.4

Para o MSOM, a estatística escore para o teste de hipótese $H_0 : \gamma = 0$ é

$$SC_i = \frac{rp_i^2}{\hat{\sigma}^2(1-h_{ii})} = r_i^2, \quad (4.15)$$

em que h_{ii} , rp_i e r_i são vistos em (4.3) e (4.5).

Considerando-se o modelo

$$g(\mu) = f(\beta, \gamma), \quad (4.16)$$

em que γ é um vetor de parâmetros. Assume-se que existe $\gamma = \gamma_0$ tal que $f(\beta, \gamma_0) = f(\beta)$. Neste caso, $g(\mu_i) = f(x_i; \beta, \gamma)$, $i = 1, \dots, n$.

Teorema 4.5

Para o modelo (4.16), a estatística escore sobre H_0 é

$$SC = \hat{\sigma}^2 \hat{\mathbf{e}}^T \hat{\mathbf{V}}^{-\frac{1}{2}} (\mathbf{H}^+ - \mathbf{H}) \hat{\mathbf{V}}^{-\frac{1}{2}} \hat{\mathbf{e}}, \quad (4.17)$$

em que \mathbf{H} foi vista no lema 4.1.1, \mathbf{H}^+ é a matriz projeção de $(\mathbf{V}^{-\frac{1}{2}}\mathbf{D}_\beta, \mathbf{V}^{-\frac{1}{2}}\mathbf{D}_\gamma)_{\hat{\beta}}$ e $\mathbf{D}_\beta = \partial\mu/\partial\beta^T$, $\mathbf{D}_\gamma = \partial\mu/\partial\gamma^T$.

Este resultado pode ser estendido para modelos em geral (Davison e Tsai, 1992). A estatística escore (4.17) pode ser vista como uma mudança na soma de quadrado dos resíduos quando o parâmetro γ é adicionado ao modelo. De fato (4.17) pode ser expresso como

$$SC = \hat{\sigma}^{-2}(SQR - SQR^+)$$

em que

$$\begin{aligned} SQR &= \hat{\mathbf{e}}^T \hat{\mathbf{V}}^{-\frac{1}{2}} (\mathbf{I}_n - \mathbf{H}) \hat{\mathbf{V}}^{-\frac{1}{2}} \hat{\mathbf{e}} \\ SQR^+ &= \hat{\mathbf{e}}^T \hat{\mathbf{V}}^{-\frac{1}{2}} (\mathbf{I}_n - \mathbf{H}^+) \hat{\mathbf{V}}^{-\frac{1}{2}} \hat{\mathbf{e}} \end{aligned}$$

4.3 Diagnóstico de influência baseado em casos apagados

Foi introduzido a distância generalizada de Cook (C_i) e a distância de verossimilhança (LD_i) na Seção 4.1; e introduzida a estatística escore (SC_i) na Seção 4.2. Nesta seção, serão resumidas as estatísticas de diagnósticos aplicáveis aos modelos não-lineares da família exponencial. As derivações são baseadas nos modelos de casos deletados argumentado por Andersen (1992).

4.3.1 Diagnóstico baseado em mínimos quadrados ponderados

A idéia básica descrita aqui é primeiramente introduzida implicitamente por Pregibon (1981) para MLG e desenvolvido por Andersen (1992) o qual discute o diagnóstico em análise de dados categóricos e obtém várias estatísticas de diagnóstico.

Denota-se o modelo (4.1) por (A) que será comparado com o modelo linear (B) obtido na forma do procedimento de repetição do estimador de máxima verossimilhança de β para o modelo (A). Os modelos (A) e (B) são dados por

$$\begin{aligned} (A) \quad &g(\mu_i) = f(x_i, \beta), \quad y_i \sim ED(\mu_i, \sigma^2) \\ (B) \quad &\mathbf{V}^{-\frac{1}{2}}\mathbf{Z} = \mathbf{V}^{-\frac{1}{2}}\mathbf{D}\beta + \varepsilon, \quad \varepsilon \sim (\mathbf{0}, \mathbf{I}_n) \end{aligned}$$

em que \mathbf{Z} é dado pelas equações (3.20) a (3.22). Para derivar a estatística de diagnóstico para o modelo (A), enfatiza-se o fato de que os modelos (A) e (B) apresentam dois pontos em comuns.

(a) O i -ésimo caso de ambos os modelos corresponde a observação (x_i, y_i) (note que \mathbf{V} é a matriz diagonal).

(b) O estimador de máxima verossimilhança de β no modelo (A) é igual ao estimador de mínimos quadrados de β na forma obtida para o modelo (B).

Através das duas propriedades acima, tem-se o seguinte resultado: avaliar a influência da i -ésima observação (x_i, y_i) no estimador de máxima verossimilhança no modelo (A) é equivalente a avaliar a influência de (x_i, y_i) no estimador de mínimos quadrados no modelo linear (B). Então pode-se achar estatísticas de diagnóstico apropriadas para o modelo (A) baseando-se no modelo (B) que pode ser considerado na forma de um modelo de regressão linear com pesos. Seguem algumas estatísticas de diagnóstico normalmente utilizadas nos modelos do tipo (A), obtidas da mesma forma dos modelos de regressão lineares.

(a) Resíduo e resíduo estudentizado

$$\mathbf{rp} = (\mathbf{V}^{-\frac{1}{2}}\mathbf{Z} - \mathbf{V}^{-\frac{1}{2}}\mathbf{D}\beta)_{\hat{\beta}} = \hat{\mathbf{V}}^{-\frac{1}{2}}\hat{\mathbf{e}},$$

$$r_i = \frac{\mathbf{rp}_i}{\sigma\sqrt{1-h_{ii}}}.$$

(b) Matriz chapéu (Matriz de projeção)

$$\mathbf{H} = \{\mathbf{V}^{-\frac{1}{2}}\mathbf{D}(\mathbf{D}^T\mathbf{V}^{-1}\mathbf{D})^{-1}\mathbf{D}^T\mathbf{V}^{-\frac{1}{2}}\}_{\hat{\beta}}.$$

Caso $(\mathbf{D}^T\mathbf{V}^{-1}\mathbf{e})_{\hat{\beta}} = 0$, tem-se

$$\mathbf{H}\mathbf{rp} = \mathbf{0}, \quad (\mathbf{I}_n - \mathbf{H})\mathbf{rp} = \mathbf{rp}.$$

(c) Distância de Cook

$$C_i = \frac{(\hat{\beta} - \hat{\beta}_i)^T (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D}) (\hat{\beta} - \hat{\beta}_i)}{p\hat{\sigma}^2},$$

$$C_i^I = \frac{h_{ii}}{1-h_{ii}} \cdot \frac{r_i^2}{p}, \quad (4.18)$$

em que C_i^I é aproximação de primeira ordem de C_i .

(d) Estatística AP

$$AP_i = \left\{ \frac{|\mathbf{X}_{(i)}^{*T} \mathbf{X}_{(i)}^*|}{|\mathbf{X}^{*T} \mathbf{X}^*|} \right\}_{\hat{\beta}},$$

em que $|\cdot|$ denota o determinante, $\mathbf{X}^* = (\mathbf{V}^{-\frac{1}{2}}\mathbf{D}, \mathbf{V}^{-\frac{1}{2}}\mathbf{Z})$ e $\mathbf{X}_{(i)}^*$ é obtida através de \mathbf{X}^* com a i -ésima linha excluída.

Teorema 4.6

Seja h_{ii}^* o elemento da diagonal da matriz de projeção \mathbf{H}^* obtida através de \mathbf{X}^* , então tem-se que

$$AP_i = 1 - h_{ii}^* = 1 - h_{ii} - \frac{rp_i^2}{(\mathbf{rp}^T \mathbf{rp})}. \quad (4.19)$$

4.3.2 Diagnóstico baseado em desvio

Os diagnósticos relacionados para os desvios introduzidos aqui são totalmente similares aos introduzidos por Pregibon (1981) para o MLG.

(a) Desvio residual

$$rD_i = \{sinal(y_i - \hat{\mu}_i)\} d_i(y_i, \mu_i(\hat{\beta})). \quad (4.20)$$

Facilmente verifica-se que rD_i reflete a qualidade do ajuste para a i -ésima observação, utilizando-se o modelo reduzido (McCullagh e Nelder, 1989).

(b) Diferença de desvio

$$\Delta_i \mathbf{D} = \mathbf{D}(\hat{\beta}) - \mathbf{D}_{(i)}(\hat{\beta}_{(i)}), \quad (4.21)$$

em que $\mathbf{D}(\beta)$ e $\mathbf{D}_{(i)}(\beta)$ são expressos como

$$D(\beta) \equiv D(\mathbf{y}, \mu(\beta)) = \sum_{j=1}^n d_j(y_j, \mu_j(\beta)), \text{ em que } \mu_j = g^{-1}(f_j(\beta)),$$

e

$$\mathbf{D}_{(i)}(\beta) \equiv \sum_{j \neq i} d_j(y_j, \mu_j(\beta)), \text{ em que } \mu_j = g^{-1}(f_j(\beta)),$$

respectivamente.

$\Delta_i \mathbf{D}$ reflete a diferença do estimador de máxima verossimilhança de σ^2 no modelo original e no modelo com casos deletados. De fato, para o modelo linear e modelo não linear, tem-se $\mathbf{D}(\beta) = SQT$ e

$$\begin{aligned} \Delta_i \mathbf{D} &= SQR - SQR_{(i)} = n\hat{\sigma}^2 - (n-1)\hat{\sigma}_{(i)}^2 \\ &= (n-1)(\hat{\sigma}^2 - \hat{\sigma}_{(i)}^2) + \hat{\sigma}^2. \end{aligned} \quad (4.22)$$

De forma similar, considerando o caso em que $s(y_i, \phi) = s(\phi) + t(y_i)$ é satisfeito no modelo (4.1). Segue de (3.13) que $\hat{\phi}$ e $\hat{\phi}_{(i)}$ satisfazem, respectivamente,

$$\begin{aligned} -ns'(\hat{\phi}) &= \mathbf{D}(\beta) + nk, \\ -(n-1)s'(\hat{\phi}_{(i)}) &= \mathbf{D}_{(i)}(\beta_{(i)}) + (n-1)k, \end{aligned} \quad (4.23)$$

que conduzem a

$$\Delta_i \mathbf{D} = -(n-1)\{s'(\hat{\phi}) - s'(\hat{\phi}_{(i)})\} - \{s'(\hat{\phi}) + k\}.$$

Que também reflete a diferença entre $\hat{\phi}$ e $\hat{\phi}_{(i)}$.

Pregibon (1981) obtém a formula de $\Delta_i \mathbf{D}$ para MLG com ligação canônica. Pode-se estender este resultado para o caso geral dos modelos não-lineares da família exponencial.

Teorema 4.7

Para o modelo (4.1), a aproximação de primeira ordem de $\Delta_i \mathbf{D}$ pode ser expressa por

$$\Delta_i \mathbf{D}^I = d_i(y_i, \hat{\mu}_i) + h_{ii} r_i^2. \quad (4.24)$$

A tabela 4.1 resume o diagnóstico de influência introduzido na seção (4.1) a (4.3) para os modelos não-lineares da família exponencial. Há oito estatísticas de diagnóstico de modelos nesta tabela. Em resumo, estatísticas rp_i , p_i , rD_i , h_{ii} , h_{ii}^* , e AP_i que refletem a influência da i -ésima observação no modelo estimado; C_i e $LD_i(\beta)$ que medem a diferença entre $\hat{\beta}$ e $\hat{\beta}_{(i)}$; enquanto $\Delta_i \mathbf{D}$ que mede a diferença entre $\hat{\sigma}^2$ e $\hat{\sigma}_{(i)}^2$. Na prática, para o diagnóstico proposto, não há grandes diferenças entre essas medidas de influência.

Tabela 4.1: Resumo das estatísticas de diagnóstico para os modelos não-lineares da família exponencial

| Diagnósticos | Fórmula | Notas |
|---------------------------------|---|----------------|
| Resíduo de Pearson | $rp_i = \hat{\mathbf{V}}^{-\frac{1}{2}} \hat{\mathbf{e}}_i$ | Lema 4.1.1 |
| Resíduo Estudentizado | $r_i = \frac{rp_i}{\hat{\sigma} \sqrt{1-h_{ii}}}$ | (4.15) |
| Desvio entre residual | $r\mathbf{D}_i = \{sinal(y_i - \hat{\mu}_i)\} d_i(y_i, \mu_i(\hat{\boldsymbol{\beta}}))$ | (4.21) |
| Matriz chapéu | $\mathbf{H} = \mathbf{V}^{-\frac{1}{2}} \mathbf{D} (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-\frac{1}{2}}$ | Lema 4.1.1 |
| Elemento diagonal | $h_{ii} = \mathbf{V}_i^{-1} \mathbf{d}_i^T (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{d}_{ii}$ | |
| Estatística AP | $AP_i = 1 - h_{ii}^* = 1 - h_{ii} - \frac{rp_i^2}{(\mathbf{rp}^T \mathbf{rp})}$ | (4.19) |
| | $h_{ii}^* = h_{ii} + \frac{rp_i^2}{(\mathbf{rp}^T \mathbf{rp})}$ | |
| Distância de Cook | $C_i^I = \frac{h_{ii}}{1-h_{ii}} \frac{r_i^2}{p}$ | (4.4) e (4.18) |
| Distância entre verossimilhança | $LD_i^I(\boldsymbol{\beta}) = p C_i$ | (4.6) e (4.8) |
| Distância de desvios | $\Delta_i \mathbf{D} = d_i + h_{ii} r_i^2$ | Teorema 4.7 |

4.4 Análise de influência local

A influência local aproximada foi apresentada por Cook (1986) e desenvolvida mais adiante por vários autores (Thomas e Cook, 1989; Wei, Lu e Shi, 1991; Escobar e Meeker, 1992; Wu e Wan, 1994). Nesta seção será revista a idéia básica e fórmulas da influência local aproximada, e então serão aplicadas aos modelos não-lineares da família exponencial considerando uma perturbação aleatória sistemática.

4.4.1 Modelos perturbados

Seja $\alpha = (\boldsymbol{\beta}^T, \sigma^2)^T$ e $l(\alpha)$ a log-verossimilhança que correspondente a (4.1), chamado de modelo postulado. Seja w um vetor n -dimensional definido em um subconjunto aberto Ω em R^n , utilizado para colocar uma perturbação no modelo (Pregibon, 1981). Seja $l(\alpha|w)$ a log-verossimilhança correspondente à perturbação do modelo para um certo valor de w e $\hat{\alpha}_w = (\hat{\boldsymbol{\beta}}_w^T, \hat{\sigma}_w^2)^T$ o estimador de máxima verossimilhança sobre $l(\alpha|w)$. Assume-se que existe w_0 em w tal que $l(\alpha|w_0) = l(\alpha)$, $\hat{\alpha}_{w_0} = \hat{\alpha}$. As derivadas de $l(\alpha|w)$ com relação a α e w são assumidas de forma a existir até a terceira ordem. Nota-se que $\hat{\alpha}_w$ satisfaz

$$\left. \frac{\partial l(\alpha|w)}{\partial \alpha} \right|_{\alpha=\hat{\alpha}_w} = 0, \quad w \in \Omega. \quad (4.25)$$

Para avaliar a influência da perturbação de w no estimador de máxima verossimilhança $\hat{\alpha} = (\hat{\boldsymbol{\beta}}^T, \hat{\sigma}^2)^T$, a distância de verossimilhança é usualmente preferida (Pregibon, 1981; Cook e

Weisberg, 1982). Segundo Cook (1986), considera-se

$$LD(w) = 2\{l(\hat{\alpha}) - l(\hat{\alpha}_w)\} \quad (4.26)$$

que mede a distância entre $\hat{\alpha}$ e $\hat{\alpha}_w$ em termos da diferença de logaritmos de verossimilhanças. Se $LD(w) > \chi^2(p+1, 1-\gamma)$, então a perturbação w resulta em um $\hat{\alpha}_w$ que fica fora de região de probabilidade de perturbação nula para α (ao nível de $1-\gamma$). Assim, um grande $LD(w)$ indica grande influência de w no ajuste para o modelo postulado. Na prática, pode-se usar $LD(w) > \frac{1}{2}\chi^2(p+1, 0.5)$ como sinal de advertência à perturbação bastante influente (Escobar e Meeker, 1992). Para um modelo com perturbação específica (4.10), se $w \rightarrow 0$, então $\hat{\alpha}_{wi} \rightarrow \hat{\alpha}_{(i)}$ e (4.26) torna-se (4.6).

Para analisar o comportamento local de $LD(W)$, o método mais direto utilizado é a expansão em série de Taylor. É facilmente verificado de (4.26) que $LD(w_0) = 0$ e

$$\left\{ \frac{\partial LD(w)}{\partial w^T} \right\}_{w_0} = (-2) \left\{ \frac{\partial l(\alpha)}{\partial \alpha} \frac{\partial \hat{\alpha}_w}{\partial w^T} \right\}_{(w_0, \hat{\alpha})} = 0. \quad (4.27)$$

Agora considere

$$\mathbf{A}'' = \left\{ \frac{\partial^2 LD(w)}{\partial w \partial w^T} \right\}_{w_0} = -2\mathbf{F}'', \text{ em que } \mathbf{F}'' = \left\{ \frac{\partial^2 l(\hat{\alpha}_w)}{\partial w \partial w^T} \right\}_{w_0}$$

Então \mathbf{A}'' e $-\mathbf{F}''$ são positivas definidas e são chamadas de matrizes de influência. Mais adiante, $w = w_0 + h$ em que h é uma direção de perturbação em Ω e pode-se assumir $\|\mathbf{h}\|^2 = \mathbf{h}^T \mathbf{h} = 1$ sem perda de generalidade. Então, tem-se o seguinte teorema.

Teorema 4.8

A aproximação de segunda ordem da distância de verossimilhança $LD(w)$, vista em (4.26), pode ser expressa por

$$LD^{II}(w) = \frac{1}{2} \mathbf{h}^T \mathbf{A}'' \mathbf{h} = -\mathbf{h}^T \mathbf{F}'' \mathbf{h}, \quad (4.28)$$

em que

$$\mathbf{F}'' = \Delta^T l''^{-1}(\hat{\alpha}) \Delta = \mathbf{G}^T l''(\hat{\alpha}) \mathbf{G}, \quad (4.29)$$

$$\Delta = \left\{ \frac{\partial^2 l(\alpha|w)}{\partial \alpha \partial w^T} \right\}_{(w_0, \hat{\alpha})} \text{ e } \mathbf{G} = \left(\frac{\partial \hat{\alpha}_w}{\partial w^T} \right)_{w_0}. \quad (4.30)$$

Este teorema pode ser usado para identificar os casos influentes. Agora seja $w = w_0 + h_i$ em que h_i é um vetor unitário contendo 1 na i -ésima posição e zeros nas outras posições, que significa que o i -ésimo caso é perturbado. Então (4.28) passa a ser

$$LD''(w) = \frac{1}{2} \mathbf{h}_i^T \mathbf{A}'' \mathbf{h}_i = \frac{1}{2} \mathbf{A}''_{ii},$$

em que \mathbf{A}''_{ii} é o i -ésimo elemento da diagonal de \mathbf{A}'' . O **Teorema 4.8** tem uma interpretação geométrica apresentada por Cook (1986) em que a idéia é baseada na direção da curva e a curvatura máxima é definida por Bates e Watts (1980). Agora será dada uma breve introdução. Seja uma perturbação w , $\eta(w) = (w^T, LD(w))^T$ pode ser considerado como uma superfície n -dimensional π_w em R^{n+1} , o qual é referido no gráfico de influência por Cook (1986). Seja $\eta_h(b) = \eta(w_0 + b\mathbf{h})$ em que b é um número real e \mathbf{h} foi definido anteriormente. Então, a curvatura normal direcional a w_0 ao longo de \mathbf{h} e a curvatura máxima são definidas, respectivamente, por

$$C_h = \frac{\|\eta_h''^N\|}{\|\eta_h''\|}, \text{ e } C_{max} = \max_{\|\mathbf{h}\|=1} C_h, \quad (4.31)$$

em que $\eta_h'' = \{d\eta_h(b)/db\}_{b=0}$, $\eta_h'' = \{d^2\eta_h(b)/db^2\}_{b=0}$ e $\eta_h''^N$ é o componente normal de η_h'' . Ambos C_h e C_{max} são chamados de curva de influência.

Teorema 4.9

A curvatura normal de π_w para w_0 definida em (4.31) pode ser expressa por

$$C_h = \mathbf{h}^T \mathbf{A}'' \mathbf{h} = -2\mathbf{h}^T \mathbf{F}'' \mathbf{h}. \quad (4.32)$$

Em resumo, para influência local aproximada descrita aqui, a matriz influência $\mathbf{A}'' = -2\mathbf{F}''$ é muito importante para identificar os casos influentes. Para um dado método de perturbação, o elemento diagonal de \mathbf{A}'' e a direção \mathbf{h}_{max} podem ser aplicados para análise de influência de diagnóstico de acordo com o **Teorema 4.8** e o **Teorema 4.9**. A seqüência do processo é: (a) calcula-se a informação observada $-l''(\hat{\alpha})$ para o modelo postulado; (b) considera-se uma perturbação específica e então calcula-se Δ e \mathbf{G} dados em (4.30) sobre o modelo perturbado; e (c) calcula-se $\mathbf{A}'' = -2\mathbf{F}''$ através de (4.29) e constroi-se o gráfico de dispersão de (i, \mathbf{A}''_{ii}) ou $(i, (\mathbf{h}_{max})_i)$, ou ambos, $i = 1, \dots, n$.

Agora esses resultados serão aplicados aos modelos não-lineares da família exponencial.

(a) Para o modelo postulado (4.1), a informação observada $-l''(\hat{\alpha})$ é expressa por

$$-l''(\hat{\alpha}) = \begin{pmatrix} \sigma^{-2}\mathbf{R}^T(\mathbf{I}_p - \mathbf{B}_\theta)\mathbf{R} & \mathbf{0} \\ \mathbf{0} & -l''_{\phi\phi} \end{pmatrix}_{(\hat{\beta}, \hat{\sigma}^2)}, \quad (4.33)$$

em que $l''_{\phi\phi}$ é um escalar.

(b) Calcula-se a matriz influência \mathbf{A}'' para uma dada perturbação no modelo. Como exemplo, considere a perturbação na variável resposta \mathbf{Y} somando-se um vetor $w \in R^n$ em \mathbf{Y} . Então a log-verossimilhança correspondente a esta perturbação no modelo é

$$l(\alpha|w) = \sum_{i=1}^n [\phi\{(y_i + w_i)\theta_i - b(\theta_i) + c(y_i + w_i)\} - \frac{1}{2}s(\phi) - \frac{1}{2}t(y_i + w_i)] \quad (4.34)$$

em que $\alpha = (\beta^T, \sigma^2)^T$, $g(\mu_i) = f(x_i, \beta)$ e $w_0 = \mathbf{0}$. Neste caso, $-l''_{\phi\phi} = \frac{1}{2}ns''(\phi)$. Segue-se de (4.34) que

$$l'_\beta = \phi\mathbf{D}^T\mathbf{V}^{-1}(\mathbf{Y} + w - \mu(\beta)), \quad l''_{\beta w} = \phi\mathbf{D}^T\mathbf{V}^{-1} = \phi\hat{\theta}^T,$$

$$\left\{ \frac{\partial^2 l}{\partial \phi \partial w_i} \right\}_{w_0, \hat{\alpha}} = \hat{\theta}_i - c'(y_i) = \hat{\theta}_i - \tilde{\theta}_i$$

Usa-se a expressão $c'(y_i) = b'^{-1}(y_i)$ e $\tilde{\theta} = (\tilde{\theta}_i) = (b'^{-1}(y_i))$. Então tem-se $l''_{\phi w} = \hat{\theta}^T - \tilde{\theta}^T$, logo

$$\Delta^T = (\phi\hat{\theta}^T, \theta - \tilde{\theta})_{(\hat{\theta}, \hat{\sigma}^2)}.$$

(c) Substituindo $-l''(\hat{\alpha})$ e Δ em (4.29), encontra-se

$$\mathbf{A}'' = 2\hat{\phi}\hat{\mathbf{V}}^{-1}\mathbf{Q}(\mathbf{I}_p - \mathbf{B}_\theta)^{-1}\mathbf{Q}^T\hat{\mathbf{V}}^{-1} + 4n^{-1}s''^{-1}(\hat{\phi})(\theta - \tilde{\theta})(\theta - \tilde{\theta})^T \quad (4.35)$$

Então calculam-se os valores de \mathbf{A}_{ii} e $(h_{max})_i$, $i = 1, \dots, n$ e encontram-se os pontos influentes.

4.4.2 Esquema de perturbação aleatória

Um modelo de perturbação aleatória foi apresentado por Wu e Wan (1994) para extrair a informação entre casos em um estudo de influência local em regressão não-linear. Este esquema de perturbação também pode ser aplicado aos modelos não-lineares da família exponencial.

O modelo postulado (4.1) pode ser representado por

$$\mathbf{Y} = \mu(\beta) + \sigma\mathbf{V}^{\frac{1}{2}}\varepsilon, \quad \varepsilon \sim (\mathbf{0}, \mathbf{I}_n), \quad (4.36)$$

em que ε é o erro aleatório padrão e $\mathbf{Y} \sim ED(\mu(\beta), \sigma^2)$ é o vetor observação. Considere a perturbação ε somado a outro vetor aleatório ρ para indicar a iteração entre casos. Então, o modelo perturbado com ruído aleatório ρ pode ser expresso por

$$\mathbf{Y} = \mu(\beta) + \sigma \mathbf{V}^{\frac{1}{2}}(\varepsilon + \rho). \quad (4.37)$$

Para discussão futura, assume-se que

$$E(\rho) = 0, \quad \text{Var}(\rho) = a^2 \mathbf{I}_n, \quad E(\rho \varepsilon) = \Omega,$$

em que a é uma constante e $\Omega = (\mathbf{w}_{kj}) = (\Omega_1, \dots, \Omega_n)$ com $\mathbf{w}_{kj} = E(\rho_k \varepsilon_j)$ reflete a iteração das observações k e j com um ruído na k -ésima observação, e $\Omega_k = E(\rho_k \varepsilon)$ reflete a iteração entre as observações considerando um ruído na k -ésima observação, em que ρ_k é o k -ésimo componente de ρ . A variância de $\varepsilon + \rho$ e \mathbf{Y} são respectivamente

$$\begin{aligned} \text{Var}(\varepsilon + \rho) &\equiv \Sigma_{RP} = \mathbf{I}_n + \Omega + \Omega^T + a^2 \mathbf{I}_n, \\ \text{Var}(\mathbf{Y}) &\equiv \sigma^2 \mathbf{V}_{RP} = \sigma^2 \mathbf{V}^{\frac{1}{2}} \Sigma_{RP} \mathbf{V}^{\frac{1}{2}}. \end{aligned} \quad (4.38)$$

O estimador $\hat{\beta}_{RP}$ de β é definido como solução da seguinte equação quase-escore

$$EQE(\beta, \Sigma_{RP}) = \mathbf{D}^T(\beta) \mathbf{V}_{RP}^{-1}(\beta) (\mathbf{Y} - \mu(\beta)) = 0. \quad (4.39)$$

Aqui, σ^2 é considerado um parâmetro de perturbação, estimado por

$$\hat{\sigma}_{RP}^2 = n^{-1} \hat{\mathbf{e}}_{RP}^T \mathbf{V}_{RP}^{-1}(\hat{\beta}_{RP}) \hat{\mathbf{e}}_{RP}$$

em que $\hat{\mathbf{e}}_{RP} = \mathbf{Y} - \mu(\hat{\beta}_{RP})$. Assim, $\hat{\beta}_{RP}$ pode ser calculado através do processo de iteração visto em (3.16). Caso a perturbação aleatória desapareça, o modelo (4.37) volta para o modelo nulo (4.36) em que temos $\Sigma_{RP} \mathbf{I}_n$, $\mathbf{V}_{RP} = \mathbf{V}$, $\hat{\beta}_{RP} = \hat{\beta}$ e $\hat{\mathbf{e}}_{RP} = \hat{\mathbf{e}}$, respectivamente.

Seguindo (4.26), a distância de verossimilhança dentro do modelo com perturbação aleatória (4.37) é denotada por

$$LD_{RP}(\Omega) = 2\{l(\hat{\beta}) - l(\hat{\beta}_{RP})\}$$

o qual é função de $\Omega = (w_{kj})$ e a . Para adquirir estatísticas de diagnóstico valiosas, considere-se o ruído ρ_k ($k = 1, \dots, n$) separadamente. Significa que $LD_{RP}(\Omega)$ é considerado como uma função da perturbação $w^{(k)} = \Omega_k = (w_{k1}, \dots, w_{kn})^T$ para o tempo, $k = 1, \dots, n$. Em outras palavras, considera-se o gráfico de influência parcial.

$$\eta(w^{(k)}) = \begin{pmatrix} w^{(k)} \\ LD_{RP}(\Omega) \end{pmatrix}, k = 1, \dots, n \quad (4.40)$$

com $w^{(k)} = \Omega_k$ e $w_0^{(k)} = 0$. De fato, $w^{(k)}$ corresponde ao modelo em que só a k -ésima observação é perturbada pelo ruído aleatório ρ_k . Além disso, a matriz de influência e a curva de influência para a equação (4.40) são denotadas por $A''^k = -sF''^{(k)}$, $C_h^{(k)}$, $C_{max}^{(k)}$ e $h_{max}^{(k)}$, respectivamente. Através do **Teorema 4.8** e da equação quase-escore (4.39), não é difícil encontrar estas quantidades. Tem-se o seguinte teorema.

Teorema 4.10

Para o gráfico de influência parcial (4.40), A''^k pode ser expresso por

$$A''^k = 2\hat{\phi}^2 \{ \tilde{h}_{kk}(\mathbf{rp})(\mathbf{rp})^T + \mathbf{rp}_k \tilde{h}_k \mathbf{rp}^T + \mathbf{rp}_k \mathbf{rp} \tilde{h}_k^T + \mathbf{rp}_k^2 \tilde{\mathbf{H}} \}, \quad (4.41)$$

em que

$$\tilde{\mathbf{H}} = (\tilde{h}_1, \dots, \tilde{h}_n) = (\tilde{h}_{kj}) = \{ \mathbf{V}^{-\frac{1}{2}} \mathbf{D} (-l''_{\beta\beta}^{-1}) \mathbf{D}^T \mathbf{V}^{-\frac{1}{2}} \}_{\hat{\beta}}.$$

A matriz de influência A''^k pode ser utilizada na análise de influência local como mostram o *Teorema 4.8* e o *Teorema 4.9*. Na prática, a informação observada $-l''_{\beta\beta}$ pode ser aproximada pela matriz de informação de Fisher $\mathbf{J}_\beta(\mathbf{Y}) = \phi \mathbf{D}^T \mathbf{V}^{-1} \mathbf{D}$ em (4.41). Assim, temos que $\tilde{\mathbf{H}} = \hat{\phi}^{-1}$ e logo

$$A''^k \approx 2\hat{\phi}^{-2} \{ h_{kk}(\mathbf{rp})(\mathbf{rp})^T + \mathbf{rp}_k h_k \mathbf{rp}^T + \mathbf{rp}_k \mathbf{rp} h_k^T + \mathbf{rp}_k^2 \mathbf{H} \}. \quad (4.42)$$

No teorema a seguir é apresentada a aproximação para a curvatura máxima $C_{max}^{(k)}$ e a direção correspondente $h_{max}^{(k)}$.

Teorema 4.11

No gráfico de influência parcial (4.40), $C_{max}^{(k)} = C^{(k)}$ e $h_{max}^{(k)} = h^{(k)}$ são, aproximadamente, expressos por

$$C^{(k)} \approx 2\hat{\phi} \{ (\mathbf{rp}_k)^2 + h_{kk}(\mathbf{rp})^T(\mathbf{rp}) \} \quad (4.43)$$

e

$$h^{(k)} \approx (2\hat{\phi})^{\frac{1}{2}} (h_k \mathbf{rp}_k + h_{kk} \mathbf{rp}) \{ h_{kk} C^{(k)} \}^{-\frac{1}{2}}. \quad (4.44)$$

O **Teorema 4.11** pode ser usado para medir interação entre casos como descritos por Wu e Wan (1994). Pela aproximação da influência local de Cook (1986), o vetor $\mathbf{h}^{(k)}$ reflete a interação entre os casos quando o k -ésimo caso é perturbado por um ruído aleatório ρ_k . O tamanho do j -ésimo elemento $\mathbf{h}_j^{(k)}$ representa a parte da contribuição para a curva de influência $C^{(k)}$ para o caso j quando o k -ésimo caso é perturbado. Para medir a influência, é introduzido a matriz interação $\mathbf{Z} = (Z_{kj})$, com elementos

$$Z_{kj} = C^{(k)}(h_j^k)^2, \quad k, j = 1, \dots, n.$$

O elemento Z_{kj} mede a influência do j -ésimo caso quando há um ruído no caso k . A matriz \mathbf{Z} é muito útil para detectar a interação e a influência em comum entre os casos. Para usar a matriz \mathbf{Z} mais efetivamente, define-se os vetores ACZ e ARZ que mensura a interação entre casos e as demais observações:

$$ACZ_j = n^{-1} \sum_{k=1}^n Z_{kj}, \quad \text{e} \quad ARZ_k = n^{-1} \sum_{j=1}^n Z_{kj}. \quad (4.45)$$

Nota-se que os vetores ACZ e ARZ possuem medidas diferentes. ACZ_j mede a contribuição média do caso j para a influência das outras observações, enquanto ARZ_k mede a contribuição média das outras observações para a influência do caso k . Uma outra forma de calcular essas medidas é apresentada no Teorema 4.12.

Teorema 4.12

Se o parâmetro de dispersão ϕ é estimado por $\hat{\phi}^{-1} = \hat{\sigma}^2 = n^{-1}(\mathbf{rp})^T(\mathbf{rp})$, tem-se que

$$\begin{aligned} ACZ_j &= 2(\mathbf{rp}^T \mathbf{rp})^{-1} \left\{ p(rp_j)^2 + \sum_{k=1}^n h_{kj}^2 (\mathbf{rp}_k)^2 h_{kk}^{-1} \right\} \\ ARZ_k &= 2 \{ h_{kk} + (rp_k)^2 (\mathbf{rp}^T \mathbf{rp})^{-1} \}. \end{aligned} \quad (4.46)$$

Para detectar a influência conjunta dos casos a e b , pode-se comparar os valores de Z_{aa} , Z_{ab} , Z_{ba} e Z_{bb} com outros valores das colunas de a e b . Assim, introduz-se as seguintes medidas

$$\begin{aligned} Z_{(a,b),a} &= (Z_{aa} + Z_{ab}) / \sum_{j=1}^n Z_{aj}, \\ Z_{(a,b),b} &= (Z_{ba} + Z_{bb}) / \sum_{j=1}^n Z_{bj}. \end{aligned} \quad (4.47)$$

Se os valores de ambos $Z_{(a,b),a}$ e $Z_{(a,b),b}$ são bastante grandes, diga-se maiores que 0.5,

e os valores dos outros elementos $Z_{aj}, Z_{bj}, j \neq a, b$, são pequenos, então os casos a e b são conjuntamente influentes.

4.5 Influência generalizada

Influência é uma parte importante do diagnóstico de modelos de regressão linear. É medida pelo elemento diagonal h_{ii} da matriz estimada \mathbf{H} e pode ser usado para avaliar a importância de observações individualmente. Na seção 4.3, foi introduzida a matriz estimada $\mathbf{H} = \mathbf{V}^{-\frac{1}{2}} \mathbf{D} (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-\frac{1}{2}}$ para os modelos não-lineares de família exponencial. Porém, esta matriz não é bem ajustada quando há existência de pontos influentes. De fato, em regressão linear, ou seja, $y_i = x_i^T \beta + \varepsilon_i, i = 1, \dots, n$, a influência tem várias propriedades boas, como $Var(y_i) = \sigma^2 h_{ii}, Var(\hat{\varepsilon}_i) = \sigma^2 (1 - h_{ii}), \rho(y_i, \hat{y}_i) = \sqrt{h_{ii}}, h_{ij} = \partial \hat{y}_i / \partial y_j$ dentre outras, em que $\hat{y}_i = x_i^T \hat{\beta}, \hat{\varepsilon}_i = y_i - x_i^T \hat{\beta}$ e $\rho(y_i, \hat{y}_i)$ é o coeficiente de correlação de y_i e \hat{y}_i . É fácil verificar que estas propriedades não são válidas para $\mathbf{H} = \mathbf{V}^{-\frac{1}{2}} \mathbf{D} (\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-\frac{1}{2}} = (h_{ii})$, em modelos não-lineares da famílias exponencial. Sendo assim, é necessário encontrar uma definição mais adequada de influência em modelos não-lineares da família exponencial. Nesta seção, é introduzida uma definição geral para classe geral de modelos. Esta definição pode ser aplicada aos modelos não-lineares da família exponencial e (a) é consistente com a influência ordinária dos modelos de regressão linear; e (b) pode ser usado como uma medida para avaliar a importância individual de observações. Também é uma conexão fechada entre esta influência e a influência local de uma perturbação que alavanca a média.

4.5.1 Definição e computação

Como apontado por muitos autores, a propriedade $h_{ii} = \partial \hat{y}_i / \partial y_i$, reflete diretamente a influência da observação y_i no ajuste (Cook e Weisberg, 1982; Hoaglin e Welsch, 1978; Yoshizoe, 1991), a definição abaixo baseia-se nesta idéia.

Definição 4.1

Seja $\mathbf{Y} = (y_1, \dots, y_n)^T$ um vetor n-dimensional de observações com a função de densidade de probabilidade dada por $p(y; \alpha)$ e $E(\mathbf{Y}) = \mu = \mu(\alpha)$, em que α é um parâmetro desconhecido. Um estimador de α é denotado por $\tilde{\alpha} = \tilde{\alpha}(\mathbf{Y})$ e $\tilde{\mathbf{Y}} = \mu(\tilde{\alpha})$ é o vetor de resposta ou vetor estimado. Então define-se

$$GL(\tilde{\alpha}) = \frac{\partial \tilde{\mathbf{Y}}}{\partial \tilde{\mathbf{Y}}^T} = \left(\frac{\partial \tilde{y}_i}{\partial y_j} \right), \quad (4.48)$$

como a influência generalizada de $\tilde{\alpha}$

Assim, a influência generalizada ainda mede a influência de observações no ajuste do modelo sobre o estimador $\tilde{\alpha}$. Obviamente, a definição (4.48) pode ser usada para muitos estimadores e modelos em geral, incluindo os modelos não-lineares da família exponencial. A observação com valor grande em $GL_{ii} = \partial \tilde{y}_i / \partial y_j$ também será chamada de ponto de influência. Além disso, é fácil verificar de $\tilde{\mathbf{Y}} = \mu(\tilde{\alpha})$ e $\tilde{\alpha} = \tilde{\alpha}(\mathbf{Y})$ que

$$GL(\tilde{\alpha}) = \left\{ \frac{\partial \mu}{\partial \alpha^T} \frac{\partial \tilde{\alpha}(\mathbf{Y})}{\partial \mathbf{Y}^T} \right\},$$

que é invariante sob reparametrização, isso é se $\alpha = \alpha(\gamma)$ é ligado um a um e $\tilde{\gamma} = \gamma(\tilde{\alpha})$, em que $\gamma(\alpha)$ é a inversa de $\alpha(\gamma)$, então

$$GL(\tilde{\alpha}) = \{(\partial \mu / \partial \gamma^T)(\partial \tilde{\gamma}(\mathbf{Y}) / \partial \mathbf{Y}^T)\}_{\tilde{\gamma}} = GL(\tilde{\gamma}).$$

Lema 4.5.1. *Seja $l(\alpha; y)$ a log-verossimilhança de \mathbf{Y} e $\hat{\alpha} = \hat{\alpha}(\mathbf{Y})$ seja o único estimador de máxima verossimilhança de α . Se $l(\alpha; y)$ possui derivada de segunda ordem com relação a α e y contínua, então tem-se*

$$GL(\hat{\alpha}) = \{(\mathbf{D}_\alpha)(-l''_{\alpha\alpha})^{-1}(l''_{\alpha Y})\}_{\hat{\alpha}}, \quad (4.49)$$

em que

$$\mathbf{D}_\alpha = \partial \mu / \partial \alpha^T, \quad l''_{\alpha\alpha} = \partial^2 l(\alpha; \mathbf{Y}) / \partial \alpha \partial \alpha^T, \quad l''_{\alpha Y} = \partial^2 l(\alpha; \mathbf{Y}) / \partial \alpha \partial \mathbf{Y}^T.$$

Este lema revela a ligação essencial entre influência generalizada e matriz de informação, que conduz à curvatura em modelos não-lineares. A equação (4.49) pode ser aplicada para classe geral de modelos. Mais adiante, será usado este lema para encontrar uma expressão útil da influência para modelos não-lineares da família exponencial.

Teorema 4.13

Para os modelos não-lineares da família exponencial (4.1), a influência do estimador de máxima verossimilhança $\hat{\alpha} = (\hat{\beta}^T, \hat{\sigma}^T)^T$ é expressa por

$$GL(\hat{\alpha}) = \{\mathbf{Q}(\mathbf{I}_p - \mathbf{B}_\theta)^{-1} \mathbf{Q}^T \mathbf{V}^{-1}\}_{\hat{\beta}}. \quad (4.50)$$

Através deste teorema, vários resultados úteis são decorrentes:

(1) Desde que $\mu = b'(\theta(\beta))$ não dependa de σ^2 o qual é ortogonal a β , tem-se

$$GL(\hat{\alpha}) = GL(\hat{\beta}) = \left\{ \frac{\partial \mu}{\partial \beta^T} (-l''_{\beta\beta})^{-1} l''_{\beta Y} \right\}_{\beta=\hat{\beta}}.$$

em que $GL(\hat{\beta})$ é a influência de $\hat{\beta}$ com σ^2 conhecido. Isto significa que a influência de $\hat{\alpha} = (\hat{\beta}^T, \sigma^2)^T$ é independente do parâmetro de dispersão σ^2 .

(2) Em (4.49), se a matriz de informação observada $-l''_{\beta\beta}$ for substituída pela informação de Fisher $\mathbf{J}_\beta(\mathbf{Y}) = \phi \mathbf{D}^T \mathbf{V}^{-1} \mathbf{D}$, então $GL(\hat{\alpha}) \approx \mathbf{D}(\mathbf{D}^T \mathbf{V}^{-1} \mathbf{D})^{-1} \mathbf{D}^T \mathbf{V}^{-1}$.

(3) Para os modelos lineares generalizados com ligação canônica, tem-se $\theta = \mathbf{X}\beta$, $\mathbf{A}_\theta^l = \mathbf{0}$ e $\mathbf{B}_\theta = \mathbf{0}$. Assim, $GL(\hat{\alpha}) = \mathbf{X}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$ e $\mathbf{H} = \mathbf{V}^{-\frac{1}{2}} \mathbf{X}(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-\frac{1}{2}}$ possui o mesmo elemento diagonal, como esperado.

(4) Para o modelo de regressão não linear normal, desde que $\mathbf{V} = \mathbf{I}_n$, $\mathbf{B}_\theta = \mathbf{B}$ tem-se $\mathbf{H} = \mathbf{D}(\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T$ e $GL(\hat{\alpha}) = \mathbf{Q}^T (\mathbf{I}_p - \mathbf{B})^{-1} \mathbf{Q}$. Além disso, para o modelo de regressão linear normal, tem-se $GL(\hat{\alpha}) = \mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$.

A definição geral de influência para uma ampla classe de modelos introduzida pode ser aplicada para outros estimadores e modelos. É importante enfatizar que a influência generalizada independe da variável resposta \mathbf{Y} e que é normalmente ligado a não linearidade do modelo.

4.5.2 Influência generalizada e influência local

Como mostrado por St. Laurent e Cook (1993), há uma ligação íntima entre influência e análise de influência local. Considere o modelo perturbado (4.34), ou seja o vetor \mathbf{Y} é perturbado somando um vetor \mathbf{w} . É fácil verificar de (4.34) que $\Delta = \partial^2 l / \partial \alpha \partial \omega^T$ visto em (4.30). Assim, segue de (4.29) e (4.49) que

$$\begin{aligned} -\mathbf{F}'' &= \Delta^T (\mathbf{D}_\alpha^T \mathbf{D}_\alpha)^{-1} (\mathbf{D}_\alpha^T \mathbf{D}_\alpha) (-l''_{\alpha\alpha}^{-1}) \Delta \\ &= \Delta^T (\mathbf{D}_\alpha^T \mathbf{D}_\alpha)^{-1} \mathbf{D}_\alpha^T GL(\hat{\alpha}) \end{aligned} \quad (4.51)$$

se $\mathbf{D}_\alpha^T \mathbf{D}_\alpha$ é não singular, e

$$\begin{aligned} GL(\hat{\alpha}) &= \mathbf{D}_\alpha (\Delta \Delta^T)^{-1} (\Delta \Delta^T) (-l''_{\alpha\alpha}^{-1}) \Delta \\ &= \mathbf{D}_\alpha (\Delta \Delta^T)^{-1} \Delta (-\ddot{\mathbf{F}}) \end{aligned} \quad (4.52)$$

se $\Delta\Delta^T$ é não singular. Agora considera-se um caso geralmente encontrado em que $s(y_i, \phi) = s(\phi) + t(y_i)$ em (4.1). Então, segue de (4.35) e (4.50) que

$$\mathbf{A}'' = 2\hat{\phi}\hat{\mathbf{V}}^{-1}GL(\hat{\alpha}) + 4n^{-1}s''^{-1}(\hat{\phi})(\hat{\theta} - \tilde{\theta})(\hat{\theta} - \tilde{\theta})^T.$$

Em particular, para os modelos não-lineares normais, tem-se $s(\phi) = -\log(\phi)$, $\mathbf{V} = \mathbf{I}_n$, $\mu = \theta$ e $\tilde{\mu} = \mathbf{Y}$. Então \mathbf{A}'' se reduz a

$$\mathbf{A}'' = 2\hat{\sigma}^{-2}GL(\hat{\alpha}) + 4n^{-1}\hat{\sigma}^{-4}\hat{e}\hat{e}^T.$$

4.6 Diagnósticos para dispersão da variância

Teste para detectar a não constância da variância (heterocedasticidade) é um problema comum em diagnóstico de regressão. Para os modelos lineares generalizados e modelos não-lineares da família exponencial, por causa da variância nominal $Var(y_i) = \sigma^2\mathbf{V}(\mu_i)$, $i = 1, \dots, n$ que sempre são não constantes (exceto quanto ao caso da distribuição normal), não é necessário descobrir a não constância da variância. Porém, o problema de discrepância ainda existe com estes modelos. A preocupação com a superdispersão foi um problema comum de nos recentes anos. O termo superdispersão média ocorre quando a variância da média estimada excede a variância nominal. Em contrapartida, o termo subdispersão média ocorre quando a variância da média estimada é menor que a variância nominal.

Considere o seguinte modelo de dispersão da variância

$$g(\mu_i) = f(x_i; \beta), \quad y_i \sim ED(\mu_i, \sigma_i^2 \omega_i^{-1}), \quad i = 1, \dots, n, \quad (4.53)$$

em que ω_i s são pesos conhecidos e $Var(y_i) = \sigma_i^2 \omega_i^{-1} V(\mu_i)$. Se $\sigma_i^2 = \sigma^2$ para todo i , então a dispersão é fixa, caso contrário é variante. Para simplificar, assume-se que $\omega_i = 1$ par todo i , pois são conhecidos. Um teste natural para dispersão da variância é

$$H_0 : \sigma_i^2 = \sigma^2, \quad i = 1, \dots, n.$$

Obviamente há muitos parâmetros ligados com o modelo (4.53), assim os parâmetros de dispersão normalmente são modelados através de outro parâmetro γ a fim de simplificar o problema (Efron, 1986; Smyth, 1989). Baseando-se nesta idéia, propõe-se o seguinte modelo da dispersão da variância para os modelos não-lineares da família exponencial:

$$\begin{cases} y_i \sim ED(\mu_i, \phi_i^{-1}), & \phi_i = \phi m_i, \\ g(\mu_i) = f(x_i, \beta), & m_i = m(z_i, \gamma); \end{cases} \quad (4.54)$$

em que z_i s são covariáveis, γ é um vetor com q parâmetros e $m(\cdot, \cdot)$ é uma função conhecida. Neste modelo, $\phi = \sigma^2$ pode ser considerado como a dispersão nominal e m_i s são adicionados para refletir a variação da dispersão. É assumido que há um único valor γ_0 de γ tal que $m(z_i; \gamma_0) = 1$ para todo i . Se $\gamma = \gamma_0$ então $Var(y_i) = \sigma^2 V(\mu_i)$ para todo i e a dispersão é constante. Logo, o teste para variação da dispersão é equivalente ao teste de hipótese

$$H_0 : \gamma = \gamma_0, H_1 : \gamma \neq \gamma_0.$$

É fácil verificar que se y_i é normalmente distribuído, ou seja, $y_i \sim N(x_i^T \beta, \sigma^2 m_i^{-1})$ no modelo (4.54), então H_0 é reduzida ao teste de heterocedasticidade em regressão linear. No restante desta seção são mostrados vários testes estatísticos para H_0 baseado no modelo de dispersão da variância.

4.6.1 Razão de verossimilhanças e estatística score

Para simplificar, assume-se que $s(y_i, \phi) = s(\phi) + t(y_i)$. Segue de (2.6) e (2.25) que a log-verossimilhança do modelo (4.54) pode ser expressa por

$$\begin{aligned} l(\gamma, \phi, \beta) &= \sum_{i=1}^n \left[\phi m_i \{y_i \theta_i - b(\theta_i) - c(y_i)\} - \frac{1}{2} s(\phi m_i) - \frac{1}{2} t(y_i) \right] \\ &= -\frac{1}{2} \sum_{i=1}^n \{ \phi m_i d'_i + s(\phi m_i) + t(y_i) \}, \end{aligned} \quad (4.55)$$

em que $d'_i = -2\{y_i \theta_i - b(\theta_i) - c(y_i)\}$. Nesta seção, fixamos $\alpha = (\gamma^T, \phi, \beta^T)^T$ e os estimadores de máxima verossimilhança de α para o modelo (4.54) são denotados por $\hat{\alpha} = (\hat{\gamma}^T, \hat{\phi}, \hat{\beta}^T)^T$.

Teorema 4.14

Para os modelos de dispersão da variância (4.54) e (4.55), a estatística da razão de verossimilhança para H_0 é

$$LR = n \{ s(\hat{\phi}_0) - \hat{\phi}_0 s'(\hat{\phi}_0) \} - \{ \mathbf{1}^T \mathbf{s}_m - \phi \mathbf{1}^T \mathbf{M} \mathbf{s}'_m \} \hat{\alpha} \quad (4.56)$$

em que $\mathbf{M} = \text{diag}(m_1, \dots, m_n)$, $\mathbf{s}_m = (s(\phi m_1), \dots, s(\phi m_n))^T$, $\mathbf{s}'_m = (s'(\phi m_1), \dots, s'(\phi m_n))^T$ e $\mathbf{1}^T = (1, \dots, 1)$, respectivamente.

Corolário 7

Para o modelo normal e normal inverso, (4.56) se resume a

$$LR = n \log(\hat{\phi} \hat{m}_g / \hat{\phi}_0), \quad m_g = \left(\prod_{i=1}^n m_i \right)^{\frac{1}{n}}. \quad (4.57)$$

No modelo (4.54), γ é o parâmetro de interesse e $(\phi, \beta^T)^T$ são os parâmetros de variação. Assim, pode-se obter a estatística escore para H_0 . Como usual, a estatística escore para H_0 é

$$SC = \left\{ \left(\frac{\partial l}{\partial \gamma} \right)^T \mathbf{J}^{\gamma\gamma} \left(\frac{\partial l}{\partial \gamma} \right) \right\}_{\hat{\alpha}_0}, \quad (4.58)$$

em que $\mathbf{J}^{\gamma\gamma}$ é o lado esquerdo superior de \mathbf{J}^{-1} e $\mathbf{J} = \mathbf{J}(\gamma, \phi, \beta)$ é a matriz de informação de fisher de \mathbf{Y} para α . Então tem-se o seguinte teorema.

Teorema 4.15

Para o modelos de dispersão da variância (4.54) e (4.55), a estatística escore para testar H_0 pode ser expressa por

$$SC = \frac{1}{2} s''^{-1}(\hat{\phi}_0) (\mathbf{d}^T \mathbf{P}_c \mathbf{d})_{\hat{\alpha}_0} \quad (4.59)$$

em que $\mathbf{d} = (d_1, \dots, d_n)^T$, $d_i = d_i(y_i, \mu_i) = d'_i + k$, \mathbf{P}_c é a matriz projeção de $\Psi_c = (\mathbf{I}_n - \mathbf{1}\mathbf{1}^T/n)\Psi$ e $\Psi = (\Psi_{ia})$ com $\Psi_{ia} = \partial m_i / \partial \gamma_a$, $i = 1, \dots, n$, $a = 1, \dots, q$.

Através deste teorema vários resultados úteis são obtidos.

(1) A equação (4.59) pode ser usada para modelos não-lineares. Para o modelo normal, tem-se $s''(\phi) = \phi^{-2} = \sigma^4$ e $d_i = e_i^2 = (y_i - \mu_i)^2$, então (4.59) torna-se

$$SC = \frac{1}{2} (\mathbf{u}^T \mathbf{P}_c \mathbf{u})_{\hat{\alpha}_0} \quad (4.60)$$

em que $\mathbf{u} = (u_i)$, $u_i = e_i^2 / \sigma^2$. Como esperado, este resultado coincide com a equação de Cook e Weisberg (1982) para o teste de heterocedasticidade em regressão linear normal. Para o modelo não linear gamma $y_i \sim GA(\mu_i, \sigma^2 m_i^{-1})$, tem-se $s(\phi) = -2\{\phi \log(\phi) - \log \Gamma(\phi)\}$, $d_i = 2\{(y_i - \mu_i) / (\mu_i - \log(y_i / \mu_i))\}$, e SC pode ser calculada através da equação (4.59).

(2) Pode-se reescrever (4.59) como uma forma padronizada de (4.60). Seja $E(d_i) = -s'(\phi) - k$, $Var(d_i) = 2s''(\phi)$ e $\mathbf{P}_c \mathbf{1} = \mathbf{0}$, se for fixado $\mathbf{d}_s = (d_{si})$ com $d_{si} = \{d_i - E(d_i)\} / \sqrt{Var(d_i)}$, então tem-se $SC = (\mathbf{d}_s^T \mathbf{p}_c \mathbf{d}_s)_{\hat{\alpha}_0}$, isso é a norma do desvio padronizado \mathbf{d}_s avaliado em $\hat{\alpha}_0$.

(3) A expressão (4.59) pode ser estendida para modelos mais gerais em que a condição $s(y_i, \phi) = s(\phi) + t(y_i)$ não é satisfeita. Através de alguns cálculos tem-se

$$SC = \frac{1}{2} \{(\mathbf{d}' + \mathbf{s}')^T \tilde{\mathbf{S}}^{-\frac{1}{2}} \tilde{\mathbf{P}} \tilde{\mathbf{S}}^{-\frac{1}{2}} (\mathbf{d}' + \mathbf{s}')\}_{\hat{\alpha}_0},$$

em que $\tilde{\mathbf{S}} = \text{diag}(\tilde{S}_i)$, $\tilde{S}_i = E\{\partial^2 s(\phi, y_i)/\partial \phi^2\}$, $s' = (s'_i)$, $s'_i = \partial s(\phi, y_i)/\partial \phi$, $\tilde{\mathbf{P}} = \tilde{\Psi}(\tilde{\Psi}^T \tilde{\Psi})^{-1} \tilde{\Psi}^T$, $\tilde{\Psi} = (\mathbf{I}_n - \tilde{\mathbf{P}}_1) \tilde{\mathbf{S}}^{\frac{1}{2}} \Psi$ e $\tilde{\mathbf{P}}_1 = \tilde{\mathbf{S}}^{-\frac{1}{2}} \mathbf{1} (\mathbf{1}^T \tilde{\mathbf{S}} \mathbf{1})^{-1} \mathbf{1}^T \tilde{\mathbf{S}}^{-\frac{1}{2}}$.

4.6.2 Razão de verossimilhanças ajustada e estatística escore

É fácil verificar que o parâmetro γ é ortogonal ao parâmetro β mas não ortogonal ao parâmetro ϕ . Para adquirir a verossimilhança perfilada ajustada, precisa-se de uma transformação para $(\phi, \beta^T)^T$ para um vetor de parâmetros λ tal que γ é ortogonal a λ .

Agora considere $\alpha = (\alpha^T_1, \alpha^T_2)^T$ em que $\alpha^T_1 = \gamma$, $\alpha^T_2 = (\phi, \beta^T)^T$ e $\alpha^T_{10} = \gamma_0$, $\hat{\alpha}^T_{20} = (\hat{\phi}_0, \hat{\beta}_0^T)^T$. Então a log-verossimilhança de \mathbf{Y} pode ser expressa por

$$l(\alpha_1, \alpha_2) = l(\alpha_1, \alpha_2(\alpha_1, \lambda)) = l^*(\alpha_1, \lambda). \quad (4.61)$$

Os estimadores de máxima verossimilhança de α_2 e λ para qualquer α_1 são denotados por $\tilde{\alpha}_2(\alpha_1)$ e $\tilde{\lambda}(\alpha_1)$, respectivamente. Então $\tilde{\alpha}_2(\hat{\alpha}_1) = \hat{\alpha}_2$, $\tilde{\alpha}_2(\hat{\alpha}_{10}) = \hat{\alpha}_{20}$ e $\tilde{\lambda}(\hat{\alpha}_1) = \hat{\lambda}$. Fixa-se $\tilde{\lambda}(\hat{\alpha}_{10}) = \hat{\lambda}_0$ e $\tilde{\alpha}(\alpha_1) = (\alpha^T_1, \tilde{\alpha}_2^T(\alpha_1))^T$, então $\tilde{\alpha}(\hat{\alpha}_1) = (\hat{\alpha}_1^T, \hat{\alpha}_2^T)^T$, $\tilde{\alpha}(\alpha_{10}) = (\alpha^T_{10}, \hat{\alpha}_{20}^T)^T$ (em que $\alpha_{10} = \gamma_0$). A matriz de informação de Fisher correspondente a $l^*(\alpha_1, \lambda)$ é denotada por $\mathbf{J}^*(\alpha_1, \lambda)$ a qual suas partições são \mathbf{J}^*_{11} , $\mathbf{J}^*_{1\lambda}$ e $\mathbf{J}^*_{\lambda\lambda}$, respectivamente. Da mesma forma, $\mathbf{J} = \mathbf{J}(\alpha_1, \alpha_2)$ é particionada por \mathbf{J}_{11} , \mathbf{J}_{12} e \mathbf{J}_{22} , respectivamente. Então, tem-se o seguinte lema.

Lema 4.6.1. *Suponha que $\alpha_2 = \alpha_2(\alpha_1, \lambda)$ transforma $\alpha = (\alpha^T_1, \alpha^T_2)^T$ para $(\alpha^T_1, \lambda^T)^T$ tal que α_1 é ortogonal a λ , então, $\alpha_2 = \alpha_2(\alpha_1, \lambda)$ satisfaz*

$$\frac{\partial \alpha_2}{\partial \alpha_1^T} = -\mathbf{J}_{22}^{-1} \mathbf{J}_{21}. \quad (4.62)$$

Mais adiante, $\mathbf{J}^*(\alpha_1, \lambda)$ e $l''^*_{\lambda\lambda}$ são expressos por

$$\mathbf{J}^*_{11} = \mathbf{J}_{11} - \mathbf{J}_{12} \mathbf{J}_{22}^{-1} \mathbf{J}_{21} = (\mathbf{J}^{11})^{-1}, \quad \mathbf{J}^*_{1\lambda} = \mathbf{0}, \quad (4.63)$$

$$\mathbf{J}^*_{\lambda\lambda} = \left(\frac{\partial \alpha_2}{\partial \lambda^T}\right)^T \mathbf{J}_{22} \left(\frac{\partial \alpha_2}{\partial \lambda^T}\right), \quad (4.64)$$

e

$$l''_{\lambda\lambda}^*(\alpha_1, \tilde{\lambda}(\alpha_1)) = \left\{ \left(\frac{\partial \alpha_2}{\partial \lambda^T} \right)^T l''_{22} \left(\frac{\partial \alpha_2}{\partial \lambda^T} \right) \right\}_{(\alpha_1, \tilde{\lambda}(\alpha_1))}. \quad (4.65)$$

Através do Lema 4.6.1, pode-se adquirir a verossimilhança perfilada ajustada para α_1 da seguinte forma:

$$l_A(\alpha_1) = l_p(\alpha_1) - \frac{1}{2} \log[\det\{-l''_{\lambda\lambda}^*(\alpha_1, \tilde{\lambda}(\alpha_1))\}], \quad (4.66)$$

$$\begin{aligned} \log[\det\{l''_{\lambda\lambda}^*(\alpha_1, \tilde{\lambda}(\alpha_1))\}] &= \log[\det\{-l''_{22}^*(\alpha_1, \tilde{\alpha}_2(\alpha_1))\}] + \\ &2 \log[\det\{\partial \alpha_2 / \partial \lambda^T\}]_{(\alpha_1, \tilde{\lambda}(\alpha_1))}, \end{aligned} \quad (4.67)$$

em que $l_p(\alpha_1) = l(\alpha_1, \tilde{\alpha}_2(\alpha_1)) = l(\tilde{\alpha})$. Obtida a verossimilhança perfilada $l_A(\alpha_1)$, outros resultados importantes são conseqüências imediatas. A estatística da razão de verossimilhança ajustada LR_A pode ser obtido de (4.66) enquanto a estatística score ajustada SC_A é a aproximação de primeira ordem de LR_A .

Para testar $H_0 : \alpha_1 = \alpha_{10}$ (isto é, $\gamma = \gamma_0$), tem-se

$$\begin{aligned} LR_A &= 2\{l_A(\hat{\alpha}_1) - l_A(\alpha_{10})\} = LR - \Delta LR, \\ \Delta LR &= \log \left[\det\{-l''_{\lambda\lambda}^*(\hat{\alpha}_1, \hat{\lambda})\} / \det\{-l''_{\lambda\lambda}^*(\alpha_{10}, \hat{\lambda}_0)\} \right], \end{aligned} \quad (4.68)$$

e

$$SC_A = SC - \Delta SC,$$

em que ΔSC é a aproximação de primeira ordem de ΔLR e SC é expresso por (4.58). Esses resultados podem ser aplicados no modelo (4.54). Porém, a estatística score pode ser obtida de forma indireta sem resolver a equação (4.62). O lema abaixo resume este resultado.

Lema 4.6.2. *A estatística score baseada em (4.66) pode ser expressa por $SC_A = CS - \Delta SC$ com*

$$\Delta SC = \sum_{k=1}^q \left[\text{tr} \left\{ \mathbf{J}_{22}^{-1} \frac{\partial}{\partial \gamma_k} \mathbf{J}_{22}(\gamma, \tilde{\alpha}_2(\gamma)) \right\} - 2 \text{tr} \left\{ \frac{\partial}{\partial \alpha_2^T} (\mathbf{J}_{22}^{-1} \mathbf{J}_{21} \delta_k) \right\} \right]_{\hat{\alpha}_0} (\hat{\gamma}_k - \gamma_{0k}), \quad (4.69)$$

em que δ_k é um vetor unitário contendo 1 para a k -ésima posição e zeros para demais, $\alpha_{10} = \gamma_0 = (\gamma_{01}, \dots, \gamma_{0q})^T$ e $\hat{\alpha}_1 = \hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_q)^T$, respectivamente.

O Lema 4.6.2 pode ser usado para modelos paramétricos em geral. Para o modelo (4.54), tem-se o seguinte teorema.

Teorema 4.16

Para o modelo de dispersão da variância (4.54) e (4.55), a estatística escore ajustada para testar H_0 pode ser expressa por

$$SC_A = \frac{1}{2} s''^{-1}(\hat{\phi}_0) \{(\mathbf{d} + 2\phi^{-1}\mathbf{h})^T \mathbf{P}_c \mathbf{d}\}_{\hat{\alpha}_0} \quad (4.70)$$

em que $\mathbf{h} = (h_{11}, \dots, h_{nn})^T$, h_{ii} é o i -ésimo elemento diagonal de \mathbf{H} e \mathbf{P}_c é definida no **Teorema 4.15**.

Este teorema pode ser aplicado para testar a dispersão da variância nos modelos não-linear normal, normal inverso e gama.

Corolário 8

Para o modelo não-linear normal e normal inversa, a estatística escore ajustada (4.70) pode ser reescrita como

$$SC_A = \frac{1}{2} \{(\mathbf{u} + 2\mathbf{h})^T \mathbf{P}_c \mathbf{u}\}_{\hat{\alpha}_0}, \quad (4.71)$$

em que \mathbf{u} é definido em (4.60).

Como um caso particular do corolário anterior, (4.71) pode ser utilizado para o teste de heterocedasticidade em modelos de regressão linear e não-linear. Para obtenção da estatística de razão de verossimilhança ajustada é necessário resolver a equação (4.62), o que é praticável somente em alguns casos especiais.

Lema 4.6.3. *Para o modelo de dispersão da variância (4.54) com distribuição normal ou normal inversa, o parâmetro γ é ortogonal ao novo parâmetro $\lambda = (\boldsymbol{\psi}, \boldsymbol{\beta}^T)^T$ segundo a transformação*

$$\boldsymbol{\psi} = \phi m_g(\gamma), \quad m_g(\gamma) = \prod_{i=1}^n \{m_i(\gamma)\}^{\frac{1}{n}}. \quad (4.72)$$

A matriz de informação de Fisher correspondente para $(\gamma^T, \boldsymbol{\psi}, \boldsymbol{\beta}^T)^T$ é

$$\mathbf{J}(\gamma, \boldsymbol{\psi}, \boldsymbol{\beta}) = \text{diag}\left(\frac{1}{2} \boldsymbol{\Psi}^T \mathbf{M}^{-1} (\mathbf{I}_n - \mathbf{1}\mathbf{1}^T/n) \mathbf{M}^{-1} \boldsymbol{\Psi}, \frac{1}{2} n \boldsymbol{\psi}^{-2}, \boldsymbol{\psi} m_g^{-1} \dot{\boldsymbol{\theta}}^T \mathbf{V} \mathbf{M} \dot{\boldsymbol{\theta}}\right). \quad (4.73)$$

Através deste lema, obtém-se a estatística da razão de verossimilhança ajustada.

Teorema 4.17

Para os modelos normal e normal inverso, a estatística da razão de verossimilhança ajustada para H_0 no modelo (4.54) pode ser expresso por

$$LR_A = \left(1 - \frac{p-2}{n}\right)LR - \log[\det\{\mathbf{R}^2(\hat{\gamma})(\mathbf{I}_p - \mathbf{B}(\hat{\gamma}))m_g^{-1}(\hat{\gamma})\}] + \log[\det\{\mathbf{R}^2(\gamma_0)(\mathbf{I}_p - \mathbf{B}(\gamma_0))\}], \quad (4.74)$$

em que LR é dado em (4.57), $\mathbf{R}(\gamma)$ é determinado pela decomposição de Cholesky

$$\boldsymbol{\theta}^T \mathbf{V}^{\frac{1}{2}} \mathbf{M}(\gamma) \mathbf{V}^{\frac{1}{2}} \boldsymbol{\theta}' = \mathbf{R}^T(\hat{\gamma}) \mathbf{R}(\hat{\gamma})$$

e $\mathbf{B}(\gamma)$ é definido como

$$\mathbf{B}(\gamma) = \{[\mathbf{e}^T \mathbf{M}(\gamma)][\mathbf{R}^{-T}(\gamma) \boldsymbol{\theta}'' \mathbf{R}^{-1}(\gamma)]\}_{\tilde{\alpha}}.$$

5 Aplicação

Na literatura, existem vários modelos não-lineares usados para modelagem de dados. Na tabela 5.2 são apresentados alguns deles

Tabela 5.1: Alguns modelos não-lineares.

| Modelo | Equação |
|-----------------------------------|--|
| Modelos de Gompertz | $Y = \alpha \exp\{-\exp[\beta - \gamma X]\}$ |
| Modelos Logístico | $Y = \frac{\alpha}{1 + \exp(\beta - \gamma X)}$ |
| Modelo de Richards | $Y = \frac{\alpha}{\{1 + \exp(\beta - \gamma X)\}^{\frac{1}{\delta}}}$ |
| Modelo Morgan-Mercer-Flodin (MMF) | $Y = \frac{\beta\gamma + \alpha X^{\delta}}{\gamma + X^{\delta}}$ |
| Modelo Weibull-type | $Y = \alpha - \beta \exp(-\gamma X^{\delta})$ |
| Modelo de regressão assintótica | $Y = \alpha - \beta \gamma^X$ |

Os modelos não-lineares da família exponencial podem ser usados em várias áreas. Abaixo, mostramos alguns estudos onde o uso destes modelos são necessários, uma vez que, a variável resposta possui uma relação não linear com as variáveis explicativas. O modelo utilizado no exemplo é o modelo de regressão assintótica (Ratkowsky, 1983), expresso por

$$Y = \alpha - \beta \gamma^X.$$

Para ilustrar a utilização do modelo de regressão assintótica, utilizamos 7 conjuntos de

dados (Ratkowsky, 1983), os quais estudam as seguintes relações:

- (1) Comprimento (Y) versus idade (X) para as espécies *Dugong dugon Muller* capturados próximos de Townsville ;
- (2) Número de folhas por lavrador por dia (Y) versus brilho refletido (X);
- (3) Rendimento de trigo (Y) versus nível de fertilizantes (X);
- (4) Reação química, quantia de $N_2 O_5$ que é decomposto (Y) versus tempo (X);
- (5) Rendimento de trigo (Y) versus taxa de aplicação de lima (X);
- (6) Rendimento de batata (Y) versus taxa de aplicação de P_2O_2 (X);
- (7) Cinturão de árvore de borracha (Y) versus taxa de aplicação de fertilizante (X).

Abaixo, temos o gráfico de dispersão para cada estudo. Verifica-se através do gráfico que a relação entre as variáveis é não-linear. Na tabela 5.2 temos as estimativas para os parâmetros do modelo em cada banco ajustado. O programa utilizado no ajuste dos dados foi o R-project.

Tabela 5.2: Estimativas dos parâmetros dos modelos ajustados para os bancos 1 a 7, respectivamente.

| Conjunto de dados | $\hat{\alpha}$ | $\hat{\beta}$ | $\hat{\gamma}$ | $\hat{\sigma}^2$ |
|-------------------|----------------|---------------|----------------|------------------------|
| (1) | 2.666630 | 0.972536 | 0.873499 | 0.778163 (10^{-2}) |
| (2) | 0.335021 | 0.295512 | 0.983830 | 0.201878 (10^{-3}) |
| (3) | 41.70790 | 15.83880 | 0.956284 | 1.454470 (10^0) |
| (4) | 33.80230 | 26.69800 | 0.752994 | 0.349154 (10^{-1}) |
| (5) | 72.43260 | 28.25190 | 0.596790 | 1.784400 (10^0) |
| (6) | 539.0780 | 307.5420 | 0.537460 | 28.06250 (10^0) |
| (7) | 22.48700 | 1..95860 | 0.705539 | 0.587643 (10^{-2}) |

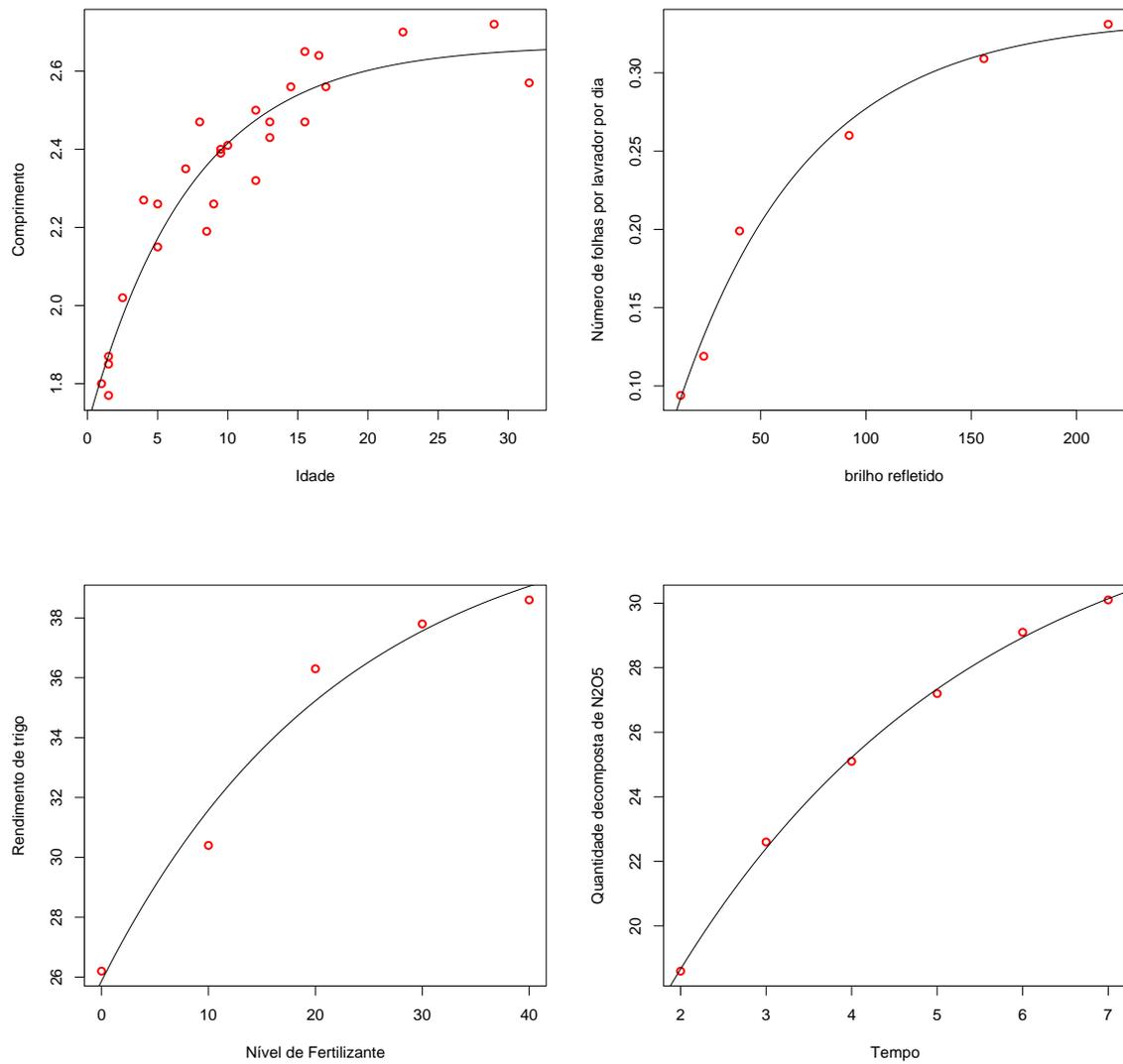


Figura 5.1: Gráfico de dispersão e reta de regressão estimada para o ajuste dos bancos de dados 1, 2, 3 e 4.

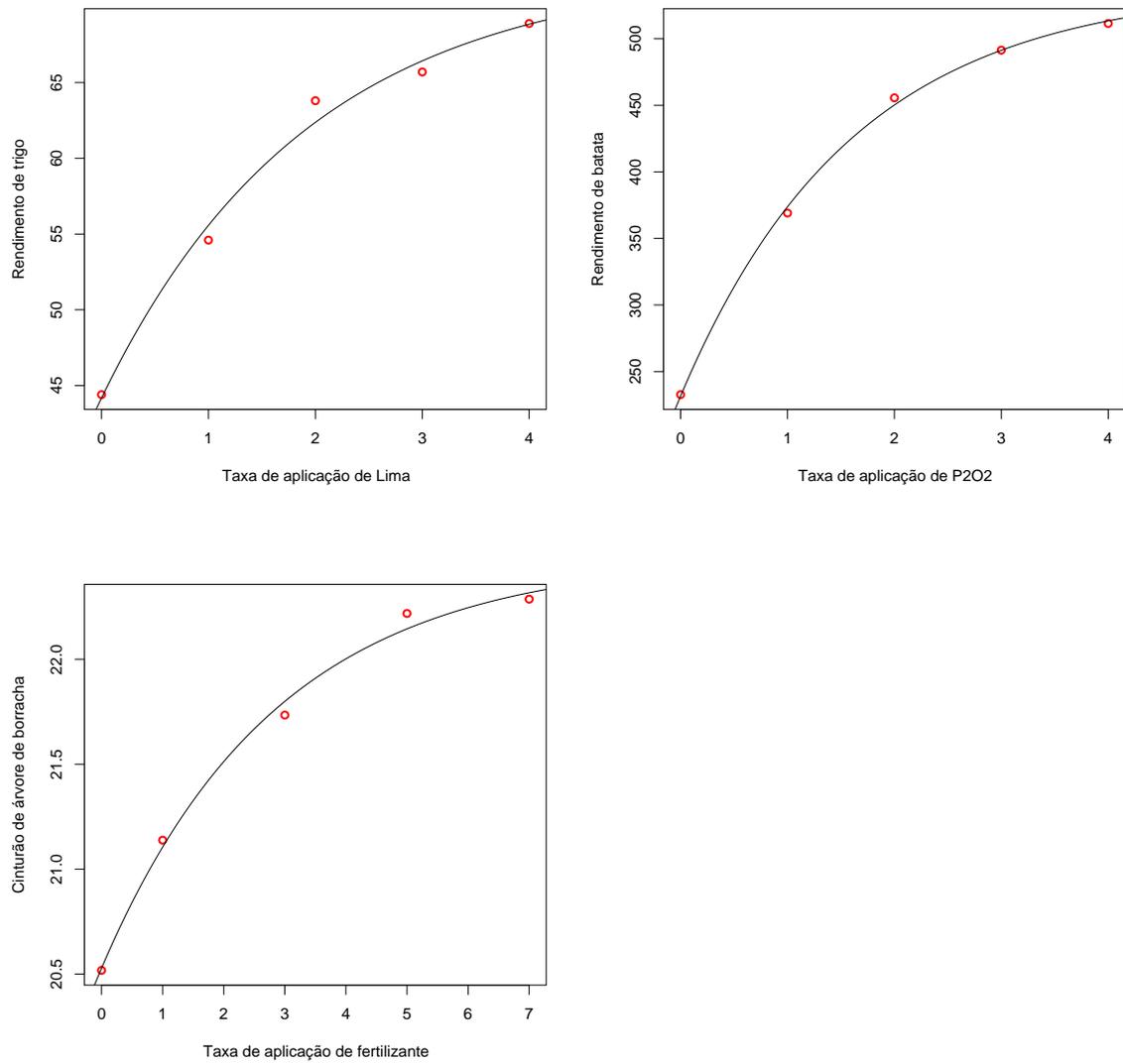


Figura 5.2: Gráfico de dispersão e reta de regressão estimada para o ajuste dos bancos de dados 5, 6 e 7.

6 Conclusões

Neste trabalho, o principal objetivo foi apresentar a teoria dos modelos não-lineares da família exponencial de uma maneira geral. Primeiramente introduziu-se a teoria dos modelos da família exponencial e seus principais resultados como: função geratriz de momentos, verossimilhança e teste razão de verossimilhança. Logo após, introduziu-se a definição de modelos não-lineares da família exponencial, suas propriedades e a estimação dos parâmetros para estes modelos. Além disso, foram apresentados métodos de análise de diagnóstico para os modelos não-lineares da família exponencial. Na aplicação, foram apresentados alguns resultados para conjunto de dados reais, mostrando que os modelos não-lineares da família exponencial abrem um leque de opções para a distribuição da variável resposta permitindo maior flexibilidade para a ligação entre a média e a componente sistemática.

Em resumo, os exemplos utilizados neste trabalho comprovam que a classe de modelagem apresentada é bastante condizente à realidade, o que garante a sua aplicabilidade em diversas situações reais. No entanto, para trabalhos futuros, é necessário verificar se alguma transformação nesta classe de modelos os deixariam mais eficazes no que diz respeito à estimação dos parâmetros e otimização do tempo de modelagem. Além disso, é importante realizar simulações para comparar os métodos de diagnósticos apresentados. Ainda, pode-se estender o estudo desses modelos em busca de correção de viés para os parâmetros, uma vez que os resultados apresentados possuem propriedades assintóticas de ordem $n-1$. Além dos possíveis estudos a serem realizados, existem uma série de idéias voltadas para o estudo dos modelos não-lineares da família exponencial que são grandes possibilidades de áreas de estudo.

Referências Bibliográficas

- Andersen, E.B. (1992). *Diagnostics in categorical data analysis*. J. Roy. Statist. Soc. Ser. **B54**, 781 - 791.
- Barndorff-Nielsen, O. E., (1978). *Information and Exponential Families in Statistical Theory*. New York: Wiley.
- Barndorff-Nielsen, O. E., (1983). *On a formula for the distribution of a maximum likelihood estimator*. Biometrika. **70**, 343 - 365.
- Bates, D.M. and Watts, D.G. (1980). *Relative curvature measures of nonlinearity*. J. Roy. Statist. Soc. Ser. **B42**, 1 - 25.
- Bates, D.M. and Watts, D.G. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.
- Belsley, D.A.; Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics*. New York: Wiley.
- Cook, R.D.. (1986). *Assessment of local influence*. J. Roy. Statist. Soc. Ser. **B48**, 133- 169.
- Cook, R.D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. London: Chapman and Hall.
- Cordeiro, G.M.. (1999). *Introdução à Teoria Assintótica*. 22º Colóquio Brasileiro de Matemática, IMPA, Rio de Janeiro.
- Cordeiro, G.M. and McCullagh, P. (1991). *Bias correction in generalized linear models*. J. Roy. Statist. Soc. Ser. **B53**, 629 - 634.
- Cordeiro, G.M. and Paula, G.A. (1989). *Improved likelihood ratio statistics for exponential family nonlinear models*. Biometrika **76**, 93 - 100.
- Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*. London: Chapman and Hall.
- Cox, D.R. and Reid, N. (1987). *Parameter Orthogonality and approximate conditional inference*. J. Roy. Statist. Soc. Ser. **B49**, 1 - 39.
- Davison, A.C. and Tsai, C.L. (1992). *Regression model diagnostics*. Internat. Statist. Rev. **60**, 337 - 355.
- Efron, B., (1986). *Double Exponential families and their use in generalized linear regression*, J. Amer. Statist. Assoc. **81**, 709 - 721.
- Escobar, L.A. and Meeker, W.Q. (1992). *Assessing influence in regression analysis with censored data*, Biometrics **48**, 507 - 528.

- Firth, D., (1991). Generalized linear models. In *Statistical Theory and Modelong*, eds Hinkley, D.V., Reid, N. and Snell, E.J. 55 - 82. London: Chapman and Hall.
- Gallant, A.R., (1987). *Nonlinear Statistical Models*. New York: Wiley.
- Hoaglim, D.C. and Welsch, R.E. (1978) *The hat matrix in regression and ANOVA*. Amer. Statist. **32**, 17 - 22.
- Jennrich, R.I. (1969). *Asymptotic properties of nonlinear least squares estimators*, Ann. Math. Statist. **40**, 633 - 643.
- Jorgensen, B., (1987). *Exponential dispersion models*. J. Roy. Statist. Soc. Ser. **B49**, 127 - 162.
- McCullagh, P., (1983). *Quasi-likelihood functions*. Ann. Statist. **11**, 59 - 67.
- McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (second edition). London: Chapman and Hall.
- Morris, C. N., (1982). *Natural exponential families with quadratic variance functions*. Ann. Statist. **10**, 65 - 80.
- Pregibon, D. (1981). *Logistic regression diagnostics*. Ann. Statist. **9**, 705 - 724.
- Ratkowsky, D.A. (1983). *Nonlinear Regression Modeling, A Unified Practical Approach*. New York: Marcel Dekker.
- Ratkowsky, D.A. (1990). *Handbook of Nonlinear Regression Models*. New York: Marcel Dekker.
- Seber, G.A.F. and Wild, C.J., (1989). *Nonlinear Regression*. New York: Wiley.
- Smyth, G.K., (1989). *Generalized linear models with varying dispersion*. J. Roy. Statist. Soc. Ser. **B51**, 47 - 60.
- St. Laurent, R.T. and Cook, R.D. (1993). *Leverage, local influence and curvature in nonlinear regression*. Biometrika. **80**, 99 - 106.
- Thomas, W. and Cook, R.D. (1989). *Assessing influence on regression coefficients in generalized linear models*. Biometrika. **76**, 741 - 749.
- Wei, B.C. (1997). *Exponential Family Nonlinear Models*. Nanjing. Springer.
- Wei, B.C. and Shi, J.Q. (1994). *On statistical models in regression diagnostics*. Ann. Statist. Math. **46**, 267 - 278.
- Wei, B.C., Lu, G. and Shi, J.Q. (1991). *Introduction to Statistical Diagnostics*. Nanjing: SE Univ. Press.
- Wu, X. and Wan, F. (1994). *A pertubation scheme for nonlinear models*. Statist. Prob. Letters **20**, 197 - 202.
- Yoshizoe, Y. (1991). *Leverage points in nonlinear regression models*. J. Japan Statist. Soc. **21**, 1 - 11.