



DORIVAL LOBATO JUNIOR

**ANÁLISE DE DIAGNÓSTICO EM MODELOS
DE REGRESSÃO NORMAL E LOGÍSTICA**

**LAVRAS – MG
2010**

DORIVAL LOBATO JUNIOR

**ANÁLISE DE DIAGNÓSTICO EM MODELOS DE REGRESSÃO
NORMAL E LOGÍSTICA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Orientador
Dr. Ruben Delly Veiga

**LAVRAS – MG
2010**

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Lobato Junior, Dorival.

Análise de diagnóstico em modelos de regressão normal e
logística / Dorival Lobato Junior. – Lavras : UFLA, 2013.
129 p. : il.

Tese (doutorado) – Universidade Federal de Lavras, 2010.
Orientador: Ruben Delly Veiga.
Bibliografia.

1. Modelos de regressão. 2. Influência local. 3. Modelos de
regressão - Técnicas de diagnósticos. 4. Modelos lineares
generalizados. I. Universidade Federal de Lavras. II. Título.

CDD – 519.536

DORIVAL LOBATO JUNIOR

**ANÁLISE DE DIAGNÓSTICO EM MODELOS DE REGRESSÃO
NORMAL E LOGÍSTICA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 13 de setembro de 2010.

Dr. Augusto Ramalho de Moraes UFLA

Dr. José Airton Rodrigues Nunes UFLA

Dr. Renato Ribeiro Lima UFLA

Dr. Ruben Delly Veiga
Orientador

**LAVRAS – MG
2010**

A minha mãe, Maria Jardelina Mendes Lobato,
por todos os valores que me ensinou, e que
contribuíram para a formação de meu caráter,
e pelo esforço de me proporcionar educação.

DEDICO

AGRADECIMENTOS

A Deus por me dar a oportunidade de me aperfeiçoar a cada dia, me tornando um ser humano melhor.

Aos meus irmãos e irmãs, que sempre estiveram ao meu lado, seja nos momentos alegres como nos tristes.

Ao meu orientador, professor Ruben Delly Veiga, pela forma paciente e dedicada com que sempre me recebeu.

Aos coorientadores, professores Renato Ribeiro de Lima e Joaquim Barbosa Queiroz, pelas sugestões e contribuições para esta tese.

Aos professores Denismar Alves Nogueira e Augusto Ramalho de Moraes, pelas sugestões e correções propostas.

A todos os meus professores do Doutorado, por todos os ensinamentos repassados durante o Curso.

Ao Coordenador do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária (UFLA), professor Dr. João Domingos Scalon.

Aos colegas de curso, Adriano, Fabrício, Verônica e Renato, pela amizade e companheirismo.

À Secretária da Pós-Graduação em Matemática da UFCG, Josi e a todos os outros funcionários do DEX-UFLA, pelo tratamento sempre cordial que nos dedicaram.

À Universidade do Estado do Pará (UEPA) e à Secretaria de Educação do Estado do Pará (SEDUC), por terem me liberado durante o transcorrer do Curso.

RESUMO

Apresentaram-se algumas técnicas de análise de diagnóstico da qualidade do ajuste em modelos de regressão normal e logística. Deu-se atenção especial ao método da influência local. Este método foi proposto, inicialmente, por Cook (1986) e tem se mostrado poderoso instrumento da análise de diagnóstico. A proposta de Cook consiste em avaliar a curvatura normal de uma superfície, com base na medida Likelihood Displacement (afastamento pela verossimilhança), sob uma pequena perturbação no modelo. Em seguida, apresentou-se a abordagem de Billor e Loynes (1993), os quais aplicaram uma medida afastamento pela verossimilhança modificada, cuja primeira derivada não se anula, exceto em casos triviais. Com isso, a inclinação máxima é utilizada como medida de influência local. Desenvolveu-se a abordagem de Billor e Loynes (1993), aplicada, inicialmente, apenas no modelo normal, para o modelo logístico. Apresentaram-se duas aplicações, nas quais foram utilizadas técnicas de análise de diagnóstico da qualidade do ajuste, por meio de gráficos e das duas propostas de influência local. No caso do modelo normal linear simples, a curvatura normal e a inclinação máxima apresentaram mesma sensibilidade na indicação de observações influentes.

Palavras-chave: Análise de regressão. Técnicas de diagnóstico. Influência local. Curvatura normal.

ABSTRACT

Presented some techniques for analyzing the diagnostic quality of fit in regression models and logistic normal. Gave special attention to the method of local influence. This method was initially proposed by Cook (1986) and has shown how powerful analysis tool for diagnosis. The proposal to Cook is to assess the normal curvature of a surface based on the likelihood measure displacement under a small perturbation in the model presented the approach of Billor and Loynes (1993), who applied a measure, the expulsion likelihood, whose first derivative is not annihilated, except in trivial cases, thus the maximum slope is used as a measure of local influence. Developed the approach Billor and Loynes (1993), used only in the normal model, for the logistic model. Presented two applications in which techniques were used to analyze the diagnostic of quality of fit, through graphics and the two proposed local influence. In the case of simple linear normal model, the maximum normal curvature and slope showed is the same sensitivity in the identification of influential observations.

Keywords: Regression analysis. Diagnostic techniques. Local Influence. Normal curvature.

SUMÁRIO

1	INTRODUÇÃO.....	9
2	MATERIAL E MÉTODOS.....	13
2.1	Análise de diagnóstico em modelos de regressão.....	13
2.1.1	Regressão normal.....	13
2.1.2	Modelos lineares generalizados.....	17
2.1.3	Regressão logística.....	22
2.2	Avaliação do modelo ajustado.....	25
2.2.1	Resíduos.....	26
2.2.2	Esquemas de perturbação.....	29
2.2.3	Influência.....	30
2.2.4	Técnicas gráficas.....	31
2.3	Avaliando a influência local através da curvatura normal..	33
2.3.1	Gráficos de influência.....	34
2.3.1.1	A curvatura normal em um gráfico de influência.....	38
2.3.1.2	A curvatura normal de uma <i>lifted line</i>	40
2.3.1.3	Uma forma operacional para a curvatura normal.....	43
2.3.1.4	Avaliando a curvatura normal em um ponto diferente de ω_0	47
2.4	Perturbação no modelo normal linear.....	52
2.4.1	Avaliando coeficientes individuais.....	59
2.4.2	Perturbação de observações na família exponencial.....	69
2.4.3	Observações individuais em regressão logística.....	74
2.5	Avaliando a influência local através da inclinação máxima de uma <i>lifted line</i>	77
2.6	Medida de afastamento pela verossimilhança modificada....	79
2.6.1	Avaliando a inclinação máxima $\alpha_1 \max^*$	90
2.6.2	Inclinação máxima no caso de perturbação de observação no modelo normal linear.....	92
2.6.3	Inclinação máxima no caso de perturbação de observação no modelo logístico.....	95
3	RESULTADOS E DISCUSSÃO.....	97
3.1	Aplicações.....	97
3.2	Estimativa do número de gansos por bando.....	97
3.3	Preferência por um novo detergente em relação a um detergente padrão.....	105
4	CONCLUSÃO.....	115
	REFERÊNCIAS.....	117
	APÊNDICE.....	119

1 INTRODUÇÃO

Modelos estatísticos são instrumentos extremamente úteis para auxiliar a compreensão de aspectos essenciais de um conjunto de dados. Porém, quase sempre, são descrições aproximadas de processos mais complicados e, conseqüentemente, uma avaliação do modelo adotado se faz necessária.

O desenvolvimento de métodos que avaliem a qualidade de um ajuste de regressão tem sido objeto da atenção dos estatísticos. Esses métodos se inserem no ramo da estatística que é chamado de análise de diagnóstico, que se iniciou com a análise de resíduos para detectar a presença de pontos extremos e avaliar a adequação da distribuição proposta para a variável resposta (BILLOR; LOYNES, 1993).

Um tópico importante na análise de diagnóstico é a detecção de pontos que exercem um peso desproporcional nas estimativas dos parâmetros do modelo. A deleção de pontos talvez seja o método mais conhecido para avaliar o impacto da retirada de uma observação particular nas estimativas da regressão (PAULA, 2004).

Belsley, Kuh e Welsch (1980) e Cook (1977) propõem métodos de diagnósticos para avaliar a influência de esquemas de perturbação por ponderação em modelos de regressão normal linear, de forma que os pesos associados às observações assumem valores 0 ou 1, indicando que a mesma foi deletada ou mantida no conjunto de dados, respectivamente. Essas ideias foram adaptadas para regressão logística por Pregibon (1981).

Contudo, a proposta mais inovadora na área foi apresentada por Cook (1986), consistindo em um método bastante geral para avaliar a influência conjunta das observações sob pequenas mudanças (perturbações) no modelo, ao invés da avaliação pela retirada individual ou conjunta de pontos. Essa

metodologia, denominada influência local, teve grande aceitação entre os pesquisadores e usuários de regressão (PAULA, 2004).

Billor e Loynes (1993) descreveram algumas dificuldades práticas associadas com a metodologia apresentada por Cook (1986) e apresentaram uma proposta alternativa para avaliar a influência local em modelos de regressão normal, procurando superar os problemas encontrados com a aplicação da metodologia sugerida por Cook (1986). Tanto a metodologia de Cook (1986) como a de Billor e Loynes (1993) utilizam a medida *likelihood displacement* (afastamento pela verossimilhança) como função objetivo.

A proposta, neste trabalho, é apresentar algumas técnicas gráficas de análise de diagnóstico e o método da influência local por meio das abordagens de Billor e Loynes (1993) e Cook (1986), expandindo-se esta última para a aplicação em modelos de regressão logísticos, mostrando-se o desenvolvimento teórico, assim como apresentando aplicações práticas de ambas. No segundo capítulo, trata-se dos modelos de regressão e de algumas técnicas de diagnóstico. Na primeira seção, enfatizam-se o modelo normal, os modelos lineares generalizados (MLGs) e, em particular, o modelo logístico linear, expondo-se as suposições básicas para cada um dos modelos e os métodos de estimação dos parâmetros. Na segunda seção, definem-se os principais resíduos utilizados e discutem-se suas aplicações no modelo normal linear e, nos MLGs, verificam-se os esquemas de perturbação e avalia-se a influência de uma observação por meio da medida *likelihood displacement* (LD) e da influência local proposta por Cook (1986). Finaliza-se o capítulo descrevendo-se os principais tipos de gráficos de diagnóstico utilizados para o modelo normal e para os MLGs em geral. No terceiro capítulo, apresenta-se a formulação da influência local avaliada por meio da curvatura normal, define-se e constrói-se a noção de gráfico de influência como uma superfície geométrica e fundamenta-se a ideia de como utilizar a curvatura normal para medir a influência causada por uma perturbação

no modelo. Ainda, avalia-se o modelo normal sob um esquema de perturbação de observação individual, mostrando-se a forma da curvatura normal para este caso. De forma geral, encontra-se a curvatura normal para os MLGs, sob o esquema de perturbação de uma observação individual e procede-se do mesmo modo para o caso do modelo logístico linear. No quarto capítulo, apresenta-se a formulação da influência local avaliada por meio da primeira derivada e define-se uma versão da medida *likelihood displacement* modificada (LD^*), de modo que sua primeira derivada não se anule.

Considerando-se a matriz de perturbação geral

$$Wg = \text{diag} \{1 + \omega_1, 1 + \omega_2, \dots, 1 + \omega_n\}$$

para o modelo normal

$$y = X\beta + \varepsilon$$

em que $\text{Var}(\varepsilon) = \delta^2 Wg^{-1}$, encontra-se o gradiente da superfície $(\omega_1, \omega_2, \dots, \omega_n, LD^*(\omega))$ e, conseqüentemente, a inclinação máxima. Em seguida, obtém-se um valor de referência para a inclinação máxima. Supondo-se o esquema de perturbação de uma observação individual no modelo normal, encontram-se a inclinação máxima e um valor de referência para ela. Finaliza-se este capítulo, obtendo-se o gradiente e, conseqüentemente, a inclinação máxima para o modelo logístico linear, sob o esquema de perturbação de uma observação individual.

No quinto capítulo apresentam-se duas aplicações. Na primeira, utiliza-se o modelo normal, acompanhado da utilização de algumas técnicas gráficas para detectar pontos influentes e verificar a validade da suposição de normalidade para a variável resposta. Também se avalia a influência local por meio da curvatura normal e da inclinação máxima, sob o esquema de perturbação de uma observação individual. Obtém-se, ainda, a inclinação máxima sob o esquema de perturbação da variância, exposto no capítulo 3. Na outra aplicação adota-se o modelo logístico linear e algumas técnicas gráficas de diagnóstico, e avalia-se a influência local por meio da curvatura normal e da

inclinação máxima, obtidas sob um esquema de perturbação de observação individual para o modelo logístico.

No apêndice A apresentam-se definições, teoremas, proposições e alguns resultados importantes, utilizados no transcorrer desta tese e, no apêndice B, descrevem-se os comandos para se obter, no aplicativo R, os resultados mais importantes que foram utilizados no capítulo 5.

2 MATERIAL E MÉTODOS

2.1 Análise de diagnóstico em modelos de regressão

Em síntese, pode-se dizer que o objetivo principal da análise de regressão é estudar o efeito que algumas variáveis exercem sobre outras, o que consiste na obtenção de uma relação funcional entre uma das variáveis e as restantes. Desse modo, podem-se identificar dois tipos de variáveis, uma variável dependente (ou variável resposta) e as outras, independentes (ou explicativas, ou explanatórias ou covariáveis). A variável dependente é aquela cujo comportamento deseja-se avaliar e as independentes são utilizadas para explicar a sua variabilidade.

2.1.1 Regressão normal

Considerar-se-á, inicialmente, o caso mais simples, em que se têm uma variável resposta Y e uma única variável independente X , admitindo-se que a relação entre a média condicional de Y dado X , é $\mu = \mu(x) = E[Y|X = x]$, ou da forma

$$\mu = \mu(x) = \alpha + \beta x . \quad (2.1)$$

Esta equação (1.1) é denominada **componente sistemática** do modelo; α e β são parâmetros desconhecidos a serem estimados. Além da componente sistemática, o modelo é composto de uma parte não determinística, devido a fatores não observáveis, denominada **componente aleatória** e denotada por uma variável aleatória ε . Dessa forma, o modelo completo fica estabelecido pela seguinte equação (MAGALHÃES, 2006):

$$Y = \alpha + \beta X + \varepsilon . \quad (2.2)$$

O **modelo de regressão normal simples** é obtido a partir das seguintes suposições assumidas para a componente aleatória \mathcal{E} :

$$E(\mathcal{E}) = 0$$

$$\text{Var}(\mathcal{E}) = \sigma^2$$

$$\mathcal{E} \sim N(0, \sigma^2).$$

Como consequência das suposições feitas para a variável aleatória \mathcal{E} , tem-se que $Y \sim N(\alpha + \beta x, \sigma^2)$.

Para a obtenção dos **estimadores** dos parâmetros α e β , considera-se uma amostra de pares de observações $(X = x_i, Y = y_i), i = 1, 2, \dots, n$. Pela expressão de (2.2), pode-se escrever, para cada par,

$$y_i = \mu(x_i) + \varepsilon_i = \alpha + \beta x_i + \varepsilon_i$$

em que os $\varepsilon_{i\text{'s}}$ são não correlacionados e $\varepsilon_i \sim N(0, \sigma^2)$, para todo i . Em busca da reta que melhor represente os n pontos amostrais, utilizou-se o método dos mínimos quadrados. Minimiza-se a função

$$S = S(\alpha, \beta) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [y_i - \alpha - \beta x_i]^2 \quad (2.3)$$

Segue daí que se precisam obter as derivadas parciais $\frac{\partial S}{\partial \alpha}$ e $\frac{\partial S}{\partial \beta}$ e, em seguida, as igualam-se a zero. Com isso, encontra-se o sistema de equações

$$-2 \sum_{i=1}^n [y_i - (\hat{\alpha} + \hat{\beta} x_i)] = 0 \quad \text{e} \quad -2 \sum_{i=1}^n [y_i - (\hat{\alpha} + \hat{\beta} x_i)] x_i = 0$$

em que $\hat{\alpha}$ e $\hat{\beta}$ são os valores que minimizam a função (2.3), ou seja, são os **estimadores mínimos quadrados** para os parâmetros α e β , respectivamente. Simplificando-se as equações desse sistema, obtém-se o sistema de **equações normais**

$$\begin{cases} n\hat{\alpha} + \hat{\beta} \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ \hat{\alpha} \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{cases} \quad (2.4)$$

Resolvendo o sistema (2.4), acha-se a solução:

$$\hat{\alpha} = \frac{1}{n} \sum_{i=1}^n y_i - \frac{\hat{\beta}}{n} \sum_{i=1}^n x_i$$

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2}$$

Segue daí que a reta de regressão estimada pode ser expressa por

$$\hat{\mu}(x) = \hat{\alpha} + \hat{\beta}x .$$

É comum encontrar, na literatura, esta equação na forma

$$\widehat{y}_i = \hat{\alpha} + \hat{\beta}x_i, \text{ para } i = 1, 2, \dots, n ;$$

em que os \widehat{y}_i são chamados de **valores preditos**.

No **modelo de regressão normal múltipla**, no qual se utiliza mais de uma variável independente, considera-se que a média condicional da variável resposta Y pode ser expressa em função das variáveis explicativas X_1, X_2, \dots, X_p , por meio da relação linear (MAGALHÃES, 2006)

$$\begin{aligned} \mu &= E[Y|X_1 = x_1, \dots, X_p = x_p] \\ &= \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p. \end{aligned}$$

A componente aleatória do modelo também é representada por uma variável aleatória ε , que segue uma distribuição normal com média zero e variância σ^2 .

Assim, para cada observação do tipo $(y_i, x_{1i}, x_{2i}, \dots, x_{pi})$, tem-se o seguinte modelo:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \varepsilon_i, \text{ para } i = 1, 2, \dots, n, \quad p \ll n, \quad (2.5)$$

em que $\varepsilon_i \sim N(0, \sigma^2)$, para todo i .

Considera-se a matriz de variáveis independentes

$$X = \begin{bmatrix} \mathbf{1} & x_{11} & \dots & x_{p1} \\ \mathbf{1} & x_{12} & \dots & x_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{1} & x_{1n} & \dots & x_{pn} \end{bmatrix},$$

e os vetores

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \text{ e } \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad (2.6)$$

Dessa forma, pode-se escrever o modelo (2.5), matricialmente, por

$$y = X\beta + \varepsilon. \quad (2.7)$$

Para o modelo múltiplo de (1.5), a função S, definida em (2.3), pode ser escrita como

$$S(\beta) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1i} - \dots - \beta_p x_{pi})^2.$$

Derivando-se a função acima em relação a $\beta_0, \beta_1, \dots, \beta_p$ e igualando-se a zero essas derivadas, obtém-se o sistema de equações normais a seguir.

$$\begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n x_{1i} + \dots + \hat{\beta}_p \sum_{i=1}^n x_{pi} = \sum_{i=1}^n y_i \\ \hat{\beta}_0 \sum_{i=1}^n x_{ji} + \hat{\beta}_1 \sum_{i=1}^n x_{1i}x_{ji} + \dots + \hat{\beta}_p \sum_{i=1}^n x_{pi}x_{ji} = \sum_{i=1}^n y_i x_{ji}, \quad j = 1, \dots, p. \end{cases}$$

Uma forma matricial para o sistema acima é a seguinte:

$$(X^t X)\hat{\beta} = X^t y,$$

Como $(X^t X)$ é uma matriz quadrada de dimensões $(p+1) \times (p+1)$, então, existe a sua matriz inversa, pois, de acordo com Graybill (1983), a matriz

X tem posto coluna completo e igual a $p + 1$. Assim, o estimador de mínimos quadrados do vetor de parâmetros β , é dado por

$$\hat{\beta} = (X^t X)^{-1} X^t y,$$

O vetor com os valores preditos pode ser escrito como

$$\begin{aligned} \hat{y} &= X\hat{\beta} \\ &= X(X^t X)^{-1} X^t y \\ &= Hy, \end{aligned}$$

em que $H = X(X^t X)^{-1} X^t$ é conhecida como matriz *hat* e representa a matriz de projeção ortogonal de \mathbf{R}^n sobre o subespaço vetorial gerado pelas colunas da matriz X .

2.1.2 Modelos lineares generalizados

Durante muitos anos, os modelos normais foram utilizados para descrever a maioria dos fenômenos aleatórios. Mesmo quando o fenômeno sob estudo não apresentava uma resposta para a qual fosse razoável a suposição de normalidade, usava-se algum tipo de transformação, no sentido de alcançar a normalidade (PAULA, 2004).

Com o desenvolvimento computacional, novos modelos têm sido propostos, dentre os quais se destacam os modelos lineares generalizados (MLGs), apresentados por Nelder e Wedderburn (1972). Essa proposta permite que a distribuição da variável resposta pertença à **família exponencial de distribuições** (ver apêndice A), significando que cada componente do vetor de observações Y tenha a função densidade ou função de probabilidade na forma

$$f(y, \theta, \phi) = \exp\{\phi[y\theta - b(\theta)] + c(y, \phi)\}, \quad (2.10)$$

em que os θ 's são os parâmetros de interesse e ϕ é o parâmetro de dispersão. Nessa classe são válidas as seguintes propriedades:

$$E[y_i] = \mu_i = b'(\theta_i)$$

$$Var[y_i] = \phi^{-1}V_i$$

$$V_i = b''(\theta_i) = \frac{d\mu_i}{d\theta_i}$$

em que b' e b'' representam, respectivamente, a primeira e a segunda derivadas da função $b(\theta)$ e V_i é definida como função de variância. A função de variância desempenha um papel importante na família exponencial, uma vez que ele caracteriza a distribuição, ou seja, dada a função de variância, tem-se uma classe de distribuições correspondentes e vice-versa (PAULA, 2004).

Com relação à componente sistemática do modelo, as variáveis independentes X_1, X_2, \dots, X_p produzem o vetor $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)'$ como **preditor linear**, que, no caso da regressão normal linear múltipla, é dada por

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \dots + \beta_p x_{pi} \\ &= \sum_{j=0}^p x_{ji} \beta_j \\ &= \mathbf{x}_i' \boldsymbol{\beta}, \text{ com } i = 1, 2, \dots, n \end{aligned} \quad (2.11)$$

em que o vetor $\mathbf{X}_i' = (1, x_{1i}, x_{2i}, \dots, x_{pi})$ é a i -ésima linha da matriz \mathbf{X} . Além disso, a ligação entre a parte sistemática e a média condicional da variável resposta se dá por meio da relação funcional entre a parte sistemática e a média condicional da variável resposta se dá por meio da seguinte relação:

$$\mu_i = g^{-1}(\eta_i), \quad i = 1, 2, \dots, n$$

em que $g(\cdot)$ é denominada de função de ligação, que deve ser monótona e diferenciável. No caso da regressão normal, é a função identidade (MACULLAGH; NELDER, 1989).

Nos MLGs, com respostas independentes, a log-verossimilhança pode ser escrita como

$$l(\beta; y) = \sum_{i=1}^n \phi\{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi). \quad (2.12)$$

Um caso particular ocorre quando

$$\theta_i = \eta_i = \sum_{j=0}^p x_{ij} \beta_j. \quad (2.13)$$

Nesse caso, tem-se que

$$l(\beta; y) = \sum_{j=0}^p s_j \beta_j - \phi \sum_{i=1}^n b(\theta_i) + \sum_{i=1}^n c(y_i, \phi),$$

em que $s_j = \phi \sum_{i=1}^n y_i x_{ij}$, com $j = 1, 2, \dots, p$.

Pelo teorema da fatoração, conforme Bolfarine e Sandoval (2001), a estatística $\mathbf{S} = (s_0, s_1, s_2, \dots, s_p)^T$ é suficiente minimal para o vetor $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^T$. As funções de ligação, tais que

$$g(\mu_i) = \sum_{j=0}^p s_j \beta_j. \quad (2.14),$$

são chamadas de ligações canônicas (DEMÉTRIO, 2002). As ligações canônicas para os modelos Normal, Poisson e Binomial são, respectivamente, dadas por

$$\mu = \eta,$$

$$\log \mu = \eta,$$

$$\log \left\{ \frac{\mu}{1 - \mu} \right\} = \eta.$$

A função score e a matriz de informação de Fisher para o vetor de parâmetros $\boldsymbol{\beta}$, nos modelos lineares generalizados, são dadas, respectivamente, por

$$U(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta}, y)}{\partial \boldsymbol{\beta}}$$

$$\begin{aligned}
&= \phi X^T W^{1/2} V^{-1/2} (y - \mu) \\
K(\beta) &= E \left\{ -\frac{\partial^2 l(\beta, y)}{\partial \beta} \right\} \\
&= \phi X^T W X \quad (2.16)
\end{aligned}$$

em que $W = \text{diag}[w_1, \dots, w_n]$ é a matriz de pesos com componentes

$$w_i = \frac{\left(\frac{d\mu_i}{d\eta_i}\right)^2}{V_i}, \quad V = \text{diag}[V_1, \dots, V_n], \quad \mu = (\mu_1, \dots, \mu_n)^T \text{ e } V_i \text{ é a função de}$$

variância. Em particular para a ligação canônica, essas quantidades tomam as formas simplificadas

$$U(\beta) = \phi X^T (y - \mu)$$

$$K(\beta) = \phi X^T V X,$$

respectivamente (PAULA, 2004).

Para se obter o estimador de máxima verossimilhança para o vetor de parâmetros β , deve-se encontrar a solução do sistema com $p + 1$ equações não lineares dado por

$$U(\beta) = 0.$$

Em geral, esse sistema não admite uma solução explícita. Segue que, expandindo-se a função escore $U(\beta)$ em série de Taylor até o termo de primeira ordem, em torno de um valor inicial $\beta^{(0)}$, tem-se

$$U(\beta) \cong U(\beta^{(0)}) + U'(\beta^{(0)})[\beta - \beta^{(0)}],$$

em que U' representa a primeira derivada de $U(\beta)$ em relação a β . Logo, tem-se que

$$U(\beta) = U'(\beta^{(0)})\tilde{\beta} - U'(\beta^{(0)})\beta^{(0)} \approx 0 \Rightarrow$$

$$U'(\beta^{(0)})\tilde{\beta} \approx U'(\beta^{(0)})\beta^{(0)} - U(\beta^{(0)}).$$

Admitindo-se que $U'(\beta^{(0)})$ é não singular, e, portanto, possui inversa $[U'(\beta^{(0)})]^{-1}$, segue que

$$[U'(\beta^{(0)})]^{-1} U'(\beta^{(0)}) \hat{\beta} \approx [U'(\beta^{(0)})]^{-1} U'(\beta^{(0)}) \beta^{(0)} - [U'(\beta^{(0)})]^{-1} U'(\beta^{(0)}) \Rightarrow$$

Aplicando-se a propriedade da matriz inversa, tem-se

$$\hat{\beta} \approx \beta^{(0)} [-U'(\beta^{(0)})]^{-1} U'(\beta^{(0)}).$$

A partir de um valor inicial para $\hat{\beta}$, a equação acima pode ser utilizada para se obter um processo iterativo, na seguinte forma:

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + [-U'(\beta^{(m)})]^{-1} U(\beta^{(m)}), \quad m = 0, 1, 2, \dots$$

Um dos critérios utilizados para que esse processo pare é dado por

$$\max_j |[\beta^{(j)'}(m+1) - \beta^{(j)'}(m)]| < \epsilon,$$

em que ϵ é um valor pré-determinado. Esse procedimento é conhecido como processo iterativo de Newton-Raphson e está descrito em Hoffman e Kunze (1979).

Como a matriz $U'(\beta)$ pode não ser positiva definida, a aplicação do método scoring de Fisher, que consiste em substituir a matriz $U'(\beta)$ pelo seu correspondente valor esperado, pode ser mais conveniente, segundo Paula (2004). Dessa forma,

$$\hat{\beta}^{(m+1)} = \hat{\beta}^{(m)} + [K(\hat{\beta}^{(m)})]^{-1} U(\beta^{(m)}), \quad m = 0, 1, 2, \dots$$

Substituindo-se os resultados (2.15) e (2.16) na equação acima, tem-se

$$\begin{aligned} \hat{\beta}^{(m+1)} &= \hat{\beta}^{(m)} + \frac{1}{\phi} (X^t W X)^{-1} \phi X^t W^{1/2} V^{1/2} (y - \mu^{(m)}) \\ &= (X^t W X)^{-1} (X^t W X) \hat{\beta}^{(m)} + (X^t W X)^{-1} X^t W^{1/2} W^{1/2} W^{-1/2} V^{1/2} (y - \mu^{(m)}) \\ &= (X^t W X)^{-1} X^t W \left\{ \eta^{(m)} + W^{-1/2} V^{1/2} (y - \mu^{(m)}) \right\} \\ &= (X^t W X)^{-1} X^t W \tilde{\eta}^{(m)} \end{aligned}$$

em que $\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \mathbf{W}^{-1/2} \mathbf{V}^{1/2} (\mathbf{y} - \boldsymbol{\mu}^{(m)})$ e $\boldsymbol{\eta}^{(m)} = \mathbf{X} \tilde{\boldsymbol{\beta}}^{(m)}$. Note que \mathbf{z} desempenha o papel de uma variável dependente modificada e \mathbf{W} é uma matriz de pesos, que muda a cada passo do processo iterativo. A convergência de (2.17) ocorre em um número finito de passos, independente dos valores iniciais utilizados.

2.1.3 Regressão logística

A regressão logística tem sido bastante utilizada. Mesmo quando a resposta de interesse não é do tipo binária, alguns pesquisadores têm dicotomizado a resposta, de modo que a probabilidade de sucesso possa ser modelada por meio da regressão logística (PAULA, 2004).

Primeiro, tratar-se-á da **regressão logística simples**. Seja Y a proporção de sucessos em n ensaios de Bernoulli independentes, cada um com probabilidade p de ocorrência de sucesso. Assume-se também que $nY \sim b(n, p)$. Com isso, a sua função de probabilidade fica estabelecida por

$$\binom{n}{ny} p^{ny} (1-p)^{n-ny} = \exp \left\{ \ln \binom{n}{ny} p^{ny} + ny \ln \left(\frac{p}{1-p} \right) = n \ln(1-p) \right\}$$

em que $0 < p < 1$ e $y \in \left\{ 0, \dots, \frac{1}{n}, \dots, 1 \right\}$.

Uma função de ligação, denominada **logit**, é definida por

$$\text{logit} : (0,1) \mapsto \mathbf{R}$$

$$t \mapsto \text{logit}(t) = \ln \frac{t}{1-t}. \quad (2.18)$$

Dessa forma, obtém-se o modelo logístico linear simples

$$\text{logit } p(x) = \eta, \quad (2.19)$$

em que $\eta = \beta_0 + \beta_1 x$ é o preditor linear, e β_0 e β_1 são parâmetros desconhecidos. Segue, de (2.18) e (2.19), que

$$\eta = \ln \left[\frac{p(x)}{1 - p(x)} \right],$$

Daí obtém-se que

$$p(x) = \frac{e^\eta}{1 + e^\eta},$$

ou, mais precisamente,

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}. \quad (2.20)$$

Os parâmetros β_0 e β_1 podem ser estimados pelo método da **máxima verossimilhança** (COOK; WEISBERG, 1982). Para isso, considerar-se-á uma amostra aleatória de n observações independentes $(y_i, x_i), i = 1, 2, \dots, n$, em que x_i é um valor variável independente, discreta ou contínua. Tem-se que $p(x_i) = P(Y = 1|X = x_i)$ é a probabilidade de sucesso; assim, $1 - p(x_i)$ é a probabilidade de fracasso, ou seja, $P(Y = 0|X = x_i)$.

A contribuição da i -ésima observação para a função de verossimilhança é

$$p(x_i)^{y_i} [1 - p(x_i)]^{1-y_i}.$$

Dessa forma, obtém-se a função de verossimilhança para observações independentes,

$$L(\beta_0, \beta_1; y_1, y_2, \dots, y_n) = \prod_{i=1}^n p(x_i)^{y_i} [1 - p(x_i)]^{1-y_i}. \quad (2.21)$$

Segue daí que a função de log-verossimilhança é dada por

$$l(\beta_0, \beta_1; y_1, y_2, \dots, y_n) = \sum_{i=1}^n \{y_i \ln[p(x_i)] + (1 - y_i) \ln[1 - p(x_i)]\}. \quad (2.22)$$

Diferenciando-se (2.22) com relação a β_0 e a β_1 e igualando-se cada um desses resultados a zero, encontra-se o sistema com as equações

$$\begin{cases} \sum_{i=1}^n \left[y_i - \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} \right] = 0 \\ \sum_{i=1}^n \left[y_i - \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 x_i}} \right] x_i = 0 \end{cases}$$

O sistema acima não admite uma solução explícita; para obter uma solução aproximada, deve-se utilizar o processo iterativo de Newton-Raphson.

A seguir aborda-se o caso da **regressão logística múltipla**.

Considere-se o modelo geral de regressão logística múltipla

$$\ln \left[\frac{p(x)}{1 - p(x)} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, \quad (2.23)$$

em que $\mathbf{x} = (1, x_1, \dots, x_p)^T$ contém os valores observados de p variáveis explicativas e o valor 1 está associado ao intercepto β_0 . A ligação *logit*, neste caso, é dada por

$$\text{logit } p(x) = \eta,$$

em que $p(x) = P[Y = \mathbf{1} | \mathbf{x}]$ é a probabilidade condicional de sucesso dado o vetor \mathbf{x} , $\eta = \mathbf{x}^T \boldsymbol{\beta}$ é o preditor linear e $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ é o vetor de parâmetros desconhecidos.

Adotando-se um procedimento análogo ao utilizado no modelo simples, obtém-se

$$p(x) = \frac{e^\eta}{1 + e^\eta} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}. \quad (2.24)$$

Para se obter o estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\beta}$, no modelo logístico múltiplo, desenvolve-se um procedimento análogo ao utilizado para o caso simples. Sendo assim, a função de verossimilhança obtida a partir de uma amostra com n observações independentes $(y_i, x_{1i}, x_{2i}, \dots, x_{pi}), i = 1, 2, \dots, n$, é dada por

$$L(\beta; y) = \prod_{i=1}^n p(x_i)^{y_i} [1 - p(x_i)]^{1-y_i}, \quad (2.25)$$

em que $x_i = (1, x_{1i}, x_{2i}, \dots, x_{pi})^T, i = 1, 2, \dots, n$. Segue que, diferenciando-se a função de log-verossimilhança $l(\beta; y)$ em relação a cada componente do vetor de parâmetros β e após igualando-se essas derivadas a zero, obtém-se o sistema

$$\begin{cases} \sum_{i=1}^n [y_i - \hat{p}(x_i)] = 0 \\ \sum_{i=1}^n [y_i - \hat{p}(x_i)] x_{ij} = 0, \quad \text{para } j = 1, 2, \dots, p. \end{cases}$$

Este sistema também não admite uma solução explícita, utilizando-se o processo iterativo de Newton-Raphson para se obter uma solução aproximada.

2.2 Avaliação do modelo ajustado

Uma das formas de avaliação de modelos ajustados visa à detecção de pontos que se encontram significativamente mais afastados dos demais. Estes pontos são, usualmente, denominados de *outliers* e formam a base para o desenvolvimento de **técnicas de diagnósticos** do ajuste de modelos de regressão.

Pode-se classificar um *outlier*, conforme sua característica, do seguinte modo (PAULA, 2004):

- a) **pontos aberrantes:** são pontos que têm certa influência sobre os valores ajustados, embora não estejam muito afastados dos demais pontos;
- b) **pontos de alavanca ou de alto leverage:** são pontos que estão mais afastados dos pontos no subespaço gerado pelas colunas da matriz \mathbf{X} . Estes pontos não influenciam de forma significativa as estimativas dos parâmetros, porém, fazem com que as variâncias dos valores ajustados

dos pontos próximos a ele sejam maiores do que as variâncias dos valores ajustados correspondentes aos demais pontos;

- c) **pontos influentes:** são pontos que, ao mesmo tempo, são de alto *leverage* e aberrantes.

2.2.1 Resíduos

O **resíduo** para a i -ésima observação é definido como uma função do tipo $r_i = r(y_i, \hat{\mu}_i)$, cujo objetivo é medir a discrepância entre o valor observado e o valor ajustado da i -ésima observação $\hat{\mu}_i$. A definição mais usual é a de **resíduo ordinário**, dada por

$$r_i = y_i - \hat{\mu}_i, \quad (2.26)$$

que mede a diferença entre o valor observado e o valor ajustado para a i -ésima observação.

Para $\mathbf{r}_i = (r_{1i}, r_{2i}, \dots, r_{ni})^T$, tem-se que, para o modelo de regressão normal linear,

$$\begin{aligned} \mathbf{r} &= \mathbf{y} - \mathbf{H}\mathbf{y} \\ &= (\mathbf{I} - \mathbf{H})\mathbf{y}, \end{aligned}$$

em que $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ é a matriz *hat* já definida na seção (2.1.1). O elemento h_{ii} , definido a seguir, pertence à diagonal da matriz \mathbf{H} e desempenha importante papel na construção de técnicas de diagnóstico

$$h_{ii} = \mathbf{x}_i^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_i. \quad (2.27)$$

Em particular, $\text{Var}(r_i) = (1 - h_{ii})\sigma^2$. Partindo do pressuposto de que todos os pontos exercem a mesma influência sobre os valores ajustados, espera-se que o valor de h_{ii} esteja próximo de $\frac{p}{n}$. Convém, então, examinar aqueles

pontos, tais que $h_{ii} \geq \frac{2p}{n}$, que podem ser diagnosticados como pontos de **alavanca** ou de **alto leverage** e, geralmente, estão localizados em regiões remotas no subespaço gerado pelas colunas da matriz \mathbf{X} (PAULA, 2004).

Em relação ao modelo de regressão normal linear, tem-se que $r_i \sim N(0, \sigma^2)$. Assim, podem-se obter os **resíduos studentizados**, denotados por t_i , dividindo-se cada r_i pelo seu respectivo desvio padrão amostral

$s(1 - h_{ii})^{1/2}$, em que $s^2 = \sum_{i=1}^n \frac{r_i^2}{n-p}$ é o estimador da variância σ^2 . Logo,

$$t_i = \frac{r_i}{s(1 - h_{ii})^{1/2}}, \quad i = 1, \dots, n \quad (2.28).$$

Como r_i e s^2 não são independentes, t_i não segue uma distribuição t de *Student*, como se poderia esperar. Paula (2004) mostra que o problema da dependência entre r_i e s^2 pode ser contornado, substituindo-se s^2 por $s_{(i)}^2$ na expressão de t_i anterior, em que $s_{(i)}^2$ é o erro quadrático médio correspondente ao modelo sem a i -ésima observação, dado por

$$s_{(i)}^2 = s^2 \left(\frac{n-p-1}{n-p-t_i^2} \right).$$

Dessa forma, tem-se que o novo resíduo studentizado

$$t_i^* = t_i \left(\frac{n-p-1}{n-p-t_i^2} \right)^{1/2}, \quad (2.29)$$

segue uma distribuição t_{n-p-1} .

A definição de resíduo studentizado para os MLGs é feita de forma análoga à utilizada na regressão normal linear. Observa-se que, para os MLGs, os estimadores dos parâmetros do modelo são dados por

$$\hat{\beta} = (X^c \bar{W} X)^{-1} X^c \bar{W} z,$$

em que $\mathbf{z} = \hat{\boldsymbol{\eta}} + \widehat{\mathbf{W}}^{-1/2} \widehat{\mathbf{V}}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}})$. Assim, pode-se definir o resíduo ordinário por

$$\begin{aligned} \mathbf{r}^* &= \widehat{\mathbf{W}}^{-1/2} (\mathbf{z} - \hat{\boldsymbol{\eta}}) \\ &= \widehat{\mathbf{V}}^{-1/2} (\mathbf{y} - \hat{\boldsymbol{\mu}}). \end{aligned} \quad (2.30)$$

Assumindo-se que $\text{Var}(\mathbf{z}) \cong \widehat{\mathbf{W}}^{-1} \phi^{-1}$ segue que $\text{Var}(\mathbf{r}^*) \cong \phi^{-1} (\mathbf{I} - \mathbf{H})$, em que $\mathbf{H} = \mathbf{W}^{1/2} \mathbf{X} (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{1/2}$. Assim, os resíduos standartizados para os MLGs são definidos por

$$t_{Si} = \frac{\phi^{1/2} (y_i - \hat{\mu}_i)}{\sqrt{\hat{v}_i (1 - \hat{h}_{ii})}}, \quad (2.31)$$

em que \hat{h}_{ii} foi definido na equação (1.27).

Os resíduos mais utilizados em MLGs são definidos a partir dos componentes da função desvio (DEMÉTRIO, 2002). A versão padronizada (DAVISON; GIGLI, 1989; MCCULLAGH, 1987) é dada por

$$t_{Di} = \frac{\phi^{1/2} d(y_i; \hat{\mu}_i)}{\sqrt{1 - \hat{h}_{ii}}}, \quad (2.32)$$

em que $d(y_i; \hat{\mu}_i) = \pm \sqrt{2} \{y_i (\hat{\theta}_i^0 - \hat{\theta}_i) + [b(\hat{\theta}_i) - b(\hat{\theta}_i^0)]\}^{1/2}$.

Para os **modelos logísticos lineares**, esse resíduo toma a forma particular

$$t_{Di} = \pm \sqrt{\frac{2}{1 - \hat{h}_{ii}}} \left[y_i \ln \left(\frac{y_i}{n_i \hat{p}_i} \right) + (n_i - y_i) \ln \left(\frac{n_i - y_i}{n_i - n_i \hat{p}_i} \right) \right]^{1/2}.$$

com $0 < y_i < n_i$. Quando $y_i = 0$, tem-se que

$$t_{Di} = - \frac{\{2n_i \ln(1 - \hat{p}_i)\}^{1/2}}{\sqrt{1 - \hat{h}_{ii}}}$$

E, quando $y_i = n_i$, segue que

$$t_{Di} = -\frac{(2n_i \ln \hat{\beta}_i)^{1/2}}{\sqrt{1 - \hat{h}_{ii}}},$$

em que \hat{h}_{ii} assume a seguinte forma:

$$\hat{h}_{ii} = n_i \hat{\beta}_i (1 - \hat{\beta}_i) \mathbf{x}_i^t (\mathbf{X}^t \hat{\mathbf{V}} \mathbf{X})^{-1} \mathbf{x}_i$$

com $\hat{\mathbf{V}} = \text{diag}[n_1 \hat{\beta}_1 (1 - \hat{\beta}_1), \dots, n_k \hat{\beta}_k (1 - \hat{\beta}_k)]$ e $n_1 + \dots + n_k = n$.

O resíduo studentizado \mathbf{t}_{Si} , para o modelo logístico linear, é dado pela expressão a seguir:

$$t_{Si} = \frac{(y_i - n_i \hat{\beta}_i)^2}{\sqrt{(1 - \hat{h}_{ii})[n_i \hat{\beta}_i (1 - \hat{\beta}_i)]}}. \quad (2.33)$$

2.2.2 Esquemas de perturbação

De maneira geral, um esquema de perturbação pode ser definido como qualquer mudança efetuada nas suposições do modelo, nos dados observados, ou em ambos, proporcionando, com isso, uma substancial modificação nos resultados da análise (BILLOR; LOYNES, 1993).

Dessa forma, podem-se dividir os esquemas de perturbações em dois grupos:

- (i) **perturbação no modelo:** quando a suposição de homocedasticidade da distribuição dos erros é substituída pela suposição de heterocedasticidade;
- (ii) **perturbação nos dados:** neste caso, os possíveis tipos de perturbação podem depender do modelo. No caso particular da regressão normal linear, pode-se perturbar tanto a matriz de variáveis explicativas como o vetor da variável resposta.

Temos duas razões essenciais para considerar a perturbação dos dados: erros de mensuração e ocorrência de *outliers*. Billor e Lynes (1993) alertam para

a importância de distinguir bem os dois tipos de perturbação, pois há consequências e interpretações diferentes para ambos os esquemas.

Nos capítulos 2 e 3, apresentam-se alguns esquemas de perturbação, como, por exemplo, a perturbação de observação individual em modelos normais e logísticos.

2.2.3 Influência

Considera-se a função de verossimilhança para o parâmetro β a seguir,

$$l_{\omega}(\beta, y) = \sum_{j=1}^n \omega_j l(\beta, y_j),$$

em que $l(\beta, y_j)$ é o logaritmo da função de verossimilhança correspondente à j -ésima observação e ω_j é um tipo de perturbação, tal que $0 \leq \omega_j \leq 1$. Quando $\omega_1 = \omega_2 = \dots = \omega_n = 1$, significa que não há perturbação no modelo e, quando $\omega_j = 0$, significa que a j -ésima observação foi excluída.

A medida de influência mais conhecida é a distância de Cook, definida por

$$D_{\omega} = \frac{(\tilde{\beta} - \tilde{\beta}_{\omega})^t (X^t X) (\tilde{\beta} - \tilde{\beta}_{\omega})}{ps^2}$$

em que $\tilde{\beta}_{\omega} = (X^t \Delta X)^{-1} X^t \Delta y$ e $\Delta = \text{diag}\{\omega_1, \dots, \omega_n\}$. Esta medida mede quanto a perturbação $\omega = (\omega_1, \dots, \omega_n)^t$ afasta $\tilde{\beta}_{\omega}$ de $\tilde{\beta}$, segundo a métrica $X^t X$.

Quando o i -ésimo ponto é excluído, Paula (2004) mostra que a distância de Cook fica expressa por

$$D_i = \frac{(\tilde{\beta} - \tilde{\beta}_{(i)})^t (X^t X) (\tilde{\beta} - \tilde{\beta}_{(i)})}{ps^2}$$

$$= t_i^2 \frac{h_{ii}}{(1 - h_{ii})^p}. \quad (2.34)$$

Portanto, D_i será grande quando o i -ésimo ponto for influente (t_i grande) e/ou quando h_{ii} for próximo de 1.

Para os MLGs, a influência exercida pela i -ésima observação no modelo ajustado é medida ao avaliar o impacto sobre o logaritmo da função de verossimilhança $l(\beta, y)$, com a retirada desta observação.

Uma medida utilizada para avaliar este impacto é conhecida como afastamento pela verossimilhança (*likelihood displacement*) (COOK; WEISBERG, 1982), definida por

$$LD_i \cong 2 [l(\hat{\beta}_{(i)}, y) - l(\hat{\beta}_{(0)}, y)],$$

em que $l(\hat{\beta}_{(i)}, y)$ é a log-verossimilhança sem a i -ésima observação. Em Demétrio (2002) é mostrado que

$$LD_i \cong \left[\frac{\hat{h}_{ii}}{1 - \hat{h}_{ii}} \right] t_{Si}^2. \quad (2.35)$$

Para medir a influência das observações nas estimativas dos coeficientes do modelo logístico, a versão da medida afastamento pela verossimilhança utilizada é dada por

$$LD_i \cong \frac{(y_i - n_i \hat{p}_i)^2}{(1 - \hat{h}_{ii})^2 [n_i \hat{p}_i (1 - \hat{p}_i)]}.$$

2.2.4 Técnicas gráficas

Uma fase importante da análise de diagnóstico é a interpretação de gráficos utilizados para detectar pontos suspeitos de serem aberrantes e/ou influentes. Para os modelos de regressão normal linear, os gráficos mais utilizados são:

- (i) gráfico de \hat{e}_i contra a ordem das observações; usado para detectar pontos aberrantes;
- (ii) gráfico de \hat{e}_i^2 contra os valores ajustados; indica a ocorrência de heterocedasticidade nos dados;
- (iii) gráfico de \hat{h}_{ii} contra a ordem das observações; também é usado para detectar observações suspeitas de serem aberrantes;
- (iv) gráfico das coordenadas do vetor $|\mathbf{e}_{\max}|$ contra a ordem das observações; destaca pontos que podem ser influentes;
- (v) gráfico normal de probabilidade com envelopes; indica se há possíveis afastamentos da normalidade na distribuição dos resíduos. É o gráfico de t_i^* contra os valores esperados das estatísticas de ordem da normal padrão $(Z' (i)s)$.

Para os MLGs, as técnicas gráficas mais recomendadas são as seguintes:

- (i) gráfico de t_{si} contra os valores ajustados; detecta a ocorrência de pontos discrepantes em relação aos demais; estes pontos podem ser influentes ou de alavanca;
- (ii) gráfico de t_{Di} , contra os valores ajustados ou contra a ordem das observações; detecta pontos influentes;
- (iii) gráfico de \hat{h}_{ii} contra os valores ajustados ou contra a ordem das observações; detecta observações suspeitas de serem aberrantes;
- (iv) gráfico de LD_i contra a ordem das observações; as observações mais discrepantes neste gráfico podem ser pontos de alavanca e/ou influentes;

- (v) gráfico normal de probabilidade para \hat{D}_i com envelopes; fornece indícios de afastamento da suposição sobre a distribuição adotada para a variável resposta.

2.3 Avaliando a influência local por meio da curvatura normal

O método da influência local é considerado um dos mais modernos, dentro da análise de diagnóstico. Proposto por Cook (1986), consiste em verificar a existência de pontos que, sob modificações modestas no modelo, causam variações desproporcionais nos resultados do ajuste (PAULA, 2004).

Considerando-se o caso geral da distância de Cook D_ω , definida na subseção (1.2.3), agora se quer estudar as mudanças produzidas nesta medida, quando $\omega_i \rightarrow 1, \forall i$. Expandindo-se D_ω em série de Taylor até segunda ordem, em torno de $\omega_0 = \mathbf{1}$, tem-se que

$$\begin{aligned} D_\omega &\cong D_{\omega_0} + (D_{\omega_0} - \omega)^T D'_{\delta_0} + \frac{1}{2}(\delta_0 - \delta)^T D''_{\delta_0}(\omega_0 - \omega) \\ &\cong \frac{1}{2}(\omega_0 - \omega)^T D''_{\omega_0}(\omega_0 - \omega). \end{aligned}$$

Pode-se mostrar que

$$D''_{\omega_0} = \text{diag}(r)H\text{diag}(r),$$

em que $\text{diag}(r) = \text{diag}\{r_1, \dots, r_n\}$. Cook (1986) estuda a maior variação de D_ω em torno de ω_0 , o que equivale a maximizar a forma quadrática

$$\text{EMBED Equation. 3 } \mathbf{r}^T F \text{ EMBED Equation. 3 } \mathbf{r}, \text{ EMBED Equation. 3 } \mathbf{r}$$

em

que

$$\text{EMBED Equation. 3 } \mathbf{r} = \omega_0 - \omega, \text{ EMBED Equation. 3 } \mathbf{r}^T \text{ EMBED Equation. 3 } \mathbf{r}$$

O máximo da forma quadrática

$$\text{EMBED Equation. 3 } \mathbf{r}^T F \text{ EMBED Equation. 3 } \mathbf{r} \text{ corresponde}$$

ao maior autovalor da matriz F , denotado por λ_{\max} . As coordenadas do autovetor \mathbf{f}_{\max} , associado a λ_{\max} , contêm as influências locais das observações na direção que cause maior alteração. Assim, o gráfico de $|\mathbf{f}_{\max}|$ contra a ordem das observações pode revelar os pontos com maior influência na vizinhança de ω_0 . Estes pontos podem causar mudanças substanciais nas estimativas dos parâmetros sob pequenas perturbações no modelo.

Para os MLGs, quando a ligação é canônica, o vetor \mathbf{f}_{\max} é o autovetor correspondente ao maior autovalor da matriz F a seguir e serve para avaliar a influência local das observações nas estimativas dos parâmetros

$$F = \text{diag}(\hat{r}_p)H\text{diag}(\hat{r}_p),$$

em que $\hat{r}_p = (\hat{r}_{p_1}, \dots, \hat{r}_{p_1})^t$ e $\hat{r}_{p_l} = \frac{\phi^{1/2}(y_l - \hat{\beta}_l)}{\tilde{V}^{1/2}}$ é o l -ésimo resíduo de Pearson avaliado em $\tilde{\beta}$.

2.3.1 Gráficos de influência

No desenvolvimento desta seção, utilizou-se o modelo de regressão linear múltipla dado na equação (2.7). Considere a seguinte forma da estatística de influência proposta por Cook (1977):

$$D_i = \frac{\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_0\|^2}{p\sigma^2}, \quad (3.1)$$

em que $\|\cdot\|$ denota a norma euclidiana de um vetor, \hat{y} e \hat{y}_ω são vetores $n \times 1$, de valores ajustados baseados no conjunto total das observações e no conjunto dos dados sem a i -ésima observação, respectivamente, e p é a dimensão do vetor de parâmetros β . Vale ressaltar que, por i -ésima observação, entende-se a $(p+1)$ -upla $(y_i, x_{1i}, x_{2i}, \dots, x_{pi})$ que corresponde ao i -ésimo valor observado da variável resposta, acompanhado dos respectivos valores observados para as variáveis explicativas.

A estatística D_i pode ser bastante útil para detectar observações discrepantes e tais ocorrências conduzem à remoção ou à correção da correspondente observação. Em geral, diagnósticos baseados na deleção de observações permitem apenas duas conclusões: ou a observação é convenientemente especificada para o modelo ou é inteiramente não observável, no sentido de que a variância da população de onde foi extraída tende a infinito (COOK, 1986).

Frequentemente, deseja-se investigar outros interesses, como, por exemplo, a heterocedasticidade. Para isso, usa-se uma versão mais geral para a medida definida em (3.1), dada por

$$D_i(\omega) = \frac{\|\hat{y} - \hat{y}_\omega\|^2}{p\sigma^2}, \quad (3.2)$$

em que \hat{y}_ω é o vetor de valores ajustados obtido quando a i -ésima observação tem peso ω , com $0 \leq \omega \leq 1$ e as outras observações têm peso 1. Observe que quando $\omega \rightarrow 0$, $D_i(\omega) \rightarrow D_i$.

Para uma avaliação mais completa da influência de uma única observação, é importante que se investigue o comportamento de $D_i(\omega)$ para valores de ω diferentes de zero.

Cook (1986) estende essas ideias, permitindo que um vetor de perturbação ω seja utilizado para perturbar outros componentes do modelo. Uma motivação para essa extensão segue da relação entre $D_i(\omega)$ e a log-verossimilhança $l(\beta)$, que será vista a seguir.

Observe que se pode escrever o numerador da equação (3.2) na seguinte forma:

$$\|\hat{y} - \hat{y}_\omega\|^2 = \|y - \hat{y}_\omega\|^2 - 2\langle y - \hat{y}_\omega, y - \hat{y} \rangle + \|y - \hat{y}\|^2,$$

em que $\langle \cdot, \cdot \rangle$ representa o produto interno usual, de forma que

$$\begin{aligned} \langle y - \hat{y}_\omega, y - \hat{y} \rangle &= \langle y - \hat{y} + \hat{y} - \hat{y}_\omega, y - \hat{y} \rangle \\ &= \langle y - \hat{y}, y - \hat{y} \rangle + \langle \hat{y} - \hat{y}_\omega, y - \hat{y} \rangle \\ &= \|y - \hat{y}\|^2 + \langle \hat{y} - \hat{y}_\omega, y - \hat{y} \rangle. \end{aligned}$$

Com isso,

$$\|\hat{y} - \hat{y}_\omega\|^2 = \|y - \hat{y}_\omega\|^2 - \|y - \hat{y}\|^2 - 2\langle \hat{y} - \hat{y}_\omega, y - \hat{y} \rangle.$$

Segue, ainda, que

$$\langle \hat{y} - \hat{y}_\omega, y - \hat{y} \rangle = \langle \hat{y}, y \rangle - \langle \hat{y}, \hat{y} \rangle - \langle \hat{y}_\omega, y \rangle + \langle \hat{y}_\omega, \hat{y} \rangle.$$

Do capítulo 1 sabemos que $\hat{y} = X(X^t X)^{-1} X^t y$ e $\hat{y}_\omega = X(X^t W X)^{-1} X^t y$, em que $W = \text{diag}(w_1, w_2, \dots, w_n)$. Com isso, tem-se que

$$\begin{aligned} \langle \hat{y}, y \rangle &= \hat{y}^t y = y^t X(X^t X)^{-1} X^t y, \\ \langle \hat{y}, \hat{y} \rangle &= \hat{y}^t \hat{y} = y^t X(X^t X)^{-1} X^t y, \\ \langle \hat{y}_\omega, y \rangle &= \hat{y}_\omega^t y = y^t W X(X^t W X)^{-1} X^t y, \\ \langle \hat{y}_\omega, \hat{y} \rangle &= \hat{y}_\omega^t \hat{y} = y^t W X(X^t W X)^{-1} X^t y. \end{aligned}$$

Assim, conclui-se que

$$\langle \hat{y} - \hat{y}_\omega, y - \hat{y} \rangle = 0.$$

Portanto, a equação em (3.3) fica resumida a

$$\|\hat{y} - \hat{y}_\omega\|^2 = \|y - \hat{y}_\omega\|^2 - \|y - \hat{y}\|^2.$$

Substituindo-se este último resultado na equação em (3.2), tem-se que

$$pD_i(\omega) = \frac{\|\hat{y} - \hat{y}_\omega\|^2 - \|y - \hat{y}\|^2}{\sigma^2}.$$

Seja $l(\beta_\omega)$ a log-verossimilhança correspondente ao modelo perturbado por um vetor $\omega \in \mathbb{R}^q$, em que \mathbb{R}^q é um conjunto aberto de pequenas perturbações em \mathbb{R}^q . Assume-se que existe um $\omega_0 \in \mathbb{R}^q$, tal que $l(\beta_\omega) = l(\beta)$, para todo vetor β .

Assumindo que a log-verossimilhança no modelo perturbado é de classe C^2 e consideram-se apenas as partes relevantes da função de log-verossimilhança para o modelo postulado $l(\beta)$ e para o modelo perturbado $l(\beta_\omega)$, em que β e β_ω são os estimadores de máxima verossimilhança nos modelos postulado e perturbado, respectivamente, obtêm-se as seguintes relações:

$$\begin{aligned} \log l(\beta) &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= -\frac{1}{2\sigma^2} \|y - \hat{y}\|^2 \end{aligned}$$

$$\begin{aligned} \log l(\beta_\omega) &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \hat{y}_{i\omega})^2 \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \hat{y}_{i\omega})^2 \\ &= -\frac{1}{2\sigma^2} \|y - \hat{y}_\omega\|^2, \end{aligned}$$

Segue daí que

$$l(\hat{\beta}) - l(\hat{\beta}_\omega) = \frac{1}{2\sigma^2} (\|y - \hat{y}_\omega\|^2 - \|y - \hat{y}\|^2),$$

ou, ainda,

$$\|y - \hat{y}_\omega\|^2 - \|y - \hat{y}\|^2 = 2\sigma^2 \{l(\hat{\beta}) - l(\hat{\beta}_\omega)\}.$$

Finalmente, substituindo-se o último resultado acima na equação (3.4), encontra-se a relação entre a estatística $D_i(\omega)$ e a função de log-verossimilhança de $\hat{\beta}$,

$$pD_i(\omega) = 2\{l(\hat{\beta}) - l(\hat{\beta}_\omega)\}.$$

Cook (1986) utiliza a medida afastamento pela verossimilhança para avaliar a influência local, que é definida por

$$LD(\omega) = 2\{l(\hat{\beta}) - l(\hat{\beta}_\omega)\},$$

quando ω varia numa vizinhança de ω_0 . Observe que a equação em (3.5), na verdade, relaciona a estatística $D_i(\omega)$ de Cook e a medida afastamento pela verossimilhança definida em (3.6). Decorre daí que um gráfico de $LD(\omega)$ versus ω contém informações essenciais da influência do esquema de perturbação utilizado. Esse gráfico pode ser interpretado como a superfície geométrica $(q + 1)$ -dimensional formada pelos valores do vetor

$$\alpha(\omega) = \begin{pmatrix} \omega \\ LD(\omega) \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_q \\ LD(\omega) \end{pmatrix},$$

quando ω varia em \square **EMBED Equation.3** \square . A superfície definida acima é chamada de **gráfico de influência**.

2.3.1.1 A curvatura normal em um gráfico de influência

Nesta seção a atenção centra-se no método de diagnóstico da influência local, avaliado através da **curvatura normal**. Este método, introduzido por Cook (1986), consiste em se avaliar o comportamento da curvatura normal em torno de um ponto ω_0 pertencente ao gráfico de influência $\alpha(\omega)$, definido na seção anterior, lembrando que ω_0 corresponde ao ponto em que $l(\hat{\beta}) = l(\hat{\beta}|\omega_0)$. Para o caso particular de ter-se $q = 1$, o gráfico de influência se reduz a $\alpha(\omega) = \{\alpha_i(\omega)\}, i = 1, 2$, o que equivale à curva plana $\alpha(\omega) = (\omega, LD(\omega))$, cuja curvatura é dada por Stoker (1969, p. 26),

$$C = \frac{|\dot{\alpha}_1 \ddot{\alpha}_2 - \dot{\alpha}_2 \ddot{\alpha}_1|}{(\dot{\alpha}_1^2 \ddot{\alpha}_2^2 + \dot{\alpha}_2^2 \ddot{\alpha}_1^2)^{3/2}}$$

em que as notações $\dot{\alpha}$ e $\ddot{\alpha}$ denotam a primeira e a segunda derivadas de α , respectivamente. Segue que

- $\dot{\alpha}_1 = \frac{d\omega}{d\omega} = 1, \ddot{\alpha}_1 = \frac{d^2\omega}{d\omega^2} = 0;$
- $\dot{\alpha}_2|_{\omega=\omega_0} = \frac{dLD(\omega)}{d\omega}|_{\omega=\omega_0} = LD(\omega)|_{\omega=\omega_0} = 0$, pois $LD(\omega)$ alcança um mínimo local em ω_0 , e $\ddot{\alpha}_2 = LD(\omega)$.

Portanto, obtém-se que

$$C = \ddot{\alpha}_2|_{\omega=\omega_0} = LD(\omega_0).$$

Quando $q > 1$, um gráfico de influência é uma superfície em \mathbf{R}^{q+1} e, dessa forma, a noção de curvatura torna-se mais complexa. Estamos interessados em avaliar como a superfície $\alpha(\omega)$ desvia-se de seu plano tangente em torno do ponto ω_0 e, para isso, será considerado o comportamento das **seções normais** da superfície próxima a esse ponto. As seções normais são obtidas pelas interseções entre a superfície $\alpha(\omega)$ e os planos contendo o vetor normal ao seu

plano tangente em ω_0 . As curvaturas dessas seções normais são denominadas de **curvaturas normais** (STOKER, 1969).

Para obter uma seção normal, considere uma direção arbitrária em \mathbf{R}^q , que será representada por um vetor ℓ de comprimento unitário, e uma reta em $\Omega \subset \mathbf{R}^q$ passando por ω_0 , que denotaremos por

$$\omega(a) = \omega_0 + a \ell, \text{ com } a \in \mathbf{R}.$$

Essa reta gera uma **linha projetada** (*lifted line*) sobre a superfície $\alpha(\omega)$ passando pelo ponto $\alpha(\omega_0)$, que consiste em uma curva sobre essa superfície. Cada direção ℓ especifica uma linha projetada e vice-versa, pois o plano tangente à superfície em ω_0 é gerado pelas colunas da matriz \mathbf{V} cujos elementos

$$\frac{\partial \alpha_i(\omega)}{\partial \omega_j}$$

são

$i = 1, 2, \dots, q+1; j = 1, 2, \dots, q$, com todas as derivadas avaliadas em ω_0 .

Note que

$$\frac{\partial \alpha_i(\omega)}{\partial \omega_j} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j. \end{cases}$$

Assim, a matriz \mathbf{V} tem a forma $\mathbf{V} = (I_q, \mathbf{0})^T$. Seja \mathbf{b}_{q+1} o vetor da base canônica de \mathbf{R}^{q+1} com $\mathbf{1}$ na última posição e zeros nas outras. Dessa forma,

$$\mathbf{V}^T \mathbf{b}_{q+1} = \mathbf{0}.$$

Logo, \mathbf{b}_{q+1} é ortogonal às colunas da matriz \mathbf{V} . Como as colunas dessa matriz formam uma base para o plano tangente à superfície em ω_0 , o espaço que contém o vetor \mathbf{b}_{q+1} é ortogonal ao plano tangente e contém o vetor $(\frac{\partial \alpha_i(\omega)}{\partial \omega_j})_{q+1}$. Como os vetores \mathbf{b}_{q+1} e $(\frac{\partial \alpha_i(\omega)}{\partial \omega_j})_{q+1}$ são linearmente independentes

e geram o espaço ortogonal ao plano tangente em ω_0 , então, eles formam uma base para este espaço.

2.3.1.2 A Curvatura normal de uma *lifted line*

A *lifted line* (linha projetada) em uma dada direção ℓ consiste na curva plana dada por ρ

$$\rho(a) = (\text{EMBED Equation. DSMT4 } \beta^T \omega_1, \text{EMBED Equation. DSMT4 } \beta^T \omega_2)^T = [a, LD\omega(a)]^T$$

e pode ser vista como o conjunto dos pontos de $\alpha(\omega)$, gerado pelos vetores b_{q+1} e $(\text{EMBED Equation. DSMT4 } \beta^T \omega^T, 0)^T$. A curvatura normal da curva $\rho(a)$, na direção ℓ , adaptada de (3.8) é escrita da seguinte forma:

$$C_1(\text{EMBED Equation. DSMT4 } \beta^T \omega) = |\text{EMBED Equation. DSMT4 } \beta^T \omega_1, \text{EMBED Equation. DSMT4 } \beta^T \omega_2| \quad (3.11)$$

avaliada em ω_0 .

Para melhor avaliar esta expressão, deve-se lembrar da suposição de que a função de log-verossimilhança perturbada $l(\beta, \omega)$ é de classe C^2 em $(\beta^T, \omega^T)^T$ e, para a superfície

$$\rho(a) = (\text{EMBED Equation. DSMT4 } \beta^T \omega_1, \text{EMBED Equation. DSMT4 } \beta^T \omega_2)^T = [a, LD\omega(a)]^T$$

, tem-se o seguinte:

- $\text{EMBED Equation. DSMT4 } \beta^T \omega_1 = \frac{da}{da} = 1$, $\text{EMBED Equation. DSMT4 } \beta^T \omega_2 = \frac{d}{da} \omega(a)$
- $\frac{d}{da} \omega(a) = \frac{d}{da} (\omega_0 + a \text{EMBED Equation. DSMT4 } \beta^T \omega)$, logo $\frac{d^2}{da^2} = 0$;

- Pela regra da cadeia,

$$\begin{aligned} \frac{d}{da} LD\omega(a) &= [\text{gradiente}(LD)]^t \left[\frac{d}{da} \omega(a) \right] \\ &= \left[\frac{\partial LD(\omega)}{\partial \omega} \right]_{\omega=\omega_0} \frac{d}{da} \omega(a); \\ \left[\frac{\partial LD(\omega)}{\partial \omega} \right]_{\omega=\omega_0} &= \left. \frac{d}{da} LD\{\omega(a)\} \right|_{a=0} = \left(\left[\frac{\partial}{\partial \omega} \right]^t \right) \end{aligned}$$

, pois ω_0 é o ponto de mínimo local da Função LD.

Substituindo-se esses resultados em (3.11), obtém-se

$$\begin{aligned} C_{\omega_0} &= \left[\frac{\partial LD(\omega)}{\partial \omega} \right]_{\omega=\omega_0} \\ &= [L\bar{D}\{\omega(a)\}], \end{aligned}$$

avaliada em $a = 0$. Utilizando novamente a regra da cadeia, segue que

$$\begin{aligned} L\bar{D}\{\omega(a)\} &= \frac{d^2}{da^2} LD\{\omega(a)\} \\ &= \frac{d}{da} \left[\frac{\partial}{\partial \omega} LD\{\omega(a)\} \right]^t \\ &= \frac{d}{da} \frac{\partial}{\partial \omega^t} LD\{\omega(a)\} \\ &= \frac{d}{da} \left(\frac{\partial}{\partial \omega_1} LD\{\omega(a)\}, \dots, \frac{\partial}{\partial \omega_q} LD\{\omega(a)\} \right) \\ &= \left(\left[\frac{\partial^2}{\partial \omega_1 \partial \omega} LD\{\omega(a)\} \right]^t \frac{d}{da} \omega(a), \dots, \left[\frac{\partial^2}{\partial \omega_q \partial \omega} LD\{\omega(a)\} \right]^t \frac{d}{da} \omega(a) \right) \end{aligned}$$

$$\begin{aligned}
&= \left[\begin{array}{c} \left[\frac{\partial^2}{\partial \omega_1 \partial \omega} LD\{\omega(\alpha)\} \right]^t \\ \vdots \\ \left[\frac{\partial^2}{\partial \omega_q \partial \omega} LD\{\omega(\alpha)\} \right]^t \end{array} \right]^t \quad \text{EMBED Equation.DSMT4} \quad \text{EMBED Equation.DSMT4} \\
&= \left[\begin{array}{c} \left(\frac{\partial^2}{\partial \omega_1 \partial \omega^t} LD\{\omega(\alpha)\} \right) \\ \vdots \\ \left(\frac{\partial^2}{\partial \omega_q \partial \omega^t} D\{\omega(\alpha)\} \right) \end{array} \right]^t \quad \text{EMBED Equation.DSMT4} \quad \text{EMBED Equation.DSMT4} \\
&= \left[\left(\frac{\partial^2}{\partial \omega \partial \omega^t} LD\{\omega(\alpha)\} \right) \right]^t \quad \text{EMBED Equation.DSMT4} \quad \text{EMBED Equation.DSMT4} \\
&= \text{EMBED Equation.DSMT4} \quad \left[\frac{\partial^2}{\partial \omega \partial \omega^t} LD\{\omega(\alpha)\} \right]^t \quad \text{EMBED Equation.DSMT4} \\
&= \text{EMBED Equation.DSMT4} \quad \frac{\partial^2}{\partial \omega \partial \omega^t} LD\{\omega(\alpha)\} \quad \text{EMBED Equation.DSMT4},
\end{aligned}$$

pois a matriz hessiana é simétrica. Avaliando-se a derivada $\alpha = \mathbf{0}$, obtém-se a curvatura normal do gráfico de influência $\alpha(\omega)$, que é dada por

$$C_{\text{EMBED Equation.DSMT4}} = \left| \text{EMBED Equation.DSMT4} \right|^t LD(\omega_0) \quad \text{EMBED Equation.DSMT4} \quad (3.12)$$

Observe que, pela definição do afastamento pela verossimilhança, tem-se que

$$LD(\omega_0) = \frac{\partial^2}{\partial \omega \partial \omega^t} [2\{l(\hat{\beta}) - l(\hat{\beta}_\omega)\}] \Big|_{\omega=\omega_0} = -2 \frac{\partial^2}{\partial \omega \partial \omega^t} l(\hat{\beta}_\omega) \Big|_{\omega=\omega_0},$$

portanto, a curvatura normal vista em (3.12) fica resumida por

$$C_{\text{EMBED Equation.DSMT4}} = 2 \left[\text{EMBED Equation.DSMT4} \right]^t \bar{F} \text{EMBED Equation.DSMT4} \quad (3.13)$$

em que \bar{F} é uma matriz quadrada de dimensão $q \times q$, cujos elementos são

$$\frac{\partial^2 l(\tilde{\beta}_\omega)}{\partial \omega_r \partial \omega_s},$$

dados pelas derivadas avaliadas em $\omega = \omega_0$.

2.3.1.3 Uma forma operacional para a curvatura normal

O elemento (r, s) da matriz \bar{F} pode ser obtido por meio da regra da cadeia

$$\begin{aligned} \bar{F}_{rs} &= \left. \frac{\partial^2}{\partial \omega_r \partial \omega_s} l(\tilde{\beta}_\omega) \right|_{\omega=\omega_0} \\ &= \left. \frac{\partial}{\partial \omega_r} \left\{ \frac{\partial}{\partial \omega_s} l(\tilde{\beta}_\omega) \right\} \right|_{\omega=\omega_0} \\ &= \left. \frac{\partial}{\partial \omega_r} \left[\left\{ \frac{\partial}{\partial \tilde{\beta}_\omega} l(\tilde{\beta}_\omega) \right\}^t \frac{\partial \tilde{\beta}_\omega}{\partial \omega_s} \right] \right|_{\omega=\omega_0} \\ &= \left[\left. \frac{\partial}{\partial \omega_r} \left\{ \frac{\partial}{\partial \tilde{\beta}_\omega} l(\tilde{\beta}_\omega) \right\}^t \frac{\partial \tilde{\beta}_\omega}{\partial \omega_s} \right] \right|_{\omega=\omega_0} + \left[\left\{ \frac{\partial}{\partial \tilde{\beta}_\omega} l(\tilde{\beta}_\omega) \right\} \frac{\partial}{\partial \omega_r} \left(\frac{\partial \tilde{\beta}_\omega}{\partial \omega_s} \right) \right] \Big|_{\omega=\omega_0} \\ &= \left(\left[\left\{ \frac{\partial^2}{\partial \tilde{\beta}_\omega \partial \tilde{\beta}_\omega} l(\tilde{\beta}_\omega) \right\}^t \frac{\partial \tilde{\beta}_\omega}{\partial \omega_r} \right] \frac{\partial \tilde{\beta}_\omega}{\partial \omega_s} \right) \Big|_{\omega=\omega_0} + \left\{ \frac{\partial}{\partial \tilde{\beta}_\omega} l(\tilde{\beta}_\omega) \frac{\partial^2 \tilde{\beta}_\omega}{\partial \omega_r \partial \omega_s} \right\} \Big|_{\omega=\omega_0} \\ &= \left[\left(\frac{\partial \tilde{\beta}_\omega}{\partial \omega_r} \right)^t \left\{ \frac{\partial^2}{\partial \tilde{\beta}_\omega \partial \tilde{\beta}_\omega} l(\tilde{\beta}_\omega) \right\} \frac{\partial \tilde{\beta}_\omega}{\partial \omega_s} \right] \Big|_{\omega=\omega_0} + \left\{ \frac{\partial}{\partial \tilde{\beta}_\omega} l(\tilde{\beta}_\omega) \frac{\partial^2 \tilde{\beta}_\omega}{\partial \omega_r \partial \omega_s} \right\} \Big|_{\omega=\omega_0}. \end{aligned}$$

Por hipótese, temos que $l(\tilde{\beta}, \omega_0) = l(\tilde{\beta})$, para todo $\tilde{\beta} \in R^p$, assim $\tilde{\beta}_{\omega_0} = \tilde{\beta}$ e, portanto,

$$\left. \frac{\partial}{\partial \hat{\beta}_\omega^t} l(\hat{\beta}_\omega) \right|_{\omega=\omega_0} = \left. \frac{\partial}{\partial \beta} l(\beta, \omega_0) \right|_{\beta=\hat{\beta}} = 0,$$

pois $\hat{\beta}$ é o ponto de máximo da função de log-verossimilhança, tanto no modelo postulado $l(\beta)$ quanto no modelo perturbado $l(\beta, \omega)$. Dessa forma, a segunda parcela da derivada \tilde{F}_{rs} acima, é nula. Logo, tem-se que

$$\tilde{F}_{rs} = \left(\frac{\partial \hat{\beta}_\omega}{\partial \omega_r} \right)^t \left(\frac{\partial^2}{\partial \hat{\beta}_\omega \partial \hat{\beta}_\omega^t} l(\hat{\beta}_\omega) \right) \left(\frac{\partial \hat{\beta}_\omega}{\partial \omega_s} \right),$$

avaliada em $\omega = \omega_0$.

Em notação matricial, pode-se escrever

$$\tilde{F} = J^t \tilde{L} J,$$

em que J é uma matriz $p \times q$ cujos elementos são dados por

$$J_{ks} = \left. \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \right|_{\omega=\omega_0}, \text{ sendo } \hat{\beta}_{k\omega} \text{ a } k\text{-ésima componente do vetor } \hat{\beta}_\omega, \text{ enquanto } (-\tilde{L}) \text{ é a matriz de informação observada de Fisher.}$$

Considerando o fato de que $\hat{\beta}_\omega$ é um ponto de máximo local da log-verossimilhança perturbada $l(\beta, \omega)$, para cada $\omega \in \Omega$,

$$\left. \frac{\partial l(\beta, \omega)}{\partial \beta_l} \right|_{\beta=\hat{\beta}} = 0, l = 1, \dots, p.$$

Deve-se observar que o primeiro membro desta igualdade é uma função das $p + q$ variáveis $\beta_{1\omega}, \dots, \beta_{p\omega}$ e $\omega_1, \dots, \omega_q$. Já o segundo membro é uma função constante, a função identicamente nula. Segue que, derivando-se os dois membros da equação (3.15) em relação a ω e, em seguida, avaliando-se a derivada no ponto ω_0 , a derivada do segundo membro é nula e a do primeiro é dada por

$$\frac{\partial}{\partial \omega_s} \left(\left. \frac{\partial l(\hat{\beta}_{\omega_0}, \omega_0)}{\partial \beta_i} \right) \right) = \sum_{k=1}^p \left(\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \hat{\beta}_{k\omega} \partial \beta_i} \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \right) \right) \Big|_{\omega=\omega_0} + \sum_{r=1}^q \left(\left. \frac{\partial^2 l(\hat{\beta}_\omega, \omega)}{\partial \omega_r \partial \beta_i} \frac{\partial \omega_r}{\partial \omega_s} \right) \right) \Big|_{\omega=\omega_0}$$

$$\begin{aligned}
&= \sum_{k=1}^p \left(\frac{\partial^2 l(\tilde{\beta}_{\omega}, \omega)}{\partial \tilde{\beta}_{k\omega} \partial \beta_i} \frac{\partial \tilde{\beta}_{k\omega}}{\partial \omega_s} \right) \Big|_{\omega=\omega_0} + \left(\frac{\partial^2 l(\tilde{\beta}_{\omega}, \omega)}{\partial \omega_s \partial \beta_i} \right) \Big|_{\omega=\omega_0} \\
&= \sum_{k=1}^p \left(\frac{\partial^2 l(\tilde{\beta}_{\omega}, \omega)}{\partial \tilde{\beta}_{k\omega} \partial \beta_i} \frac{\partial \tilde{\beta}_{k\omega}}{\partial \omega_s} \right) \Big|_{\omega=\omega_0} + \frac{\partial^2 l(\tilde{\beta}_{\omega_0}, \omega_0)}{\partial \omega_s \partial \beta_i}.
\end{aligned}$$

Dessa forma, para $i = 1, \dots, p$ vale a seguinte relação

$$\sum_{k=1}^p \left(\frac{\partial^2 l(\tilde{\beta}_{\omega}, \omega)}{\partial \tilde{\beta}_{k\omega} \partial \beta_i} \frac{\partial \tilde{\beta}_{k\omega}}{\partial \omega_s} \right) \Big|_{\omega=\omega_0} + \frac{\partial^2 l(\tilde{\beta}_{\omega_0}, \omega_0)}{\partial \omega_s \partial \beta_i} = 0.$$

Como as duas funções de log-verossimilhança são iguais em ω_0 , para todo vetor β , tem-se que

$$\frac{\partial^2 l(\tilde{\beta}_{\omega}, \omega)}{\partial \tilde{\beta}_{k\omega} \partial \beta_i} \Big|_{\omega=\omega_0} = \frac{\partial^2 l(\tilde{\beta}_{\omega_0}, \omega_0)}{\partial \tilde{\beta}_{k\omega} \partial \beta_i} = \frac{\partial^2 l(\tilde{\beta})}{\partial \beta_k \partial \beta_i} = \frac{\partial^2 l(\tilde{\beta})}{\partial \beta_k \partial \beta_i} \Big|_{\beta=\tilde{\beta}}.$$

Segue daí, e pela equação (3.16), que para $i = 1, \dots, p$, e $s = 1, \dots, q$,

$$\sum_{k=1}^p \left\{ \left(\frac{\partial^2 l(\beta)}{\partial \beta_k \partial \beta_i} \right) \Big|_{\beta=\tilde{\beta}} \left(\frac{\partial \tilde{\beta}_{k\omega}}{\partial \omega_s} \right) \Big|_{\omega=\omega_0} \right\} + \frac{\partial^2 l(\tilde{\beta}_{\omega_0}, \omega_0)}{\partial \omega_s \partial \beta_i} = 0. \quad (2.17)$$

Agora, vamos definir uma matriz Δ , de dimensão $p \times q$, dada por

$$\Delta = \frac{\partial^2 l(\beta, \omega)}{\partial \beta \partial \omega^t},$$

avaliada em $\beta = \tilde{\beta}$ e $\omega = \omega_0$. Com isso, a equação (3.17) pode ser escrita na forma

$$\sum_{k=1}^p \tilde{L}_{ki} J_{ks} + \tilde{\Delta}_{is} = 0,$$

ou, ainda, usando a simetria da matriz \tilde{L} ,

$$\sum_{k=1}^p \tilde{L}_{ik} J_{ks} + \tilde{\Delta}_{is} = 0.$$

Assim, obtém-se a forma mais resumida para a matriz \mathbf{J} ,

$$\mathbf{J} = -\mathbf{L}^{-1} \tilde{\Delta} \quad (3.18).$$

Substituindo \mathbf{J} na equação (3.14), obtém-se

$$\tilde{F} = \left(-\tilde{\Delta} \tilde{L}^{-1} \right) \tilde{L} \left(-\tilde{L}^{-1} \tilde{\Delta} \right)$$

$$\dot{F} = \left(\frac{\partial LD(\omega)}{\partial \omega_j} \right) = \left(-2 \frac{\partial l(\hat{\beta}_\omega)}{\partial \omega_j} \right),$$

com $j = 1, \dots, q$, então, pode-se escrever a matriz V como

$$V^t = (I_q, \dot{F}).$$

Seja ω^{jk} o vetor de ordem $(q+1) \times 1$ com elementos dados por $\frac{\partial^2(\alpha_i)}{\partial \omega_j \partial \omega_k}$, com $i = 1, \dots, q+1$. A seguir obtêm-se a primeira e a segunda derivadas da superfície $\alpha(\omega(\alpha))$ em ω^* , que denotaremos por $\dot{\alpha}$ e $\ddot{\alpha}$, respectivamente. Assim, tem-se que

$$\dot{\alpha} = \left(\frac{\partial \left(\omega^* \cdot \mathbf{1} + a \right)}{\partial a} \right) \quad (3.21).$$

Mas,

$$\begin{aligned} \frac{\partial LD\{\omega(\alpha)\}}{\partial a} &= \frac{\partial LD\{\omega_1\}}{\partial \omega_1} \frac{\partial \omega_1(\alpha)}{\partial a} + \dots + \frac{\partial LD\{\omega_q\}}{\partial \omega_q} \frac{\partial \omega_q(\alpha)}{\partial a} \\ &= (-2) \frac{\partial l(\hat{\beta}_\omega)}{\partial \omega_1} + \dots + (-2) \frac{\partial l(\hat{\beta}_\omega)}{\partial \omega_q} \\ &= \left((-2) \frac{\partial l(\hat{\beta}_\omega)}{\partial \omega_1}, \dots, (-2) \frac{\partial l(\hat{\beta}_\omega)}{\partial \omega_q} \right) \begin{pmatrix} \mathbf{1} \\ \vdots \\ \mathbf{1} \end{pmatrix} \\ &= \dot{F} \cdot \mathbf{1}. \end{aligned} \quad (3.22)$$

Segue que, substituindo-se o resultado acima na expressão em (3.21), tem-se

$$\dot{\alpha} = \left(\mathbf{1}, \dots, \mathbf{1} \right)$$

Pode-se, ainda, escrever este vetor utilizando a notação de matriz em bloco a seguir:

$$\dot{\alpha} = \left[\begin{array}{c} \text{EMBED Equation.DSMT4} \\ \dot{F}^t \text{EMBED Equation.DSMT4} \end{array} \right] = \left[\begin{array}{c} I_q \\ \dot{F}^t \end{array} \right] \text{EMBED Equation.DSMT4} \quad (3.23)$$

Para se obter-se a segunda derivada, observe que

$$\begin{aligned} \ddot{\alpha}_\ell &= \frac{\partial \dot{\alpha} \text{EMBED Equation.DSMT4}}{\partial \mathbf{a}} \\ &= \left(\frac{\partial \ell_1}{\partial \mathbf{a}}, \dots, \frac{\partial \ell_q}{\partial \mathbf{a}}, \frac{\partial \dot{F}^t \ell}{\partial \mathbf{a}} \right)^t \\ &= \left(0, \dots, 0, \frac{\partial \dot{F}^t \ell}{\partial \mathbf{a}} \right)^t. \end{aligned} \quad (3.24)$$

Note que se pode escrever $\dot{F}^t \ell = \sum_{r=1}^q \dot{F}_r \ell_r$, em que $\dot{F}_r = -2 \frac{\partial l(\beta_\omega)}{\partial \mathbf{a}}$. Assim,

$$\begin{aligned} \frac{\partial \dot{F}^t \ell}{\partial \mathbf{a}} &= \sum_{r=1}^q \frac{d \dot{F}_r \ell_r}{d \mathbf{a}} \\ &= \sum_{r=1}^q \ell_r \frac{d}{d \mathbf{a}} \left[-2 \frac{\partial l(\beta_\omega)}{\partial \mathbf{a}} \right] \\ &= -2 \sum_{r=1}^q \ell_r \left[\frac{\partial^2 l(\beta_\omega)}{\partial \omega_r \partial \omega_1} \frac{\partial \omega_1(a)}{\partial \mathbf{a}} + \dots + \frac{\partial^2 l(\beta_\omega)}{\partial \omega_r \partial \omega_q} \frac{\partial \omega_q(a)}{\partial \mathbf{a}} \right] \\ &= -2 \sum_{r=1}^q \ell_r \left[\frac{\partial^2 l(\beta_\omega)}{\partial \omega_r \partial \omega_1} \epsilon_1 + \dots + \frac{\partial^2 l(\beta_\omega)}{\partial \omega_r \partial \omega_q} \epsilon_q \right] \\ &= -2 \sum_{r=1}^q \ell_r \sum_{j=1}^q \frac{\partial^2 l(\beta_\omega)}{\partial \omega_r \partial \omega_j} \epsilon_j \\ &= -2 \sum_{r=1}^q \sum_{j=1}^q \ell_r F_{rj} \epsilon_j \\ &= -2 \ell^t F \ell. \end{aligned} \quad (3.25)$$

Substituindo-se o resultado acima no vetor acima $\hat{\alpha}_\ell$, em (3.24), obtém-se a segunda derivada,

$$\hat{\alpha}_\ell = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -2\ell^t \bar{F} \ell \end{pmatrix} = -2\ell^t \bar{F} \ell \mathbf{b}_{q+1} \quad (3.26),$$

em que \mathbf{b}_{q+1} é o vetor definido na seção (2.3).

A curvatura normal $C_{\hat{\tau}}$ associada à direção ℓ pode ser escrita como (COOK, 1986)

$$C_{\hat{\tau}} = \frac{\|\bar{P}_V \hat{\alpha}_\ell\|}{\|\hat{\alpha}_\ell\|^2} \quad (3.27),$$

em que P_V é o operador de projeção ortogonal do espaço coluna da matriz V e $\bar{P}_V = I - P_V$. A curvatura normal $C_{\hat{\tau}}$ será avaliada a seguir, por meio das formas obtidas para a primeira e a segunda derivadas, calculadas anteriormente.

Desenvolvendo-se o denominador da equação (3.27), tem-se que

$$\begin{aligned} \|\hat{\alpha}_\ell\|^2 &= \hat{\alpha}_\ell^t \hat{\alpha}_\ell \\ &= (V\ell)^t (V\ell) \\ &= \ell^t V^t V \ell \\ &= \ell^t \begin{bmatrix} I_q & \dot{F} \\ & \dot{F}^t \end{bmatrix} \ell \\ &= \ell^t [I_q + \dot{F} \dot{F}^t] \ell. \end{aligned} \quad (3.28)$$

Para se obter o numerador de (3.27), observe que

$$\begin{aligned} P_V &= V(V^t V)^{-1} V^t \\ &= \begin{bmatrix} I_q \\ \dot{F}^t \end{bmatrix} [I_q + \dot{F} \dot{F}^t]^{-1} [I_q \dot{F}]. \end{aligned}$$

Pela proposição (A.3.1), tem-se que

$$[I_q + \dot{F} \dot{F}^t]^{-1} = I_q - \frac{1}{1 - \|\dot{F}\|^2} \dot{F} \dot{F}^t.$$

Segue daí que se pode escrever

$$P_V = \begin{bmatrix} I_q \\ \dot{F}^t \end{bmatrix} [I_q + \lambda F \dot{F}^t]^{-1} [I_q \dot{F}],$$

em que $\lambda = -\frac{1}{1 - \|F\|^2}$. Dessa forma, obtém-se

$$P_V = \begin{bmatrix} I_q + \lambda \dot{F} \dot{F}^t & \dot{F} + \lambda \dot{F} \dot{F}^t \dot{F} \\ \dot{F}^t + \lambda \dot{F}^t F \dot{F}^t & \dot{F}^t \dot{F} + \lambda \dot{F}^t F \dot{F}^t \dot{F} \end{bmatrix}.$$

Logo,

$$\bar{P}_V = I_{q+1} - P_V = \begin{bmatrix} -\lambda \dot{F} \dot{F}^t & -\dot{F} - \lambda \dot{F} \dot{F}^t \dot{F} \\ -\dot{F}^t - \lambda \dot{F}^t F \dot{F}^t & 1 - \dot{F}^t F - \lambda \dot{F}^t F \dot{F}^t \dot{F} \end{bmatrix}.$$

Pode-se escrever

$$\bar{P}_V = \begin{bmatrix} A & B \\ B^t & C \end{bmatrix}, \quad (3.29)$$

com $A = (-\lambda \dot{F} \dot{F}^t)_{q \times q}$, $B = (-\dot{F} - \lambda \dot{F} \dot{F}^t \dot{F})_{q \times 1}$ e $C = (1 - F^t F - \lambda F^t F F^t F)_{1 \times 1}$.

Note, ainda, que

$$\begin{aligned} \|\bar{P}_V \bar{\alpha}_\ell\| &= (\bar{P}_V \bar{\alpha}_\ell)^t (\bar{P}_V \bar{\alpha}_\ell) \\ &= \bar{\alpha}_\ell^t \bar{P}_V^t \bar{P}_V \bar{\alpha}_\ell \\ &= [0(-2)\ell^t \dot{F} \ell] \begin{bmatrix} A^t & B \\ B^t & C \end{bmatrix} \begin{bmatrix} A & B \\ B^t & C \end{bmatrix} \begin{bmatrix} 0 \\ (-2)\ell^t \dot{F} \ell \end{bmatrix} \\ &= [0(-2)\ell^t \dot{F} \ell] \begin{bmatrix} A^t A + B B^t & A^t B + B C \\ B^t A + C B^t & B^t B + C^2 \end{bmatrix} \begin{bmatrix} 0 \\ (-2)\ell^t \dot{F} \ell \end{bmatrix} \\ &= (-2\ell^t \dot{F} \ell) (B^t B + C^2) (-2\ell^t \dot{F} \ell) \\ &= (-2\ell^t \dot{F} \ell)^2 (B^t B + C^2) \end{aligned} \quad (3.30)$$

Pela proposição (A.3.2) no apêndice A, têm-se os seguintes resultados

$$\begin{aligned} B^t B &= \frac{\|\dot{F}\|^2}{(1 + \|\dot{F}\|^2)^2} \\ C^2 &= \frac{1}{(1 + \|\dot{F}\|^2)^2}. \end{aligned}$$

Substituindo-se estes resultados em (2.30), tem-se que

$$\begin{aligned}
\|\bar{P}_V \ddot{\alpha}_\ell\|^2 &= (-2\ell^t \dot{F} \ell)^2 \left[\frac{\|\dot{F}\|^2}{(1 + \|\dot{F}\|^2)^2} + \frac{1}{(1 + \|\dot{F}\|^2)^2} \right] \\
&= (-2\ell^t \dot{F} \ell)^2 \left[\frac{1 + \|\dot{F}\|^2}{(1 + \|\dot{F}\|^2)^2} \right] \\
&= (-2\ell^t \dot{F} \ell)^2 \frac{1}{1 + \|\dot{F}\|^2}, \tag{3.31}
\end{aligned}$$

Segue daí que

$$\|\bar{P}_V \ddot{\alpha}_\ell\| = |2\ell^t \dot{F} \ell| \frac{1}{(1 + \|\dot{F}\|^2)^{1/2}}. \tag{3.32}$$

Finalmente, substituindo-se os resultados de (3.28) e (3.32) em (3.27), obtém-se a expressão a seguir para a curvatura da seção normal de uma superfície a partir de um ponto ω^* :

$$C_i^* = \frac{|-2\ell^t \dot{F} \ell|}{\left(1 + \|\dot{F}\|^2\right)^{\frac{1}{2}} \ell^t \left[I_q + \dot{F}^t \dot{F} \right] \ell}. \tag{3.33}$$

2.4 Perturbação no modelo normal linear

Considera-se uma amostra y_1, y_2, \dots, y_n , em que $Y_i \sim N(\mu_i, \sigma^2)$, com σ^2 conhecido. Seja ω um vetor de perturbação de ordem $n \times 1$ para o modelo normal linear múltiplo dado em (1.7). Tem-se que a função de verossimilhança fica estabelecida por

$$L(\beta, \omega) = \frac{1}{(\sigma^2)^{\frac{n}{2}} (2\pi)^{\frac{n}{2}}} e^{-\sum_{i=1}^n \omega_i (y_i - x_i^t \beta)^2},$$

consequentemente, a função de log-verossimilhança perturbada é

$$l(\beta, \omega) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \sum_{i=1}^n \frac{1}{2\sigma^2} \omega_i (y_i - x_i^t \beta)^2, \quad (3.34)$$

em que ω_i e y_i são os elementos de ordem i dos vetores ω e y , respectivamente e x_i^t é a i -ésima linha da matriz de variáveis explicativas X .

Diferenciando-se a função de log-verossimilhança em relação ao vetor β , tem-se que

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= -\frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i \frac{\partial}{\partial \beta} (y_i - x_i^t \beta)^2 \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i 2(y_i - x_i^t \beta) \frac{\partial}{\partial \beta} (y_i - x_i^t \beta) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n \omega_i (y_i - x_i^t \beta) x_i \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n x_i^t r_i \omega_i \\ &= \frac{1}{\sigma^2} X^t \text{diag}(r_i) \omega, \end{aligned} \quad (3.35)$$

em que $r_i = y_i - x_i^t \beta$, com $i = 1, \dots, n$ são os resíduos. Diferenciando-se, agora, este resultado em relação ao vetor ω^t , segue que

$$\begin{aligned} \frac{\partial}{\partial \omega^t} \left[\frac{\partial l}{\partial \beta} \right] &= \frac{\partial^2 l}{\partial \beta \partial \omega^t} \\ &= \frac{1}{\sigma^2} X^t \text{diag}(r_i). \end{aligned}$$

Observa-se que, pela definição dada na seção (3.3), tem-se que

$$\Delta = \left. \frac{\partial^2 l}{\partial \beta \partial \omega^t} \right|_{\beta = \hat{\beta}, \omega = \omega_0},$$

Segue daí que

$$\Delta = \frac{1}{\sigma^2} X^t D(e), \quad (3.36)$$

em que $D(e) = \text{diag}(e_i)$ e $e_i = r_i \Big|_{\beta = \hat{\beta}} = y_i - x_i^t \hat{\beta}$.

Diferenciando-se a equação (2.35) em relação ao vetor β^t , e avaliando-se o resultado em $\hat{\beta}$ e ω_0 , obtém-se a matriz $\bar{L}(\hat{\beta})$ definida na subseção (3.3.2). Segue que

$$\begin{aligned} \frac{\partial}{\partial \beta} \left[\frac{\partial l}{\partial \beta^t} \right] &= \frac{\partial^2 l}{\partial \beta \partial \beta^t} \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n \omega_i \frac{\partial}{\partial \beta^t} (y_i - x_i^t \beta) x_i \\ &= -\frac{1}{\sigma^2} \sum_{i=1}^n \omega_i \left(0 - \frac{\partial}{\partial \beta^t} x_i x_i^t \beta \right) \\ &= -\frac{1}{\sigma^2} \sum_{i=1}^n \omega_i x_i x_i^t \\ &= -\frac{1}{\sigma^2} \sum_{i=1}^n x_i \omega_i x_i^t \\ &= -\frac{1}{\sigma^2} X^t W X, \end{aligned}$$

em que $W = \text{diag}(\omega_i)$.

Dessa forma,

$$\bar{L}(\hat{\beta}) = \frac{\partial^2 l}{\partial \beta \partial \beta^t} \Big|_{\beta = \hat{\beta}, \omega = \omega_0} = -\frac{1}{\sigma^2} X^t I_n X = -\frac{1}{\sigma^2} X^t X,$$

Consequentemente, tem-se que

$$L^{-1} = -\sigma^2 (X^t X)^{-1}. \quad (3.37)$$

Substituindo-se os resultados dados em (3.36) e (3.37) na equação da curvatura em (3.20), tem-se

$$\begin{aligned}
C_i &= 2 \left| \ell^t \frac{1}{\sigma^2} D(\mathbf{e}) X (-\sigma^2) (X^t X)^{-1} \frac{1}{\sigma^2} X^t D(\mathbf{e}) \ell \right| \\
&= 2 |\mathbf{1}| \left| \frac{1}{\sigma^2} \ell^t D(\mathbf{e}) X (X^t X)^{-1} X^t D(\mathbf{e}) \ell \right| \\
&= \frac{2 \ell^t D(\mathbf{e}) P_X D(\mathbf{e}) \ell}{\sigma^2}, \quad (3.38)
\end{aligned}$$

em que $P_X = X(X^t X)^{-1} X^t$ é a matriz de projeção ortogonal de \mathbf{R}^n sobre o subespaço vetorial gerado pelas colunas da matriz X .

Quando σ^2 não é conhecido, o vetor de parâmetros desconhecidos é denotado por $\theta = (\beta^t, \sigma^2)^t$. Observe que a parte relevante da função de log-verossimilhança do modelo perturbado fica expressa por

$$l(\theta, \omega) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i (y_i - x_i^t \beta)^2.$$

Tem-se, ainda, que

$$\frac{\partial l}{\partial \theta} \begin{bmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \sigma^2} \end{bmatrix}.$$

e, assim,

$$\begin{aligned}
\Delta &= \left[\frac{\partial}{\partial \omega^t} \left(\frac{\partial}{\partial \theta} \right) \right] \Big|_{\theta=\hat{\theta}, \omega=\omega_0} \\
&= \begin{bmatrix} \frac{\partial^2 l}{\partial \beta \partial \omega^t} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \omega^t} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} \end{bmatrix}, \quad (2.39)
\end{aligned}$$

Note que o bloco superior da matriz acima já foi calculado na equação (3.37). Precisa-se, ainda, de uma expressão útil para o bloco inferior. Assim,

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} - \sum_{i=1}^n \left(-\frac{1}{2\sigma^4} \right) \omega_i (y_i - x_i^t \beta)^2$$

$$\begin{aligned}
&= -\frac{n}{2\sigma^2} + \sum_{i=1}^n \left(\frac{1}{2\sigma^4}\right) \omega_i r_i^2 \\
&= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n \omega_i r_i^2 \\
&= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \omega^t r_{sq} ,
\end{aligned}$$

em que $r_{sq} = (r_1^2, r_2^2, \dots, r_n^2)^t$. Diferenciando-se esta derivada em relação ao vetor ω^t , tem-se

$$\begin{aligned}
(\partial^2 l) / (\partial \sigma^2 \partial \omega^t) &= \mathbf{0} + 1/(2\sigma^4) \partial / (\partial \omega^t) (\omega^t r_{sq}) \\
&= \frac{1}{2\sigma^4} r_{sq} .
\end{aligned}$$

Logo,

$$\frac{\partial^2 l}{\partial \sigma^2 \partial \omega^t} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} = \frac{1}{2\hat{\sigma}^4} e_{sq}^t , \quad (3.40)$$

em que $e_{sq} = (e_1^2, \dots, e_n^2)$ e $e_i^2 = r_i^2 \Big|_{\theta=\hat{\theta}}$.

Substituindo-se os resultados obtidos em (3.36) e (3.40) na equação (3.39), tem-se

$$\Delta = \begin{bmatrix} \frac{1}{\hat{\sigma}^2} X^t D(e) \\ \frac{1}{2\hat{\sigma}^4} e_{sq}^t \end{bmatrix} , \quad (3.41)$$

Observe que

$$\frac{\partial^2 l}{\partial \theta \partial \theta^t} = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta \partial \beta^t} & \frac{\partial^2 l}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \beta^t} & \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^t} \end{bmatrix} ,$$

assim,

$$\bar{L}(\hat{\theta}) = \begin{bmatrix} \frac{\partial^2 l}{\partial \beta \partial \beta^t} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} & \frac{\partial^2 l}{\partial \beta \partial \sigma^2} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \beta^t} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} & \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} \end{bmatrix}. \quad (3.42)$$

Dessa forma, segue que

$$(i) \quad \frac{\partial^2 l}{\partial \beta \partial \beta^t} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} = -\frac{1}{\hat{\sigma}^2} X^t X$$

(ii) Diferenciando-se a equação (3.35) tem-se que

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial \sigma^2} &= -\frac{1}{\sigma^4} \sum_{i=1}^n \omega_i (y_i - x_i^t \beta) x_i \\ &= -\frac{1}{\sigma^4} \sum_{i=1}^n \omega_i (x_i y_i - x_i x_i^t \beta) x_i \\ &= -\frac{1}{\sigma^4} \left\{ \sum_{i=1}^n x_i \omega_i y_i - \sum_{i=1}^n x_i \omega_i x_i^t \beta \right\} \\ &= -\frac{1}{\sigma^4} \{ X^t W y - (X^t W X) \beta \}, \end{aligned}$$

assim,

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta \partial \sigma^2} \Big|_{\hat{\theta}=\hat{\theta}, \omega=\omega_0} &= -\frac{1}{\hat{\sigma}^4} \{ X^t I y - (X^t I X) \hat{\beta} \} \\ &= -\frac{1}{\hat{\sigma}^4} \{ X^t y - (X^t X) \hat{\beta} \}, \end{aligned}$$

mas, para o modelo de regressão linear normal múltipla dado na equação (2.7),

tem-se que $\hat{\beta} = (X^t X)^{-1} X^t y$, logo,

$$\begin{aligned}\frac{\partial^2 l}{\partial \beta \partial \sigma^2} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} &= -\frac{1}{\hat{\sigma}^4} \{X^t y - (X^t X)(X^t X)^{-1} X^t y\} \\ &= -\frac{1}{\hat{\sigma}^4} \{X^t y - X^t y\} \\ &= \mathbf{0}_{p \times 1}\end{aligned}$$

$$(iii) \quad \frac{\partial^2 l}{\partial \sigma^2 \partial \beta^t} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} = \left[\frac{\partial^2 l}{\partial \beta \partial \sigma^2} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} \right]^t = \mathbf{0}_{p \times 1}^t = \mathbf{0}_{1 \times p}$$

(iv) Derivando-se a parte relevante da função de log-verossimilhança em relação a σ^2 , tem-se

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n w_i (y_i - X_i^t \beta)^2,$$

logo,

$$\frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n w_i (y_i - X_i^t \beta)^2.$$

Segue daí que

$$\begin{aligned}\frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \Big|_{\theta=\hat{\theta}, W=W_0} &= \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} \sum_{i=1}^n w_i (y_i - X_i^t \hat{\beta})^2 \\ &= \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} \sum_{i=1}^n \frac{(y_i - X_i^t \hat{\beta})^2}{n}.\end{aligned}$$

Note que, na verdade, a expressão $\frac{(y_i - X_i^t \hat{\beta})^2}{n}$ consiste no estimador de máxima verossimilhança do parâmetro σ^2 , ou seja, é a estimativa $\hat{\sigma}^2$. Portanto,

$$\begin{aligned}\frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \Big|_{\theta=\hat{\theta}, W=W_0} &= \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} \hat{\sigma}^2 \\ &= -\frac{n}{2\hat{\sigma}^4}.\end{aligned}$$

Substituindo-se os resultados nos itens (i), (ii), (iii) e (iv) na equação dada em (3.42), tem-se

$$L(\hat{\theta}) = \begin{bmatrix} -\frac{1}{\hat{\sigma}^2} X^t X & \mathbf{0}_{p \times 1} \\ \mathbf{0}_{1 \times p} & -\frac{n}{2\hat{\sigma}^4} \end{bmatrix}.$$

Logo, tem-se que

$$L^{-1}(\hat{\theta}) = - \begin{bmatrix} (X^t X)^{-1} \hat{\sigma}^2 & \mathbf{0}_{p \times 1} \\ \mathbf{0}_{1 \times p} & \frac{2\hat{\sigma}^4}{n} \end{bmatrix} \quad (3.43).$$

Finalmente, substituindo-se os resultados em (3.41) e (3.43) na fórmula da curvatura normal dada em (3.20), tem-se

$$\begin{aligned} C_l &= 2 \left| \ell^t \begin{bmatrix} \frac{1}{\hat{\sigma}^2} D(\mathbf{e}) X & \frac{1}{2\hat{\sigma}^4} \end{bmatrix} \begin{bmatrix} -(X^t X)^{-1} \hat{\sigma}^2 & \mathbf{0}_{p \times 1} \\ \mathbf{0}_{1 \times p} & -\frac{2\hat{\sigma}^4}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\sigma}^2} X^t D(\mathbf{e}) \\ \frac{1}{2\hat{\sigma}^4} e_{sq}^t \end{bmatrix} \right| \\ &= 2 \left| \ell^t \left[\frac{-D(\mathbf{e}) X (X^t X)^{-1} X^t D(\mathbf{e})}{\hat{\sigma}^2} - \frac{e_{sq} e_{sq}^t}{2n\hat{\sigma}^4} \right] \right| \\ &= \frac{2}{\hat{\sigma}^2} \left| \ell^t \left[D(\mathbf{e}) P_X D(\mathbf{e}) + \frac{e_{sq} e_{sq}^t}{2n\hat{\sigma}^2} \right] \right|. \end{aligned} \quad (3.44).$$

Vale ressaltar que não são conhecidas expressões analíticas para autores ζ_{\max} da matriz \bar{F} , quando a curvatura normal for dada pelas fórmulas estabelecidas nas equações (3.38) e (3.44) (COOK, 1986).

2.4.1 Avaliando coeficientes individuais

Frequentemente, torna-se interessante a análise individual dos coeficientes em um modelo linear (COOK, 1986). Nesta seção, estabeleceremos uma fórmula para a curvatura normal ser usada na avaliação da influência local sobre um coeficiente determinado.

Considere a seguinte participação para a matriz de variável explicativas $X = [X_1 X_2]$, em que a variável de interesse é representada pela primeira coluna

X_1 e o seu respectivo coeficiente é o β_1 . Seja r_{X_1} o vetor de resíduos da regressão da coluna X_1 sobre as colunas restantes, ou seja, a matriz X_2 . Assim,

$$r_{X_1} = X_1 - \hat{X}_1,$$

em que $\hat{X}_1 = X_2(X_2^t X_2)^{-1} X_2^t X_1$. Logo, pode-se escrever

$$\begin{aligned} r_{X_1} &= X_2(X_2^t X_2)^{-1} X_2^t X_1 \\ &= [I - X_2(X_2^t X_2)^{-1} X_2^t] X_1 \\ &= [I - P_{X_2}] X_1 \\ &= \hat{P}_{X_2} X_1, \end{aligned}$$

(3.45),

em que $\hat{P}_{X_2} = X_2(X_2^t X_2)^{-1} X_2^t$ é a matriz de projeção ortogonal sobre as colunas da matriz X_2 e $\hat{P}_{X_2} = I - P_{X_2}$.

Consideraremos a equação da curvatura normal a seguir (COOK, 1986):

$$C_t(\beta_1) = 2|\ell^t [\Delta^t \bar{L}^{-1} \Delta - \Delta^t B_{22} \Delta] \ell|. \quad (3.46)$$

em que $B_{22} = \begin{bmatrix} 0 & 0 \\ 0 & L_{22}^{-1} \end{bmatrix}$, e as matrizes Δ e \bar{L} foram definidas na seção anterior.

Pelos cálculos feitos para se obter a fórmula dada na equação (3.44), temos que

$$\Delta^t \bar{L}^{-1} \Delta = \frac{-D(e)X(X^t X)^{-1} X^t D(e)}{\bar{\sigma}^2} - \frac{e_{sq} e_{sq}^t}{en \bar{\sigma}^4}. \quad (3.47)$$

Note que podemos escrever a matriz Δ e a matriz $\bar{L}(\bar{\theta})$ por meio das seguintes partições:

$$\Delta = \begin{bmatrix} \frac{1}{\hat{\sigma}^2} X_1^t D(\mathbf{e}) \\ \frac{1}{\hat{\sigma}^2} X_2^t D(\mathbf{e}) \\ \frac{1}{2\hat{\sigma}^4} e_{sq}^t \end{bmatrix}.$$

$$\tilde{L}(\hat{\theta}) = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix},$$

em que

$$L_{11} = -\frac{1}{\hat{\sigma}^2} X_1^t X_1, L_{12} = \mathbf{0}_{1 \times p}, L_{21} = \mathbf{0}_{p \times 1}, L_{22} = \begin{bmatrix} -\frac{1}{\hat{\sigma}^2} X_2^t X_2 & \mathbf{0}_{(p-1) \times 1} \\ \mathbf{0}_{1 \times (p-1)} & -\frac{n}{2\hat{\sigma}^4} \end{bmatrix}.$$

De onde obtém-se que

$$L_{22}^{-1} = \begin{bmatrix} (X_2^t X_2)^{-1} \hat{\sigma}^2 & \mathbf{0}_{(p-1) \times 1} \\ \mathbf{0}_{1 \times (p-1)} & -\frac{2\hat{\sigma}^4}{n} \end{bmatrix}.$$

Quando substituímos a matriz B_{22} , juntamos com a partição apresentada para a matriz Δ anteriormente, na expressão $\Delta^t B_{22} \Delta$, e obtém-se que

$$\Delta^t B_{22} \Delta = \frac{-D(\mathbf{e}) X (X_2^t X_2)^{-1} X_2^t D(\mathbf{e})}{\hat{\sigma}^2} - \frac{e_{sq} e_{sq}^t}{en\hat{\sigma}^4}.$$

3.48)

Substituindo-se os resultados obtidos em (3.47) e (3.48) na equação (3.46), tem-se que

$$\begin{aligned} C_i(\beta_1) &= 2 \left| \ell^t \frac{-D(\mathbf{e})}{\hat{\sigma}^2} \left[X(X^t X)^{-1} X^t - X_2 (X_2^t X_2)^{-1} X_2^t \right] D(\mathbf{e}) \ell \right| \\ &= \frac{2}{\hat{\sigma}^2} \left| \ell^t D(\mathbf{e}) \left[X(X^t X)^{-1} X^t - (X_1 X_2) \begin{pmatrix} \mathbf{0}_{1 \times 1} & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & (X_2^t X_2)^{-1} \end{pmatrix} \begin{pmatrix} X_1^t \\ X_2^t \end{pmatrix} \right] D(\mathbf{e}) \ell \right| \\ &= \frac{2}{\hat{\sigma}^2} \left| \ell^t D(\mathbf{e}) \left[X(X^t X)^{-1} X^t - X A_{22} X^t \right] D(\mathbf{e}) \ell \right| \\ &= \frac{2}{\hat{\sigma}^2} \left| \ell^t D(\mathbf{e}) X \left[(X^t X)^{-1} - A_{22} \right] X^t D(\mathbf{e}) \ell \right| \end{aligned}$$

$$= \frac{2}{\sigma^2} |e^t D(e) N D(e) e|$$

$$N = X[(X^t X)^{-1} - A_{22}]X^t e A_{22} = \begin{pmatrix} \mathbf{0}_{1 \times 1} & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & (X_2^t X_2)^{-1} \end{pmatrix}.$$

em que

Utilizando-se a partição da matriz X mencionada no início desta subseção, tem-se que

$$N = [X_1 X_2] [(X^t X)^{-1} - A_{22}] [X_1^t X_2^t]^t$$

$$X^t X = \begin{bmatrix} X_1^t X_1 & X_1^t X_2 \\ X_2^t X_1 & X_2^t X_2 \end{bmatrix}$$

Seja

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = X^t X,$$

então $R_{11} = X_1^t X_1$, $R_{12} = X_1^t X_2$, $R_{21} = X_2^t X_1$ e $R_{22} = X_2^t X_2$, denotando a inversa da matriz R

$$R^{-1} = \begin{bmatrix} R^{11} & R^{12} \\ R^{21} & R^{22} \end{bmatrix}.$$

Agora, vamos aplicar a forma padrão da inversa de uma matriz particionada nas mesmas condições da R , mostrada no teorema (A.2.2), no Apêndice (A), de onde segue que

$$(i) \quad R^{11} = [R_{11} - R_{12} R_{22}^{-1} R_{21}]$$

$$= [X_1^t X_1 - X_1^t X_2 (X_2^t X_2)^{-1} X_2^t X_1]^{-1}$$

$$= [X_1^t \{I - P_1(X_2)\} X_1]^{-1}$$

$$= [X_1^t \{I - P_1(X_2)\} X_1]^{-1}$$

$$= [X_1^t \{I - P_1(X_2)\} X_1]^{-1}$$

como $\bar{\mathbf{P}}_{\mathbf{X}_2}$ é uma matriz idempotente, tem-se que

$$\mathbf{R}^{11} = \left[\mathbf{X}_1^t \bar{\mathbf{P}}_{\mathbf{X}_2} \bar{\mathbf{P}}_{\mathbf{X}_3} \mathbf{X}_1 \right]^{-1}$$

Logo, usando o resultado (3.45), tem-se

$$\begin{aligned} \mathbf{R}^{11} &= \left[r_{\mathbf{X}_1}^t r_{\mathbf{X}_1} \right]^{-1} \\ &= \left[\left\| r_{\mathbf{X}_1} \right\|^2 \right]^{-1} \\ &= \frac{1}{\left\| r_{\mathbf{X}_1} \right\|^2} \end{aligned} \quad (ii)$$

$$\begin{aligned} \mathbf{R}^{22} &= \left[\mathbf{R}_{22} - \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \mathbf{R}_{12} \right]^{-1} \\ &= \left[\mathbf{R}_{22} + \mathbf{R}_{21} \left(-\mathbf{R}_{11}^{-1} \right) \mathbf{R}_{12} \right]^{-1}, \end{aligned}$$

Aplicando-se o resultado do teorema (A.2.4), descrito no apêndice A, e operando algumas simplificações, tem-se

$$\mathbf{R}^{22} =$$

$$\mathbf{R}_{22}^{-1} - \mathbf{R}_{22}^{-1} \mathbf{R}_{21} \mathbf{E} \mathbf{R}_{12} (\mathbf{R}_{22}^t)^{-1} + \mathbf{R}_{22}^{-1} \mathbf{R}_{21} \mathbf{E} \left[\mathbf{E} + \mathbf{R}_{22}^{-1} \right]^{-1} \mathbf{E}^t \mathbf{R}_{12} (\mathbf{R}_{22}^t)^{-1}$$

em que $\mathbf{E} = \left[\mathbf{R}_{12} \mathbf{R}_{22}^{-1} \mathbf{R}_{21}^t \right]^{-1}$. Segue daí que

$$\begin{aligned} \mathbf{E} &= \left[\mathbf{X}_1^t \mathbf{X}_2 (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1 \right]^{-1} \\ &= \left[\mathbf{X}_1^t \mathbf{P}_{\mathbf{X}_2} \mathbf{X}_1 \right]^{-1}. \end{aligned}$$

Usando o fato de $\mathbf{P}_{\mathbf{X}_2}$ ser idempotente, pode-se escrever

$$\mathbf{E} = \left[\mathbf{X}_1^t \mathbf{P}_{\mathbf{X}_2} \mathbf{P}_{\mathbf{X}_2} \mathbf{X}_1 \right]^{-1}.$$

Denotando os valores preditos da regressão de \mathbf{X}_1 sobre \mathbf{X}_2 por $\hat{\mu}_{\mathbf{X}_1}$, segue, do capítulo (1), que

$$\begin{aligned}
\mathbf{E} &= \left[\left[\hat{\mu}_{X_1}^t \hat{\mu}_{X_1} \right] \right]^{-1} \\
&= \left[\left[\|\hat{\mu}_{X_1}\|^2 \right] \right]^{-1} \\
&= \frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2}.
\end{aligned}$$

Substituindo o resultado acima e o restante dos elementos na última expressão para a matriz \mathbf{R}^{2z} , tem-se

$$\begin{aligned}
\mathbf{R}^{2z} &= \left[(\mathbf{X}_2^t \mathbf{X}_2)^{-1} - \left[(\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1 \left(\frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} \right) \mathbf{X}_1^t \mathbf{X}_2 \right] (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \right. \\
&\quad \left. + (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1 \left(\frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} \right) \left[\frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} - \left[(\mathbf{X}_1^t \mathbf{X}_1)^{-1} \right] \right]^{-1} \frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} \mathbf{X}_1^t \mathbf{X}_2 \right] (\mathbf{X}_2^t \mathbf{X}_2)^{-1}
\end{aligned}$$

note que $\mathbf{X}_1^t \mathbf{X}_1 = \|\mathbf{X}_1\|^2$. Assim, pode-se escrever

$$\begin{aligned}
\mathbf{R}^{2z} &= \left[(\mathbf{X}_2^t \mathbf{X}_2)^{-1} - \left[(\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1 \right] \left[(\mathbf{X}_1^t \mathbf{X}_1)^{-1} \right] \left[(\mathbf{X}_1^t \mathbf{X}_1)^{-1} \right] \right. \\
&\quad \left. \left\{ \frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} - \frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} \left[\frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} - \frac{\mathbf{1}}{\|\mathbf{X}_1\|^2} \right]^{-1} \frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} \right\} \right].
\end{aligned}$$

Seja $\hat{\beta}_{X_1}$ o vetor de parâmetros estimados da regressão de \mathbf{X}_1 por \mathbf{X}_2 ,

ou seja, $\hat{\beta}_{X_1} = \left[(\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1 \right]$. Logo,

$$\begin{aligned}
\mathbf{R}^{2z} &= \\
&\left[(\mathbf{X}_2^t \mathbf{X}_2)^{-1} - \hat{\beta}_{X_1} \hat{\beta}_{X_1}^t \right] \left\{ \frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} - \frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} \frac{\|\mathbf{X}_1\|^2 \|\hat{\mu}_{X_1}\|^2}{\|\mathbf{X}_1\|^2 - \|\hat{\mu}_{X_1}\|^2} \frac{\mathbf{1}}{\|\hat{\mu}_{X_1}\|^2} \right\}
\end{aligned}$$

$$= \left[(\mathbf{X}_2^t \mathbf{X}_2)^{-1} - \hat{\boldsymbol{\beta}}_{\mathbf{X}_1} \hat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \frac{1}{\|\mathbf{X}_1\|^2 - \|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} \right].$$

Observe, ainda, que, pelo item (i) anterior, tem-se que

$$\frac{1}{\|r_{\mathbf{X}_1}\|^2} = \left[\mathbf{X}_1^t \mathbf{X}_1 - \mathbf{X}_1^t \mathbf{P}_{\mathbf{X}_1} \mathbf{X}_1 \right]^{-1}.$$

Utilizando, novamente, a idempotência da matriz $\mathbf{P}_{\mathbf{X}_1}$, segue que

$$\begin{aligned} \frac{1}{\|r_{\mathbf{X}_1}\|^2} &= \left[\|\mathbf{X}_1\|^2 - \mathbf{X}_1^t \mathbf{P}_{\mathbf{X}_1} \mathbf{P}_{\mathbf{X}_1} \mathbf{X}_1 \right]^{-1} \\ &= \left[\|\mathbf{X}_1\|^2 - \hat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \hat{\boldsymbol{\beta}}_{\mathbf{X}_1} \right]^{-1} \\ &= \frac{1}{\|\mathbf{X}_1\|^2 - \|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2}. \end{aligned}$$

Dai conclui-se que

$$\|r_{\mathbf{X}_1}\|^2 = \|\mathbf{X}_1\|^2 - \|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2.$$

Para finalizar, substitui-se o resultado acima na última expressão determinada para \mathbf{R}^{22} , de onde se obtém

$$\begin{aligned} \mathbf{R}^{22} &= \left[(\mathbf{X}_2^t \mathbf{X}_2)^{-1} + \hat{\boldsymbol{\beta}}_{\mathbf{X}_1} \hat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \frac{1}{\|r_{\mathbf{X}_1}\|^2} \right]. \\ \text{(ii) } \mathbf{R}^{12} &= -\mathbf{R}_{11}^{-1} - \mathbf{R}_{12} \mathbf{R}^{22} \\ &= -(\mathbf{X}_1^t \mathbf{X}_1)^{-1} \mathbf{X}_1^t \mathbf{X}_2 \left\{ (\mathbf{X}_2^t \mathbf{X}_2)^{-1} + \frac{1}{\|r_{\mathbf{X}_1}\|^2} \hat{\boldsymbol{\beta}}_{\mathbf{X}_1} \hat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \right\} \\ &= -\frac{1}{\|r_{\mathbf{X}_1}\|^2} \left[\mathbf{X}_1^t (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \right] - \frac{1}{\|\mathbf{X}_1\|^2 \|r_{\mathbf{X}_1}\|^2} \mathbf{X}_1^t \hat{\boldsymbol{\beta}}_{\mathbf{X}_1} \hat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \end{aligned}$$

$$= \left[\frac{\|r_{X_1}\|^2 + X_2^t X_2 \hat{\beta}_{X_1}}{\|X_1\|^2} \right] \left(-\frac{\hat{\beta}_{X_1}^t}{\|r_{X_1}\|^2} \right).$$

Denota-se por b o numerador da expressão entre colchetes no resultado acima. Logo, tem-se que

$$b = r_{X_1}^t r_{X_1} + X_1^t X_2 \hat{\beta}_{X_1}.$$

Substituindo r_{X_1} pela sua expressão dada na equação (3.45) e $\hat{\beta}_{X_1}$ pela forma definida no item (ii), no resultado acima, tem-se

$$\begin{aligned} b &= X_1^t P_{X_2} P_{X_2} X_1 + X_1^t X_2 (X_2^t X_2)^{-1} X_2^t X_1 \\ &= X_1^t P_{X_2} X_1 + X_1^t P_{X_2} X_1 \\ &= X_1^t [P]_{X_2} + [P_{X_2}] X_1 \\ &= X_1^t I X_1 \\ &= \|X_1\|^2. \end{aligned}$$

Agora, substituindo-se o resultado de b acima na última expressão encontrada para R^{12} , segue que

$$R^{12} = -\frac{\hat{\beta}_{X_1}^t}{\|r_{X_1}\|^2}.$$

$$(iv) \quad R^{21} = -R_{22}^{-1} - R_{21} R^{11}$$

$$\begin{aligned} &= -(X_2^t X_2)^{-1} X_2^t X_1 \left(\frac{1}{\|r_{X_1}\|^2} \right) \\ &= -\frac{\beta_{X_1}}{\|r_{X_1}\|^2}. \end{aligned}$$

Substituindo-Se os resultados encontrados nos itens (i), (ii), (iii) e (iv) nas entradas da matriz R^{-1} , conclui-se que

$$\mathbf{R}^{-1} = \begin{bmatrix} \frac{1}{\|r_{X_1}\|^2} & -\frac{\hat{\beta}_{X_1}^t}{\|r_{X_1}\|^2} \\ -\frac{\hat{\beta}_{X_1}}{\|r_{X_1}\|^2} & [(\mathbf{X}_2^t \mathbf{X}_2)^{-1} + \frac{1}{\|r_{X_1}\|^2} \hat{\beta}_{X_1} \hat{\beta}_{X_1}^t] \end{bmatrix}$$

Substituindo o resultado anterior, a inversa da matriz \mathbf{R} , na equação de \mathbf{N} e fazendo-se algumas simplificações, obtém-se

$$\mathbf{N} = \frac{1}{\|r_{X_1}\|^2} \{ [\mathbf{X}_1 - \mathbf{X}_2 \hat{\beta}_{X_1}] [\mathbf{X}_1^t - \hat{\beta}_{X_1}^t \mathbf{X}_2^t] \}$$

Note que $\mathbf{X}_2 \hat{\beta}_{X_1}$ corresponde ao vetor de valores preditos da regressão de \mathbf{X}_1 sobre \mathbf{X}_2 , ou seja $\hat{\mu}_{X_1}$. Assim, pode-se escrever

$$\mathbf{N} = \frac{1}{\|r_{X_1}\|^2} \{ [\mathbf{X}_1 - \hat{\mu}_{X_1}] [\mathbf{X}_1^t - \hat{\mu}_{X_1}^t] \}.$$

Tem-se, ainda, que os elementos do vetor $\mathbf{X}_1 - \hat{\mu}_{X_1}$, na última equação, representam os resíduos da regressão. Portanto, segue que

$$\mathbf{N} = \frac{1}{\|r_{X_1}\|^2} \{ r_{X_1} r_{X_1}^t \}.$$

Finalmente, substituindo-se a expressão encontrada para \mathbf{N} , na equação (3.49), obtém-se a fórmula para a curvatura normal a seguir

$$C_i(\beta_1) = \frac{2}{\sigma^2} \left| \frac{\ell^t D(\mathbf{e}) (r_{X_1} r_{X_1}^t) D(\mathbf{e}) \ell}{\|r_{X_1}\|^2} \right|.$$

Para se obter o valor máximo que a curvatura normal C_i pode assumir, quando está sujeita à condição de que $\|\ell\| = 1$, deve-se observar que, na

verdade, a expressão $r_{X_1}^t D(e)\ell$ da equação (3.50) é um escalar. Logo, é igual ao seu transposto. Portanto, podemos reescrever a referida equação como

$$C_i(\beta_1) = \frac{2(r_{X_1}^t D(e)\ell)^2}{\|r_{X_1}\|^2 \hat{\sigma}^2}.$$

Veja, ainda, que o estimador de máxima verossimilhança de σ^2 é dado

por $\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_i^2}{n}$ e que $\|r_{X_1}\|^2 = \sum_{i=1}^n r_i^2$. Dessa forma, segue que

$$\begin{aligned} C_i(\beta_1) &= \frac{2(r_{X_1}^t D(e)\ell)^2}{\sum_{i=1}^n r_i^2 \frac{\sum_{i=1}^n e_i^2}{n}} \\ &= \frac{2n(r_{X_1}^t D(e)\ell)^2}{\sum_{i=1}^n r_i^2 \sum_{i=1}^n e_i^2}. \end{aligned}$$

Note que, para obter o valor máximo da curvatura normal acima, basta analisar a expressão $(r_{X_1}^t D(e)\ell)^2$, pois o resultante da fórmula é constante. Assim, seja a função vetorial

$$f(\ell) = (r_{X_1}^t D(e)\ell)^2.$$

Aplicando-se a desigualdade de Cauchy-Schwartz descrita no teorema (A.2.1), no Apêndice (A), na função f definida acima, tem-se que

$$f(\ell) = (r_{X_1}^t D(e)\ell)^2 \leq (r_{X_1}^t D(e)D(e)r_{X_1})(\ell^t \ell),$$

como o vetor ℓ está sujeito a $\ell^t \ell = 1$, segue que

$$\begin{aligned} f(\ell) &\leq (r_{X_1}^t D(e)D(e)r_{X_1}) \\ &\leq \|D(e)r_{X_1}\|^2 \end{aligned}$$

$$\leq \sum_{i=1}^n e_i^2 r_i^2,$$

sendo que a igualdade na expressão acima ocorre somente quando é satisfeita a seguinte condição:

$$\boldsymbol{\ell} = c\mathbf{D}(\mathbf{e})\mathbf{r}_{X_1},$$

em que c é um escalar. Dessa forma, a função f assume o valor máximo, ou seja,

$$f(\boldsymbol{\ell}) = \sum_{i=1}^n e_i^2 r_i^2$$

quando o vetor $\boldsymbol{\ell}$ está na direção do vetor $\mathbf{D}(\mathbf{e})\mathbf{r}_{X_1}$. Portanto, conclui-se, daí, que

$$C_{\max(\beta_1)} = \frac{2n \sum_{i=1}^n e_i^2 r_i^2}{\sum_{i=1}^n r_i^2 \sum_{i=1}^n e_i^2} \quad (3.51)$$

é a curvatura máxima, que obviamente ocorre na direção

$$\boldsymbol{\ell}_{\max} = \mathbf{D}(\mathbf{e})\mathbf{r}_{X_1} \quad (3.52).$$

Elementos do autovetor $\boldsymbol{\ell}_{\max}$ relativamente grandes ou pequenos correspondem às observações que possuem $|r_j|$ e $|e_j|$ simultaneamente grandes. Portanto, por meio de uma inspeção do vetor $\boldsymbol{\ell}_{\max}$ podemos identificar quais observações contribuem substancialmente com o autovalor C_{\max} (COOK, 1986).

2.4.2 Perturbação de observações na família exponencial

Nesta seção, serão estendidos alguns resultados anteriores para outros modelos além do normal linear, cujas funções densidades ou de probabilidade sejam membros da família exponencial.

Considere, agora, uma amostra aleatória y_1, y_2, \dots, y_n de observações independentes retiradas de uma variável cuja função densidade ou probabilidade pertença à família exponencial, isto é, seja dada pela equação (2.10). Logo, pode-se escrever a função de log-verossimilhança do vetor $\mathbf{y} = [(y_1, y_2, \dots, y_n)]^t$, quando atribuímos um peso ω_i à i -ésima observação, da seguinte forma:

$$l(\boldsymbol{\beta}, \phi | \omega) = \phi \sum_{i=1}^n \omega_i (y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n \omega_i c(y_i, \phi),$$

em que ϕ^{-1} é o parâmetro de escala ($\phi > 0$), $\boldsymbol{\beta}$ é o vetor de parâmetros desconhecidos e $\theta_i = \theta_i(\boldsymbol{\beta})$.

Observe os seguintes resultados:

(i) Temos que $\boldsymbol{\theta} = \theta_i, i = 1, \dots, n$, assim,

$$\dot{\boldsymbol{\theta}} = \frac{\partial \boldsymbol{\theta}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \left(\frac{\partial \theta_i}{\partial \beta_j} \right), i = 1, \dots, n; j = 1, \dots, p$$

e

$$\dot{\theta}_i = \frac{\partial \theta_i}{\partial \boldsymbol{\beta}} = \left(\frac{\partial \theta_i}{\partial \beta_j} \right), j = 1, \dots, p.$$

Segue que

$$\ddot{\theta}_i = \frac{\partial^2 \theta_i}{\partial \boldsymbol{\beta}^t \partial \boldsymbol{\beta}} = \left(\frac{\partial^2 \theta_i}{\partial \beta_j \partial \beta_k} \right), j = 1, \dots, p; k = 1, \dots, p.$$

(ii) Temos que $\mathbf{b} = (b_i(\theta_i)), i = 1, \dots, n$. Daí segue que

$$\dot{\mathbf{b}} = \frac{\partial \mathbf{b}}{\partial \boldsymbol{\theta}} = \left(\frac{\partial b_i}{\partial \theta_j} \right), i = 1, \dots, n; j = 1, \dots, n.$$

Mas, observe que $\frac{\partial b_i}{\partial \theta_j} = \begin{cases} 1, & \text{se } i = j \\ 0, & \text{se } i \neq j \end{cases}$. Assim,

pode-se escrever

$$\dot{\mathbf{b}} = \text{diag} \left(\frac{\partial b_i}{\partial \theta_j} \right), i = 1, \dots, n.$$

Logo, tem-se que

$$\ddot{\mathbf{b}} = \frac{\partial^2 \mathbf{b}}{\partial \theta^2} = \text{diag} \left(\frac{\partial^2 b_i}{\partial \theta_j^2} \right), i = 1, \dots, n.$$

Pela equação (3.19) tem-se a seguinte relação:

$$\ddot{\mathbf{F}} = \mathbf{\Delta}^t \ddot{\mathbf{L}}^{-1} \mathbf{\Delta},$$

em que o elemento (r,s) da matriz $\mathbf{\Delta}$ é dado por

$$\Delta_{rs} = \frac{\partial^2 l}{\partial \beta_r \partial \omega_s} \Big|_{\beta = \hat{\beta}, \omega = \omega_0}.$$

E a matriz $\ddot{\mathbf{L}}$ é

$$\ddot{L}_{rs} = \frac{\partial^2 l}{\partial \beta_r \partial \beta_s} \Big|_{\beta = \hat{\beta}, \omega = \omega_0}$$

Segue que, derivando-se a função de log-verossimilhança dada na equação (3.53), em relação a ω_s , obtém-se

$$\frac{\partial l}{\partial \omega_s} = \phi y_i \theta_i - \phi b(\theta_s) + c(y_s, \phi).$$

Agora derivamos a equação acima em relação a β_r , de onde se tem

$$\begin{aligned} \Delta_{rs} &= \phi y_s \frac{\partial \hat{\theta}_s}{\partial \hat{\beta}_r} - \phi \frac{\partial b_s}{\partial \hat{\theta}_s} \frac{\partial \hat{\theta}_s}{\partial \hat{\beta}_r} \\ &= \phi \left(y_s - \phi \frac{\partial b_s}{\partial \hat{\theta}_s} \right) \frac{\partial \hat{\theta}_s}{\partial \hat{\beta}_r}. \end{aligned}$$

Portanto, a matriz $\mathbf{\Delta}$ pode ser escrita da seguinte forma:

$$\begin{aligned} \Delta &= \begin{bmatrix} \phi \left(y_1 - \frac{\partial b_1}{\partial \theta_1} \right) \frac{\partial \theta_1}{\partial \beta_1} & \cdots & \phi \left(y_n - \frac{\partial b_n}{\partial \theta_n} \right) \frac{\partial \theta_n}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ \phi \left(y_1 - \frac{\partial b_1}{\partial \theta_1} \right) \frac{\partial \theta_1}{\partial \beta_p} & \cdots & \phi \left(y_n - \frac{\partial b_n}{\partial \theta_n} \right) \frac{\partial \theta_n}{\partial \beta_p} \end{bmatrix} \\ &= \phi \begin{bmatrix} \frac{\partial \theta_1}{\partial \beta_1} & \cdots & \frac{\partial \theta_n}{\partial \beta_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \theta_1}{\partial \beta_p} & \cdots & \frac{\partial \theta_n}{\partial \beta_p} \end{bmatrix} \begin{bmatrix} y_1 - \frac{\partial b_1}{\partial \theta_1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & y_n - \frac{\partial b_n}{\partial \theta_n} \end{bmatrix} \\ &= \phi \tilde{\theta}^t \text{diag} \left(y_i - \frac{\partial b_i}{\partial \theta_i} \right), \text{ com } i = 1, \dots, n, \end{aligned}$$

em que $\tilde{\theta}$ é a matriz θ expressa no item (i), avaliada em $\hat{\beta}$ e ω_0 . Seja

$$u_i = y_i - \frac{\partial b_i}{\partial \theta_i} \text{ e } D(u) = \text{diag}(u_i), \text{ então se pode escrever}$$

$$\Delta = \phi \tilde{\theta}^t D(u). \quad (3.54)$$

Quando derivamos a função de log-verossimilhança na equação (3.53) em relação a β_r , obtém-se

$$\frac{\partial l}{\partial \beta_s} = \phi \sum_{i=1}^n \left(y_i - \frac{\partial b_i}{\partial \theta_i} \right) \frac{\partial \theta_i}{\partial \beta_r}.$$

Segue que, derivando este resultado em relação a β_s , tem-se

$$\tilde{L}_{rs} = \phi \sum_{i=1}^n \left\{ - \frac{\partial^2 b_i}{\partial \theta_i^2} \frac{\partial \tilde{\theta}_i}{\partial \beta_r} \frac{\partial \tilde{\theta}_i}{\partial \beta_s} + u_i \frac{\partial^2 \theta_i}{\partial \beta_r \partial \beta_s} \right\}.$$

Note que $\frac{\partial^2 \theta_i}{\partial \beta_r \partial \beta_s}$ corresponde ao elemento (r,s) da matriz

$$\tilde{\theta}_i |_{\beta=\hat{\beta}, \omega=\omega_0} = \tilde{\theta}_i^t \frac{\partial \tilde{\theta}_i}{\partial \beta_r}$$

$$e \frac{\partial \hat{\theta}_i}{\partial \hat{\beta}_s}$$

São, respectivamente, os elementos (i, r) e (i, s) da matriz

$\hat{\theta}_i \Big|_{\hat{\beta}=\hat{\beta}_0, \omega=\omega_0} = \hat{\theta}_i$, em que $\hat{\theta}_i$ e $\hat{\theta}_i$ foram expressas no item (i) desta seção.

Observe, ainda, que $\frac{\partial^2 b_i}{\partial \hat{\theta}_i^2}$ é o elemento (i, i) da matriz \hat{b} , dada no item (ii), no início desta seção. Assim, pode-se escrever

$$\hat{L}_{rs} = \varphi \sum_{i=1}^n \left\{ u_i (\hat{\theta}_i)_{rs} - \hat{\theta}_{ir} \hat{b}_{ii} \hat{\theta}_{is} \right\}.$$

Pela simetria da matriz $\hat{\theta}$, tem-se que $\hat{\theta}_{ir} = \hat{\theta}_{ri}$. Logo,

$$\hat{L}_{rs} = \varphi \sum_{i=1}^n u_i (\hat{\theta}_i)_{rs} - \varphi \sum_{i=1}^n \hat{\theta}_{ri} \hat{b}_{ii} \hat{\theta}_{is}.$$

Portanto, a matriz \hat{L} pode ser escrita na seguinte forma matricial

$$\hat{L} = \varphi \sum_{i=1}^n u_i \hat{\theta}_i - \varphi \hat{\theta}^c \hat{b} \hat{\theta} \quad (3.55)$$

Para finalizar, substituem-se os resultados das equações (3.54) e (3.55) em (3.19). Com isso, obtém-se

$$\begin{aligned} \hat{F} &= \varphi D(u) \hat{\theta} \left[\varphi \left(\sum_{i=1}^n u_i \hat{\theta}_i - \hat{\theta}^c \hat{b} \hat{\theta} \right) \right]^{-1} \varphi \hat{\theta}^c D(u) \\ &= \varphi D(u) \hat{\theta} \left[\sum_{i=1}^n u_i \hat{\theta}_i - \hat{\theta}^c \hat{b} \hat{\theta} \right]^{-1} \hat{\theta}^c D(u). \end{aligned} \quad (3.56)$$

Conseqüentemente, obtém-se a curvatura normal, substituindo-se o resultado acima na equação (3.13), dada na seção (2.3).

Muitos MLGs são casos especiais da função de log-verossimilhança em (3.53) (COOK, 1986). Se considerarmos a condição $\theta_i(\boldsymbol{\beta}) = g(\mathbf{x}_i^t \boldsymbol{\beta})$, em que $i = 1, \dots, n$ e g é uma função de ligação, seguem as seguintes adaptações dos resultados anteriores desta seção.

Seja $\eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$ o preditor linear, definido no capítulo (1). Observe que $\dot{\boldsymbol{\theta}} = \frac{\partial \boldsymbol{\theta}}{\partial \boldsymbol{\beta}} \left(\frac{\partial g(\eta_i)}{\partial \beta_i} \right), i = 1, \dots, n; j = 1, \dots, p$.

Usando a regra da cadeia, tem-se que o elemento (r,s) da matriz $\dot{\boldsymbol{\theta}}$ pode ser escrito como

$$\dot{\theta}_{rs} = \frac{\partial g(\eta_r)}{\partial \beta_s} = \frac{\partial g}{\partial \eta_r} \frac{\partial \eta_r}{\partial \beta_s}$$

mas, $\eta_r = x_{r1}\beta_1 + \dots + x_{rs}\beta_s + \dots + x_{rp}\beta_p$. Logo, $\frac{\partial \eta_r}{\partial \beta_s} = x_{rs}$. Assim, tem-se que

$$\dot{\theta}_{rs} = \dot{g}_r x_{rs},$$

em que \dot{g}_r é a derivada $\frac{\partial g}{\partial \eta_r}$. Logo, pode-se expressar a matriz $\dot{\boldsymbol{\theta}}$ como

$$\begin{aligned} \dot{\boldsymbol{\theta}} &= \begin{bmatrix} \dot{g}_1 x_{11} & \dots & \dot{g}_1 x_{1p} \\ \vdots & \ddots & \vdots \\ \dot{g}_n x_{n1} & \dots & \dot{g}_n x_{np} \end{bmatrix} \\ &= \begin{bmatrix} \dot{g}_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \dot{g}_n \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \\ &= \text{diag}(\dot{g}_i) \mathbf{X}, i = 1, \dots, n. \end{aligned} \quad (3.57)$$

Quando g é a ligação canônica, tem-se que $\dot{\theta}_i(\boldsymbol{\beta}) = \eta_i = \mathbf{x}_i^t \boldsymbol{\beta}$, $\dot{\boldsymbol{\eta}} = \mathbf{X}$ e, conseqüentemente, $\ddot{\boldsymbol{\eta}} = \mathbf{0}$.

Agora iremos obter a matriz $\ddot{\boldsymbol{\theta}}_i$. Note que

$$\bar{\theta}_i = \frac{\partial^2 \theta_i}{\partial \beta \partial \beta^t} = \left(\frac{\partial^2 \theta_i}{\partial \beta_k \partial \beta_j} \right), k = \dots, p; j = 1, \dots, p.$$

Assim, o elemento (r,s) da matriz acima pode ser expresso por

$$\begin{aligned} (\bar{\theta}_i)_{rs} &= \frac{\partial}{\partial \beta_s} \left(\frac{\partial \theta_i}{\partial \beta_r} \right) \\ &= \frac{\partial}{\partial \beta_s} \left(\frac{\partial g}{\partial \eta_i} x_{ir} \right) \\ &= \frac{\partial^2 g}{\partial \eta_i^2} \frac{\partial \eta_i}{\partial \beta_s} x_{ir}. \end{aligned}$$

Observe que $\frac{\partial^2 g}{\partial \eta_i^2}$ é o i-ésimo termo elemento do vetor $\bar{\mathbf{g}}$. Logo, temos que

$$(\bar{\theta}_i)_{rs} = \bar{g}_i x_{is} x_{ir}.$$

Dessa forma, podemos escrever a matriz $\bar{\theta}_i$ como

$$\bar{\theta}_i = \bar{g}_i \mathbf{x}_i \mathbf{x}_i^t. \quad (3.58)$$

2.4.3 Observações individuais em regressão logística

Seja y_1, \dots, y_n uma amostra aleatória de variáveis independentes, com $y_i \sim \text{Binomial}(m_i^i, p_i^i)$, para $i = 1, \dots, n$. Assim a função de log-verossimilhança perturbada do vetor $\mathbf{y} = (y_1, \dots, y_n)^t$ é dada por

$$l(\beta, \mathbf{y} | \omega) = \sum_{i=1}^n \left[\omega_i y_i \ln \left(\frac{p_i}{1 - p_i} \right) + \omega_i m_i \ln(1 - p_i) \right],$$

em que ω é o vetor de parâmetros desconhecidos do modelo e $p_i^i = P[y_i^i = 1 | \mathbf{x}_i^i]$. Pela seção (2.3) do capítulo (2), sabe-se que

$$\ln \left[\frac{p_i}{1 - p_i} \right] = \eta_i.$$

Segue daí que

$$p_i = \frac{e^{n_i}}{1 + e^{n_i}}$$

e

$$1 - p_i = \frac{1}{1 + e^{n_i}}.$$

Portanto, pode-se escrever a função de log-verossimilhança como

$$l(\beta, y | \omega) = \sum_{i=1}^n \left[\omega_i y_i n_i + \omega_i m_i l_n \left(\frac{1}{1 + e^{n_i}} \right) \right].$$

Derivando a log-verossimilhança acima em relação a β_s , tem-se

$$\begin{aligned} \frac{\partial l}{\partial \beta_s} &= \sum_{i=1}^n \omega_i y_i \frac{\partial n_i}{\partial \beta_s} + \sum_{i=1}^n \omega_i m_i \frac{\partial}{\partial \beta_s} l_n \left(\frac{1}{1 + e^{n_i}} \right) \\ &= \sum_{i=1}^n \omega_i y_i x_{is} - \sum_{i=1}^n \omega_i y_i x_{is} \left(\frac{1}{1 + e^{n_i}} \right) \\ &= \sum_{i=1}^n \omega_i y_i x_{is} - \sum_{i=1}^n \omega_i m_i x_{is} p_i \end{aligned} \quad (3.59)$$

Agora derivamos o resultado acima em relação a ω_r , de onde se obtém que

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_s \partial \omega_r} &= \omega_r y_r x_{rs} - \omega_r m_r x_{rs} p_r \\ &= \omega_r x_{rs} (y_r - m_r p_r). \end{aligned}$$

Avaliando a derivada acima em $\hat{\theta} = (\hat{\beta}^t, \hat{p}_i)^t$ e $\omega = \omega_0$, obtém-se o elemento (r, s) da matriz Δ , ou seja,

$$\begin{aligned} \text{EMBED Equation. DSMT4} \quad \Delta_{rs} &= \left. \frac{\partial^2 l}{\partial \beta_s \partial \omega_r} \right|_{\theta = \hat{\theta}, \omega = \omega_0} \\ &= x_{rs} (y_r - m_r \hat{p}_r). \end{aligned}$$

Seja $v_i = m_i \hat{\beta}_i (1 - \hat{\beta}_i)$, $i = 1, \dots, n$. Logo, podemos escrever

$$\begin{aligned} \square \text{ EMBED Equation. DSMT4 } \square \square \Delta_{rs} &= x_{rs} v_r^{1/2} \left(\frac{y_r - m_r \hat{\beta}_r}{v_r^{1/2}} \right) \\ &= x_{rs} v_r^{1/2} x_{ri} \end{aligned}$$

em que $x_i = \frac{y_i - m_i \hat{\beta}_i}{v_i^{1/2}}$. Assim, tem-se que a matriz Δ pode ser escrita matricialmente por

$$\begin{aligned} \Delta_{rs} &= \mathbf{X}^t \text{diag} \left\{ v_i^{1/2} \right\} \text{diag} \{ x_i \} \\ &= \mathbf{W}^t \mathbf{D} \{ x \} \end{aligned} \quad (3.60)$$

em que $\mathbf{W} = \mathbf{D} \{ v \}^{1/2} \mathbf{X}$, $\mathbf{D} \{ v \}^{1/2} = \text{diag} \{ v_i^{1/2} \}$ e $\mathbf{D} \{ x \} = \text{diag} \{ x_i \}$, $i = 1, \dots, n$.

Derivando o resultado da equação (3.59) em relação β_r , tem-se

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_s \partial \beta_r} &= \sum_{i=1}^n \omega_i m_i x_{is} \left[\frac{e^{n_i} (1 + e^{n_i}) - e^{n_i} e^{n_i}}{(1 + e^{n_i})^2} \right] x_{ir} \\ &= \sum_{i=1}^n \omega_i x_{is} x_{ir} m_i \left(\frac{e^{n_i}}{1 + e^{n_i}} \right) \left(\frac{1}{1 + e^{n_i}} \right) \\ &= \omega_i x_{is} x_{ir} m_i p_i (1 - p_i). \end{aligned}$$

Avaliando-se a derivada acima em $\hat{\theta} = (\hat{\beta}^t, \hat{p}_i)^t$ e ω_0 , obtém-se o elemento (r, s) da matriz $\bar{\mathbf{L}}$, ou seja,

$$\begin{aligned} \bar{L}_{rs} &= \left. \frac{\partial^2 l}{\partial \beta_s \partial \beta_r} \right|_{\theta = \hat{\theta}, \omega = \omega_0} \\ &= \sum_{i=1}^n x_{is} v_i x_{ir} \\ &= \sum_{i=1}^n x_{is} v_i^{1/2} v_i^{1/2} x_{ir}. \end{aligned}$$

Dessa forma, a matriz \bar{L} pode ser escrita como

$$\begin{aligned}\bar{L} &= X^t \text{d}lag \{v_i^{1/2}\} \text{d}lag \{v_i^{1/2}\} X \\ &= W^t W.\end{aligned}\quad (3.61)$$

Substituindo-se os resultados das expressões (3.60) e (3.61) na equação (3.19) tem-se

$$\begin{aligned}\bar{F} &= D(x)W(x)W[W^t W]^{-1}W^t D(x) \\ &= D(x)P_W D(x),\end{aligned}$$

em que $P_W = W[W^t W]^{-1}W^t$, ou seja, é a projeção ortogonal do espaço \mathbf{R}^n sobre as colunas da matriz W . Segue que, substituindo-se o resultado da matriz \bar{F} acima na equação (3.13), obtém-se a curvatura normal para avaliar a perturbação de observações individuais em regressão logística

$$C_i = 2|F^t D(x)P_W D(x)F|. \quad (3.62)$$

2.5 Avaliando a influência local através da inclinação máxima de uma *lifted line*

A abordagem de Cook para diagnóstico de influência local apresenta certas dificuldades práticas, em algumas situações. Billor e Loynes (1993) destacam quatro dessas dificuldades, listadas a seguir:

- (i) a escolha de um diferencial. A escolha de um referencial para comparar com o valor da curvatura máxima obtida não é bem clara. O valor 2 foi sugerido por Cook (1986) como um referencial, ou seja, se a curvatura exceder 2, indica sensibilidade local, porém, esse valor é independente não apenas dos dados e do modelo como também das dimensões n , p e q . Portanto, não parece convincente a escolha desse valor como referencial. Em parte, esse problema acontece pela natureza do valor de C_{\max} , que, na verdade, é

uma derivada. Logo, indica a razão de mudança provocada pelo esquema de perturbação adotado, embora o interessante fosse uma medida de mudança absoluta;

- (ii) o cálculo da curvatura máxima (C_{\max}). Em situações relativamente simples, não há uma fórmula explícita disponível para a curvatura máxima, por exemplo, no caso de regressão normal linear;
- (iii) a falta de invariância da curvatura sob reparametrização do esquema de perturbação. Schall e Dunne (1992) propõem uma solução para o problema de falta de invariância da curvatura máxima, sob reparametrização do esquema de perturbação, usando a relação entre o fator de inflação da variância e a curvatura máxima;
- (iv) a falta de definição dos parâmetros. Billor e Loynes (1993) mostram, por meio de um exemplo prático, uma dificuldade decorrente da falta de definição dos parâmetros.

Considere o modelo de regressão normal linear

$$y_t = x_t^T \beta + \sigma_t \varepsilon_t, t = 1, \dots, n$$

e as duas versões de um esquema de perturbação a seguir:

a) Versão 1:

$$\sigma_t = \begin{cases} \sigma(1 + \omega), & t = 1 \\ \sigma, & t \geq 2 \end{cases} \quad (4.1)$$

b) Versão 2:

$$\sigma_t = \begin{cases} \sigma, & t = 1 \\ \frac{\sigma}{1 - \omega}, & t \geq 2 \end{cases} \quad (4.2)$$

Quando σ não é conhecido, os modelos (4.1) e (4.2) são idênticos. Contudo, Billor e Loynes (1993) constataram a ocorrência de certa diferença no cálculo da curvatura máxima para as duas versões apresentadas.

2.6 Medida de afastamento pela verossimilhança modificada

A medida apresentada por Tsai (1986 apud BILLOR; LOYNES, 1993) para medir a influência local, que definiremos mais adiante, tem, de imediato, a vantagem de estar de acordo com a falta de definição dos parâmetros, como discutido no início deste capítulo.

Considere a medida afastamento pela verossimilhança modificada, definida por

$$LD^*(\omega) = -2[l(\hat{\theta}) - (\hat{\theta}_\omega|\omega)], \quad (4.3)$$

em que $\hat{\theta}$ e $\hat{\theta}_\omega$ são os estimadores de máxima verossimilhança do vetor de parâmetros desconhecidos θ , sob os modelos original e perturbado, respectivamente.

Note que, com a medida LD^* , o problema da escolha dos parâmetros desaparece, visto que $l(\hat{\theta}_\omega|\omega)$ é a máxima verossimilhança sob o modelo perturbado, a qual independe de uma particular escolha de parametrização.

Em relação à deleção de observações, Billor e Loynes (1993) levantam a seguinte questão: se nosso interesse é perturbar o modelo, devemos conservar os dados inalterados, mas, no caso de deleção de observações, ocorre, na verdade, uma modificação na dimensão dos dados. Sendo assim, a omissão de um ponto dos dados parece ser mais do que uma simples perturbação, podendo ser mais bem entendida como outro tipo de operação.

Para lidar com essa dificuldade, Billor e Loynes (1993) usam o fato de que as estimativas dos parâmetros $\hat{\beta}^{(i)}$ e $\hat{\sigma}^2_{(i)}$, baseados no conjunto dos dados sem a i -ésima observação, são iguais às obtidas quando usamos o modelo *outlier mean shift* (deslocamento de média) com o conjunto dos dados completo.

Considere o modelo de regressão linear definido em (2.7), com todas as suposições usuais, exceto que a variância dos erros seja conhecida e igual a σ^2 . A influência local de Cook foi motivada pela estatística D_i , definida na equação (3.1). Para a medida afastamento pela verossimilhança LD^* , definimos a estatística D_i^* como uma análoga natural da estatística de Cook, D_i . Dessa forma, D_i^* mede a diferença entre $\hat{\beta}$ e $\hat{\beta}_i$, em que $\hat{\beta}$ é o estimador de máxima verossimilhança de β .

Se $l(\beta)$ denota a função de log-verossimilhança do modelo original e $l_{ms}(\beta, \phi)$ é a log-verossimilhança do modelo deslocamento de média, dado por

$$y = X\beta + d_i\phi + \varepsilon \quad (4.4)$$

em que d_i é um vetor $n \times 1$, com o i -ésimo elemento igual a 1 e os restantes iguais a zero, defina

$$LD_i^* = -2[l(\hat{\beta}) - l_{ms}(\hat{\beta}_{(i)}, \phi)]. \quad (4.5)$$

A função de log-verossimilhança para o modelo deslocamento de média é

$$l_{ms}(\hat{\beta}_{(i)}, \phi) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log (2\pi)^{1/2} - \frac{1}{2\sigma^2} (y_i - x_i^T \hat{\beta} - \phi)^2 - \frac{1}{2\sigma^2} \sum_{j=1}^n (y_j - x_j^T \hat{\beta})^2. \quad (4.6)$$

Derivando l_{ms} em relação ao vetor β , tem-se

$$\frac{\partial l_{ms}}{\partial \beta} = -\frac{1}{\sigma^2} (y_i - x_i^T \hat{\beta} - \phi)(-x_i) - \frac{1}{\sigma^2} \sum_{j=1}^n (y_j - x_j^T \hat{\beta})(-x_j).$$

Igualando-se a zero esta derivada, tem-se que

$$y_i x_i - x_i^T \hat{\beta}_{ms} x_i - \phi x_i = - \sum_{j=1}^n y_j x_j + \sum_{j=1}^n x_j^T \hat{\beta}_{ms} x_j \Rightarrow$$

$$x_i y_i + \sum_{j=1}^n x_j y_j - \phi x_i = (x_i x_i^T) \hat{\beta}_{ms} + \sum_{j=1}^n (x_j x_j^T) \hat{\beta}_{ms} \Rightarrow$$

$$X^t y - \phi x_i = \left[x_i x_i^t + \sum_{j \neq i} x_j x_j^t \right] \hat{\beta}_{ms} \Rightarrow$$

$$X^t y - \phi x_i = (X^t X) \hat{\beta}_{ms} \Rightarrow$$

$$\hat{\beta}_{ms} = (X^t X)^{-1} X^t y - (X^t X)^{-1} \phi x_i.$$

Considerando que $\hat{\beta} = (X^t X)^{-1} X^t y$, obtém-se o estimador de máxima verossimilhança do vetor de parâmetros β , para o modelo deslocamento de média

$$\hat{\beta}_{ms} = \hat{\beta} - \phi (X^t X)^{-1} x_i. \quad (4.7)$$

Agora vamos obter o estimador de máxima verossimilhança do parâmetro ϕ . Segue que, derivando-se a equação (4.6) em relação a ϕ , tem-se

$$\frac{\partial l_{ms}}{\partial \phi} = -\frac{1}{\sigma^2} (y_i - x_i^t \hat{\beta} - \phi) (-1).$$

Igualando-se a zero este resultado, obtém-se que

$$\hat{\phi} = y_i - x_i^t \hat{\beta}_{ms}.$$

Substituindo-se o valor de $\hat{\beta}_{ms}$ dado na equação (4.7) no resultado acima, tem-se que

$$\begin{aligned} \hat{\phi} &= y_i - x_i^t [\hat{\beta} - \hat{\phi} (X^t X)^{-1} x_i] \\ &= y_i - x_i^t \hat{\beta} + \hat{\phi} x_i^t (X^t X)^{-1} x_i \\ &= y_i - \hat{\mu}_i + \hat{\phi} h_{ii}, \end{aligned}$$

em que $\hat{\mu}_i = x_i^t \hat{\beta}$ é o valor predito para y_i da regressão de y sobre X e $h_{ii} = x_i^t (X^t X)^{-1} x_i$ é o elemento (i, i) da matriz *hat*. Pode-se, ainda, escrever

$$\begin{aligned} \hat{\phi} &= r_i + \hat{\phi} h_{ii} \\ &= \frac{r_i}{1 - h_{ii}}, \quad (4.8) \end{aligned}$$

em que $r_i = y_i - \hat{\mu}_i$ é o i -ésimo resíduo ordinário da regressão de \mathbf{Y} sobre \mathbf{X} .

Observe que se pode escrever o estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\beta}$, sem a i -ésima observação no modelo, como

$$\tilde{\boldsymbol{\beta}}_{(i)} = (\mathbf{X}_{(i)}^t \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^t \mathbf{Y}_{(i)}. \quad (4.9)$$

em que $\mathbf{X}_{(i)}$ é a matriz de variáveis explicativas sem a i -ésima linha e $\mathbf{Y}_{(i)}$ é o vetor de variável resposta sem o i -ésimo elemento. Segue que

$$\mathbf{X}_{(i)}^t \mathbf{X}_{(i)} = \mathbf{X}^t \mathbf{X} - \mathbf{x}_i \mathbf{x}_i^t.$$

Aplicando-se o resultado do teorema (A.2.4), descrito no Apêndice (A), ao segundo membro da equação acima, tem-se que (COOK; WEISBERG, 1982, p. 210)

$$(\mathbf{X}_{(i)}^t \mathbf{X}_{(i)})^{-1} = (\mathbf{X}^t \mathbf{X})^{-1} + \frac{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1}}{1 - \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i}.$$

Substituindo-se o resultado acima na equação (4.9), tem-se

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_{(i)} &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(i)}^t \mathbf{Y}_{(i)} + \frac{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1}}{1 - \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i} \mathbf{X}_{(i)}^t \mathbf{Y}_{(i)} \\ &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(i)}^t \mathbf{Y}_{(i)} + \frac{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(i)}^t \mathbf{Y}_{(i)}}{1 - \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i} + (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i y_i - (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i y_i \\ &= (\mathbf{X}^t \mathbf{X})^{-1} \left[\mathbf{X}_{(i)}^t \mathbf{Y}_{(i)} + \mathbf{x}_i y_i \right] + (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \left[\frac{\mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(i)}^t \mathbf{Y}_{(i)}}{1 - \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i} - y_i \right]. \end{aligned}$$

Note que $\mathbf{X}^t \mathbf{y} = \mathbf{X}_{(i)}^t \mathbf{Y}_{(i)} + \mathbf{x}_i y_i$. Logo,

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_{(i)} &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} + (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \left[\frac{\mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(i)}^t \mathbf{Y}_{(i)} - y_i + \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i y_i}{1 - h_{ii}} \right] \\ &= \tilde{\boldsymbol{\beta}} + \frac{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \left[\mathbf{X}_{(i)}^t \mathbf{Y}_{(i)} + \mathbf{x}_i y_i \right] - (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i y_i}{1 - h_{ii}} \end{aligned}$$

$$\begin{aligned}
&= \widehat{\beta} + \frac{(X^T X)^{-1} x_i x_i^T (X^T X)^{-1} X^T y - (X^T X)^{-1} x_i y_i}{1 - h_{ii}} \\
&= \widehat{\beta} + \frac{(X^T X)^{-1} x_i x_i^T \widehat{\beta} - (X^T X)^{-1} x_i y_i}{1 - h_{ii}} \\
&= \widehat{\beta} + \frac{(X^T X)^{-1} x_i [\widehat{\mu}_i - y_i]}{1 - h_{ii}} \\
&= \widehat{\beta} - \frac{r_i}{1 - h_{ii}} (X^T X)^{-1} x_i. \quad (4.10)
\end{aligned}$$

Vamos considerar, agora, apenas a parte relevante da log-verossimilhança do modelo original. Assim,

$$l(\beta) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i^T \beta)^2.$$

Segue daí que

$$\begin{aligned}
l(\beta) &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - x_i^T \beta)^2 \\
&= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 \\
&= -\frac{1}{2\sigma^2} \sum_{i=1}^n r_i^2. \quad (4.11)
\end{aligned}$$

Do mesmo modo, considere-se, agora, apenas a parte relevante da log-verossimilhança do modelo deslocamento de média em (4.6). Tem-se que

$$l_{ms}(\widehat{\beta}_{(0)}, \phi) = -\frac{1}{2\sigma^2} (y_i - x_i^T \beta - \phi)^2 - \frac{1}{2\sigma^2} \sum_{j \neq i} (y_j - x_j^T \widehat{\beta}_{(0)})^2.$$

Substituindo-se os resultados em (4.8) e (4.10) na equação acima, tem-se

$$l_{ms}(\widehat{\beta}_{(0)}, \phi) = -\frac{1}{2\sigma^2} \left\{ y_i - x_i^T \left[\widehat{\beta} - \frac{r_i}{1 - h_{ii}} (X^T X)^{-1} x_i \right] - \frac{r_i}{1 - h_{ii}} \right\}^2 - \frac{1}{2\sigma^2} \sum_{j \neq i} \left\{ y_j - x_j^T \left[\widehat{\beta} - \frac{r_i}{1 - h_{ii}} (X^T X)^{-1} x_i \right] \right\}^2$$

$$\begin{aligned}
&= -\frac{1}{2\sigma^2} \left\{ y_i - \beta_i + \frac{r_i}{1-h_{ii}} x_i^T (X^T X)^{-1} x_i - \frac{r_i}{1-h_{ii}} \right\}^2 - \frac{1}{2\sigma^2} \sum_{j \neq i} \left\{ y_j - \beta_j + \frac{r_i}{1-h_{ii}} x_j^T (X^T X)^{-1} x_i \right\}^2 \\
&= -\frac{1}{2\sigma^2} \left\{ r_i + \frac{r_i}{1-h_{ii}} h_{ii} - \frac{r_i}{1-h_{ii}} \right\}^2 - \frac{1}{2\sigma^2} \sum_{j \neq i} \left\{ r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right\}^2
\end{aligned}$$

em que $h_{ij} = x_j^T (X^T X)^{-1} x_i$ é o elemento (i, j) da matriz *hat*. Segue que

$$l_{ms}(\hat{\beta}_{(i)}, \phi) = -\frac{1}{2\sigma^2} (r_i - r_i)^2 - \frac{1}{2\sigma^2} \sum_{j \neq i} \left\{ r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right\}^2.$$

Somando e subtraindo a parcela $\frac{1}{2\sigma^2} \left\{ r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right\}^2$ no resultado

acima tem-se que

$$\begin{aligned}
l_{ms}(\hat{\beta}_{(i)}, \phi) &= -\frac{1}{2\sigma^2} \left\{ \sum_{j \neq i} \left\{ r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right\}^2 + \left\{ r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right\}^2 \right\} - \left(r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right)^2 \\
&= -\frac{1}{2\sigma^2} \sum_{j \neq i} \left(r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right)^2 + \frac{1}{2\sigma^2} \left(r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right)^2. \quad (4.12)
\end{aligned}$$

Substituindo em (4.11) e (4.12) na equação (4.5), segue que

$$\begin{aligned}
LD_{\hat{\beta}} &= -2 \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n r_i^2 - \left[-\frac{1}{2\sigma^2} \sum_{j=1}^n \left(r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right)^2 + \frac{1}{2\sigma^2} \left(r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right)^2 \right] \right\} \\
&= -2 \left(-\frac{1}{2\sigma^2} \right) \left\{ \sum_{i=1}^n r_i^2 - \sum_{j=1}^n \left(r_j^2 + 2r_j \frac{r_i}{1-h_{ii}} h_{ij} + \frac{r_i^2}{(1-h_{ii})^2} h_{ij}^2 \right) + \frac{r_i^2}{(1-h_{ii})^2} \right\} \\
&= \frac{1}{\sigma^2} \left\{ -2 \frac{r_i}{1-h_{ii}} \sum_{j=1}^n h_{ij} r_j + \frac{r_i^2}{(1-h_{ii})^2} \sum_{j=1}^n h_{ij}^2 + \frac{r_i^2}{(1-h_{ii})^2} \right\}. \quad (4.13)
\end{aligned}$$

Observe que

$$\begin{aligned}
Hr &= X(X^T X)^{-1} X^T (y - X\beta) \\
&= X(X^T X)^{-1} X^T y - X(X^T X)^{-1} X^T X\beta \\
&= X\beta - X\beta
\end{aligned}$$

$= \mathbf{0}$,

Logo, todos os elementos do produto \mathbf{Hr} são iguais a zero. Ou seja,

$$\sum_{j=1}^n h_{ij} r_j = 0, \forall i. \quad (4.14)$$

Além do mais, como \mathbf{H} é uma matriz simétrica e idempotente, tem-se que

$$\begin{aligned} \sum_{j=1}^n h_{ij}^2 &= \sum_{j=1}^n h_{ij} h_{ij} \\ &= \sum_{j=1}^n h_{ij} h_{ji} \\ &= h_{ii}. \end{aligned} \quad (4.15)$$

Substituindo-se os resultados em (4.14) e (4.15) na equação (4.13), tem-se

$$\begin{aligned} LD_i^* &= \frac{1}{\sigma^2} \left(0 - \frac{r_i^2}{(1-h_{ii})^2} h_{ii} + \frac{r_i^2}{(1-h_{ii})^2} \right) \\ &= \frac{1}{\sigma^2} \left\{ \frac{r_i^2(1-h_{ii})}{(1-h_{ii})^2} \right\} \\ &= \frac{r_i^2}{\sigma^2(1-h_{ii})}. \end{aligned} \quad (4.16)$$

Note que a medida acima depende apenas de duas quantidades, os resíduos r_i e os elementos da diagonal principal da matriz hat .

No capítulo anterior viu-se que uma motivação para a influência local de Cook (1986) é a relação, em (3.6), entre a estatística D_i e a medida de afastamento pela verossimilhança LD . Billor e Loynes (1993) consideram uma relação análoga, entre a estatística D_i^* e a medida afastamento pela verossimilhança modificada LD^* , dada por

$$pD_i^*(\omega) = -2[l(\hat{\beta}) - l(\hat{\beta}_\omega | \omega)]. \quad (4.17)$$

Por meio de um raciocínio semelhante ao utilizado por Cook (1986) para estudar a influência local, tem-se que $(\omega, LD^*(\omega))$ representa uma superfície $q + 1$ dimensional e, com isso, é possível estabelecer **linhas projetadas** (*lifted line*) em várias direções dessa superfície. Considerando a reta em \mathbf{R}^q definida por $\omega(\alpha) = \omega_0 + \alpha \ell$, em que ω_0 representa um vetor de não perturbação do modelo, ℓ é um vetor fixo em \mathbf{R}^q e α é um escalar, segue que o comportamento de $LD^*(\omega_0 + \alpha \ell)$ proporciona informações importantes sobre a influência do esquema de perturbação adotado.

Como a primeira derivada de LD^* avaliada em ω_0 não se anula, exceto para alguns valores particulares do vetor ℓ , a primeira derivada de LD^* em relação a ℓ produz informações valiosas sobre o comportamento local de $LD^*(\omega_0)$. Em particular, a inclinação máxima e a correspondente direção máxima, denotadas, respectivamente, por $d_{1\max}^{\uparrow*}$ e $\ell_{1\max}^{\uparrow*}$, são importantes quando $q \neq 1$, em que q é a dimensão do vetor ω (BILLOR; LOYNES, 1993).

A direção máxima da função LD^* no ponto ω_0 ocorre na direção do gradiente $\nabla LD^*(\omega_0)$. Logo, a inclinação máxima $(d_{\max}^{\uparrow*})$ é a medida por (LIMA, 1981)

$$d_{1\max}^{\uparrow*} = \|\nabla LD^*(\omega_0)\|.$$

Pela regra da cadeia,

$$\begin{aligned} \frac{\partial LD^*}{\partial \alpha} &= \frac{\partial \omega}{\partial \alpha} \frac{\partial LD^*}{\partial \omega} \\ &= \ell^T \left\{ -2 \left[\frac{\partial l(\hat{\beta})}{\partial \omega} - \left(\frac{\partial l(\beta|\omega)}{\partial \beta} \frac{\partial l\beta_\omega}{\partial \omega} + \frac{\partial l(\hat{\beta}_\omega|\omega)}{\partial \omega} \right) \right] \right\}. \end{aligned}$$

De imediato, tem-se que $\frac{\partial l(\hat{\beta})}{\partial \omega} = 0$ e que $\frac{\partial l(\beta|\omega)}{\partial \beta}$, quando avaliada em $\hat{\beta}$ e ω_0 , também se anula. Assim, a equação acima fica resumida a

$$\frac{\partial LD^*}{\partial \alpha} \Big|_{\beta=\hat{\beta}, \omega=\omega_0} = 2\ell^T \frac{\partial l(\beta|\omega)}{\partial \omega} \Big|_{\beta=\hat{\beta}, \omega=\omega_0}.$$

Portanto, a inclinação máxima pode ser escrita como

$$d_{l, \max}^{\uparrow *} = 2 \|\nabla l(\hat{\beta}|\omega)\|. \quad (4.18)$$

Para uma aplicação dessa teoria, vamos considerar o modelo perturbado de regressão normal linear dado a seguir

$$y = X\beta + \varepsilon$$

em que $\text{Var}(\varepsilon) = \sigma^2 W^{-1}$ e $W = \text{diag}(1 + \omega_{1,1,1}, \dots, 1)$, com o modelo não perturbado sendo obtido quando $\omega_1 = 0$. A inclinação da linha projetada sobre a superfície $(\omega, LD^*(\omega_0))$ é calculada por meio da primeira derivada de $LD^*(\omega)$ em relação a ω_1 . A log-verossimilhança para o modelo acima fica estabelecida por

$$l(\beta, \omega) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi + \frac{n}{2} \ln(1 + \omega_1) - \frac{(1 + \omega_1)(y_1 - x_1^T \beta)^2}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{i=2}^n (y_i - x_i^T \beta)^2.$$

segue daí que

$$\begin{aligned} \frac{\partial LD^*}{\partial \omega_1} &= 2 \frac{\partial l(\beta|\omega)}{\partial \omega_1} \\ &= 2 \left[\frac{1}{2(1 + \omega_1)} - \frac{(y_1 - x_1^T \beta)^2}{2\sigma^2} \right]. \end{aligned}$$

Assim,

$$\begin{aligned} \frac{\partial LD^*}{\partial \omega_1} \Big|_{\beta=\hat{\beta}, \omega_1=0} &= 1 - \frac{(y_1 - x_1^T \hat{\beta})^2}{2\sigma^2} \\ &= 1 - \frac{r_1^2}{\sigma^2}, \quad (4.19) \end{aligned}$$

em que $r_1 = y_1 - x_1^T \hat{\beta}$ é o resultado ordinário relativo à primeira observação da regressão de \mathbf{y} sobre \mathbf{X} . Têm-se as seguintes considerações para o resultado do resíduo r_1 acima:

1. se $|r_1| = \sigma$, não existe razão para preferir o modelo perturbado ao não perturbado e vice-versa,
2. se $|r_1| \ll \sigma$ deve-se optar pelo modelo perturbado caso $\omega_1 < 0$, isto é, se $\text{Var}(y_1) < \sigma^2$.

Observa-se, ainda, que $1 - \frac{r_1^2}{\sigma^2}$ tem uma cota superior finita quando $|r_1| \rightarrow 0$, enquanto sua cota é inferior quando $|r_1| \rightarrow +\infty$ e $-\infty$. Essa assimetria está inteiramente de acordo com a teoria vista (BILLOR; LOYNES, 1993).

Agora, considere-se a matriz de perturbação geral dada por

$$W_g = \text{diag}(1 + \omega_{1,1}, \omega_{2,1}, \dots, 1 + \omega_{n,1}),$$

para o mesmo modelo anterior. Sabe-se que a taxa de variação da função LD^* é medida pelo gradiente

$$\nabla LD^*(\omega_0) = \left(\frac{\partial LD^*}{\partial \omega_1}, \frac{\partial LD^*}{\partial \omega_2}, \dots, \frac{\partial LD^*}{\partial \omega_n} \right).$$

Utilizando-se um procedimento análogo ao realizado para obter o resultado em (4.19), tem-se que

$$\left. \frac{\partial LD^*(\omega)}{\partial \omega_i} \right|_{\omega=\omega_0} = 1 - \frac{r_i^2}{\sigma^2}$$

quando $\omega_i = 0$ com $i = 1, \dots, n$. Dessa forma, a razão máxima de crescimento de LD^* é na direção do vetor gradiente $\nabla LD^*(\omega_0)$, ou seja,

$$d_{LD^*}^{\max} \hat{*} = \|\nabla LD^*(\omega_0)\|$$

$$\begin{aligned}
&= \sqrt{\left(\frac{\partial LD^*}{\partial \omega_1}\right)^2 + \left(\frac{\partial LD^*}{\partial \omega_2}\right)^2 + \dots + \left(\frac{\partial LD^*}{\partial \omega_n}\right)^2} \\
&= \sum_{j=1}^n \left(1 - \frac{r_j^2}{\sigma^2}\right)^2. \quad (4.20)
\end{aligned}$$

O mesmo desenvolvimento adotado nesta seção pode ser aplicado quando estamos interessados em avaliar o comportamento de um gráfico de influência em um ponto ω^* diferente de ω_0 , visto que, para isso, faz-se necessária apenas a primeira derivada da função LD^* . A única diferença existente é que, neste caso, a razão máxima de crescimento deve ser avaliada agora no ponto ω^* .

2.6.1 Avaliando a inclinação máxima $d_{1\max}^{\dagger *}$

Uma das dificuldades na interpretação dos resultados em um estudo da influência local, na abordagem de Cook (1986), é justamente o de determinar um valor de referência para comparar com a curvatura máxima obtida. Segundo Billor e Loynes (1993), essa questão deve ser respondida pelo julgamento e a intuição em escolher um valor apropriado para servir de referência, e que seja adequado ao problema.

Billor e Loynes (1993) esclarecem que existem várias maneiras para determinar tal referência. Por exemplo, considere-se o modelo de regressão normal linear perturbado pela matriz $W_g = \text{diag}(1 + \omega_{1,1} + \omega_2, \dots, 1 + \omega_n)$, já mencionado anteriormente. Já foi demonstrado que, para esse problema, tem-se

$$d_{1\max}^{\dagger *} = A^{\dagger}(1/2),$$

em que $A = \sum_{j=1}^n \left(1 - \frac{r_j^2}{\sigma^2}\right)^2$. Esta expressão é útil para avaliar os momentos de A , porém, para a obtenção de um valor de referência, ela se torna um tanto

complicada. Segue daí que, substituindo-se os resíduos r_j em A pelos respectivos erros aleatórios ε_j , os quais são assumidos variáveis aleatórias independentes normalmente distribuídas, com média zero e variância σ^2 , pode-se obter a esperança e a variância de A , como segue. Tem-se que

$$A = \sum_{j=1}^n \left[\left(\frac{\varepsilon_j}{\sigma} \right)^2 - 1 \right]^2.$$

Note que

$$\varepsilon_j \sim N(0, \sigma^2) \Rightarrow \left(\frac{\varepsilon_j}{\sigma} \right)^2 \sim X_{(1)}^2, \forall j$$

em que $X_{(1)}^2$ representa uma distribuição qui-quadrado com um grau de liberdade.

Como $E(X_{(1)}^2) = 1$, segue que

$$\begin{aligned} E(A) &= E \left\{ \sum_{j=1}^n \left[\left(\frac{\varepsilon_j}{\sigma} \right)^2 - E \left(\frac{\varepsilon_j}{\sigma} \right)^2 \right]^2 \right\} \\ &= \sum_{j=1}^n E \left[\left(\frac{\varepsilon_j}{\sigma} \right)^2 - E \left(\frac{\varepsilon_j}{\sigma} \right)^2 \right]^2 \\ &= \sum_{j=1}^n \text{Var} \left(\frac{\varepsilon_j}{\sigma} \right)^2 \\ &= 2n, \end{aligned} \quad (4.21)$$

uma vez que $\text{Var}(X_{(1)}^2) = 2$.

Pela proposição (A.3.3) do Apêndice (A), tem-se

$$E[A^2] = 56n + 4n^2 \quad (4.22).$$

Substituindo-se os resultados em (4.21) e (4.22) na fórmula da variância a seguir, tem-se que

$$\begin{aligned} \text{Var}[A] &= E[A^2] - (E[A])^2 \\ &= 56n + 4n^2 - (2n)^2 \end{aligned}$$

$$= 56n \quad (4.23).$$

Dessa forma, um referencial conveniente seria, então, dado por $E[A] + 2DP[A]$, em que $DP[A]$ representa o desvio padrão de A . Segue que, para o nosso exemplo, poderia se utilizar como referencial a quantidade

$$2n + 4\sqrt{14n}.$$

Para uma melhor aproximação, pode-se usar uma transformação de A . Porém, o valor $2n + 4\sqrt{14n}$ pode ser considerado como uma referência para avaliar a influência local do problema proposto.

2.6.2 Inclinação máxima no caso de perturbação de observação no modelo normal linear

Considerando apenas a parte relevante da função de log-verossimilhança perturbada como dada em (3.35), tem-se que

$$l(\beta|\omega) = \sum_{i=1}^n \frac{1}{2\sigma^2} \omega_i (y_i - x_i^T \beta)^2, \quad (4.24)$$

em que ω_i e y_i são as i -ésimas componentes dos vetores ω e y , respectivamente.

Derivando a função (4.24) em relação a ω_i , tem-se que

$$\frac{\partial l}{\partial \omega_i} = \frac{1}{2\sigma^2} (y_i - x_i^T \beta)^2.$$

Aplicando-se a derivada acima em $\hat{\beta}$ e $\hat{\sigma}^2$, tem-se que

$$\begin{aligned} \left. \frac{\partial l}{\partial \omega_i} \right|_{\beta=\hat{\beta}, \sigma^2=\hat{\sigma}^2} &= \frac{1}{2\hat{\sigma}^2} (y_i - x_i^T \hat{\beta})^2 \\ &= \frac{1}{2\hat{\sigma}^2} (y_i - \mu_i)^2 \\ &= \frac{1}{2\hat{\sigma}^2} r_i^2. \end{aligned}$$

Da seção (3.2), tem-se que

$$\begin{aligned} \frac{\partial LD^*}{\partial \omega_i} &= 2 \frac{\partial l(\beta|\omega)}{\partial \omega_i} \Big|_{\beta=\hat{\beta}, \sigma^2=\hat{\sigma}^2} \\ &= 2 \left\{ \frac{1}{2\sigma^2} r_i^2 \right\} \\ &= \frac{r_i^2}{\sigma^2}. \quad (4.25) \end{aligned}$$

Vimos, na seção (3.2), que a direção da inclinação máxima $\ell_{i \max}^{\uparrow *}$ é a direção do gradiente. Assim, pode-se escrever

$$\ell_{i \max}^{\uparrow *} = \nabla [LD]^{\uparrow *} (\omega_i \mathbf{0}) = ((r_1^2)/\sigma^2, \dots, (r_n^2)/\sigma^2) \quad (4.26).$$

Segue que a inclinação máxima é obtida por

$$\begin{aligned} d_{i \max}^{\uparrow *} &= \|\nabla [LD]^{\uparrow *} (\omega_i \mathbf{0})\| \\ &= \sqrt{\sum_{j=1}^n \left(\frac{r_j^2}{\sigma^2}\right)^2} \quad (4.27). \end{aligned}$$

Para obter um valor de referência para a inclinação máxima acima, pode-se

desenvolver um raciocínio análogo ao realizado na seção (3.3). Seja $A = \sum_{j=1}^n \frac{r_j^2}{\sigma^2}$, substituindo em A os r_j^2 pelos erros ε_j^2 que são assumidos variáveis aleatórias independentes normalmente distribuídas com média zero e variância σ^2 , segue que

$$A = \sum_{j=1}^n \frac{\varepsilon_j^2}{\sigma^2} = \sum_{j=1}^n \left(\frac{\varepsilon_j}{\sigma}\right)^2.$$

Tem-se que $\left(\frac{\varepsilon_j}{\sigma}\right)^2 \sim \chi_1^2$ e $E[X_1^2] = 1$. Fazendo $x_j = \left(\frac{\varepsilon_j}{\sigma}\right)^2$, segue que

$$E[A] = \sum_{j=1}^n E[x_j]$$

$$= \sum_{j=1}^n \mathbf{1} \\ = n \quad (4.28).$$

Observe que

$$A^2 = \sum_{j=1}^n x_j^2 + \sum_{j \neq 1} x_i x_j.$$

Logo,

$$E[A^2] = \sum_{j=1}^n E[x_j^2] + \sum_{j \neq 1} E[x_i x_j],$$

pela independência entre x_i e x_j , para todo $i \neq j$, tem-se que

$$E[A^2] = \sum_{j=1}^n E[x_j^2] + \sum_{j \neq 1} E[x_i]E[x_j]. \quad (4.29)$$

Pela proposição (A.3.3) no Apêndice (A), tem-se que a função geradora de momentos de x_j é dada por

$$M_{x_j}^*(t) = 3(1 - 2t)^{-3/2} \Rightarrow M_{x_j}^*(0) = E[x_j^2] = 3. \quad (4.30)$$

Substituindo-se este resultado em (4.29), obtém-se

$$E[A^2] = \sum_{j=1}^n 3 + \sum_{j=1}^n (n-1) \\ = 3n + n(n-1) \\ = n^2 + 2n \quad (4.31).$$

Substituindo-se o resultado das equações (4.28) e (4.31) na fórmula da variância a seguir, obtém-se

$$\text{Var}[A] = E[A^2] - (E[A])^2 \\ = n^2 + 2n - (n^2) \\ = 2n.$$

Portanto, pode-se considerar, como um valor de referência para a inclinação máxima obtida em (4.27), o resultado de

$$E[A] + 2\sqrt{\text{Var}[A]} = n + 2\sqrt{2n}. \quad (4.32)$$

2.6.3 Inclinação máxima no caso de perturbação de observação no modelo logístico

Considere a função de log-verossimilhança perturbada para uma amostra aleatória de uma variável aleatória com distribuição Binomial (m_i, p_i) , como dado na subseção (3.6.1)

$$l(\beta, y|\omega) = \sum_{i=1}^n \omega_i y_i \ln \left[\frac{p_i}{1-p_i} \right] + \sum_{i=1}^n \omega_i m_i \ln[1-p_i].$$

Derivando a função em relação a ω_i , obtém-se

$$\frac{\partial l}{\partial \omega_i} = y_i \ln \left[\frac{p_i}{1-p_i} \right] + m_i \ln[1-p_i].$$

Avaliando-se a expressão acima em \hat{p}_i , tem-se

$$\left. \frac{\partial l}{\partial \omega_i} \right|_{p_i=\hat{p}_i} = y_i \ln \left[\frac{\hat{p}_i}{1-\hat{p}_i} \right] + m_i \ln[1-\hat{p}_i].$$

Dessa forma, as coordenadas do vetor gradiente $\ell_{\mathbf{1}, \max}^{\uparrow *} = \nabla [LD]^{\uparrow *} (\omega_{\mathbf{1}} \mathbf{0})$ são dadas por

$$\begin{aligned} \ell_i^* &= \frac{\partial LD^*}{\partial \omega_i} \\ &= 2 \left. \frac{\partial l(\beta|\omega)}{\partial \omega_i} \right|_{\beta=\hat{\beta}, \sigma^2=\hat{\sigma}^2} \\ &= 2 \left\{ y_i \ln \left[\frac{\hat{p}_i}{1-\hat{p}_i} \right] + m_i \ln[1-\hat{p}_i] \right\} \quad (4.33) \end{aligned}$$

A inclinação máxima é obtida pela aplicação imediata de

$$d_{\mathbf{1}, \max}^{\uparrow *} = \|\nabla [LD]^{\uparrow *} (\omega_{\mathbf{1}} \mathbf{0})\|.$$

3 RESULTADOS E DISCUSSÕES

3.1 Aplicações

Neste capítulo discutem-se dois exemplos de análise de dados reais via pacote estatístico **R**. O objetivo principal destas aplicações é colocar em prática as técnicas de diagnóstico, com ênfase na influência local de Cook (1986) e na abordagem de Billor e Loynes (1993), descritas nesta tese.

Utilizam-se também algumas **técnicas gráficas** para detectar a fuga das suposições adotadas para os modelos sugeridos e a ocorrência de possíveis *outliers*. Estes conceitos foram descritos no capítulo 1; para maiores detalhes consultar Paula (2004).

No primeiro exemplo (seção (5.1), utilizou-se o modelo de regressão normal linear simples para os dados propostos. A análise foi desenvolvida de quatro maneiras: técnicas gráficas de diagnóstico; curvatura normal de Cook para regressão normal linear sob perturbação de observação; inclinação máxima sob perturbação na variância e inclinação máxima sob perturbação de observação.

No segundo exemplo (seções (4.2), foi utilizado o modelo de regressão logística múltipla para os dados apresentados. Analisou-se o modelo proposto sob três aspectos: técnicas gráficas de diagnóstico, curvatura normal de Cook sob esquema de perturbação de observação e a inclinação máxima também sob esquema de perturbação de observação.

3.2 Estimativa do número de gansos por bando

Métodos de inspeção aérea são regularmente empregados para estimar a população de gansos, durante o verão, no oeste da baía de Hudson, no Canadá

(WEISBERG, 1985). Para isso, uma pequena aeronave sobrevoa a região e, quando avista um bando de gansos, uma pessoa experiente estima a quantidade dessas aves. Para investigar a eficiência desse método, um experimento foi conduzido, no qual uma aeronave carregando dois observadores sobrevoou 45 bandos e cada um deles estimou, independentemente, o número de pássaros por bando. Além disso, uma fotografia foi retirada de cada bando e, posteriormente, uma contagem do número de gansos foi realizada por meio dessa fotografia.

Os dados coletados neste experimento encontram-se em Weisberg (1985). Aplicou-se um modelo normal linear simples para os dados, de modo que Y seja o número de gansos por bando, computados a partir da fotografia aérea e X , o número de gansos, estimado pelo observador I do experimento. O modelo ficou, portanto, dado por

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, 45 \quad (5.1)$$

com a suposição de que $\varepsilon_i \sim N(0, \sigma^2)$, com os ε_i s, mutuamente independentes. As estimativas dos parâmetros, fornecidas pelo **R**, foram $\hat{\alpha} = 26,650$ e $\hat{\beta} = 0,883$, com respectivos erros padrões dados por 8,614 e 0,078, ambos indicando que os coeficientes apresentaram-se significativos ao modelo.

A estimativa da variância é dada por 44,41 e a porcentagem de variação de Y explicada pelo modelo, ou seja, o valor do **coeficiente de determinação**, foi de 74,45%. A regressão, por meio do teste F, mostrou-se significativa.

O **diagrama de dispersão** (plot de X versus Y) mostrado na Figura 5.1 sugere forte evidência de heterocedasticidade. Assim, pode-se esperar que a introdução de uma pequena perturbação revele alguma espécie de sensibilidade no modelo.

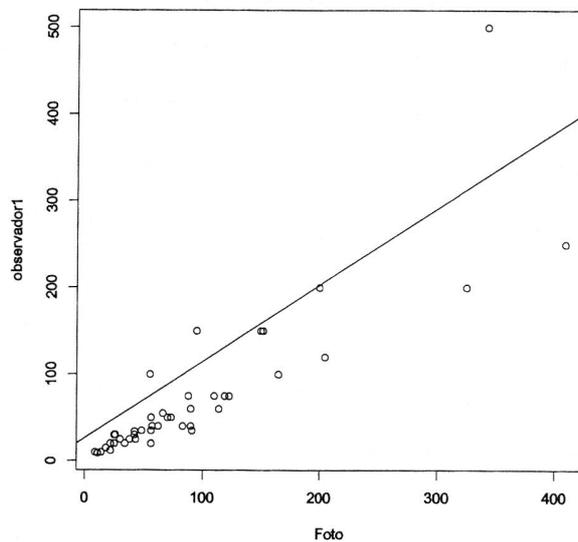


Figura 1 Número de gansos contados a partir da foto *versus* o número de gansos estimado pelo observador 1

A seguir, faz-se uma análise de alguns gráficos de diagnósticos, com o objetivo de identificar pontos influentes e verificar a validade das suposições adotadas para o modelo.

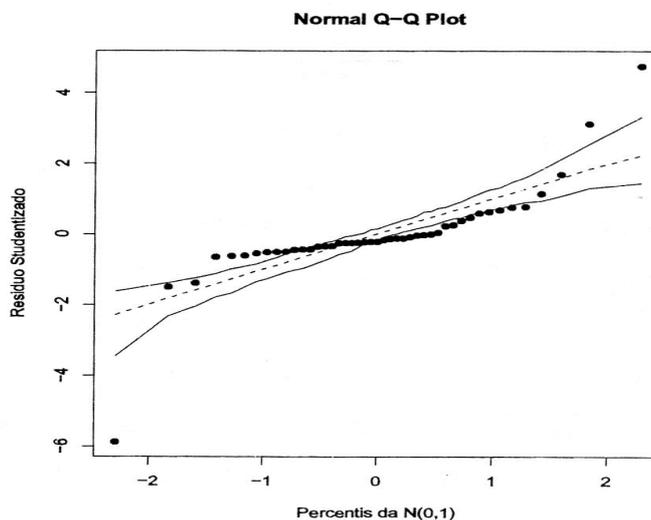


Figura 2 Gráfico de probabilidades com envelopes

Na Figura 2 tem-se o gráfico normal de probabilidade com envelopes para os resíduos t_{si} , que avalia a validade da suposição de normalidade da variável resposta, para o modelo adotado em (5.1). Pode-se notar, por meio deste gráfico, um forte indício de não validade desta suposição, pois há uma grande quantidade de pontos fora das bandas de confiança do gráfico, configurando, talvez, a ausência de algum termo extra no modelo.

Outro gráfico importante a ser interpretado é o gráfico dos valores \hat{h}_{ii} em função das ordens das observações dadas na Figura 3. Por meio dele destacamos

dois pontos acima da linha $\frac{2 * p}{n}$, que correspondem às observações 28 e 29, sendo, portanto, candidatos a pontos influentes.

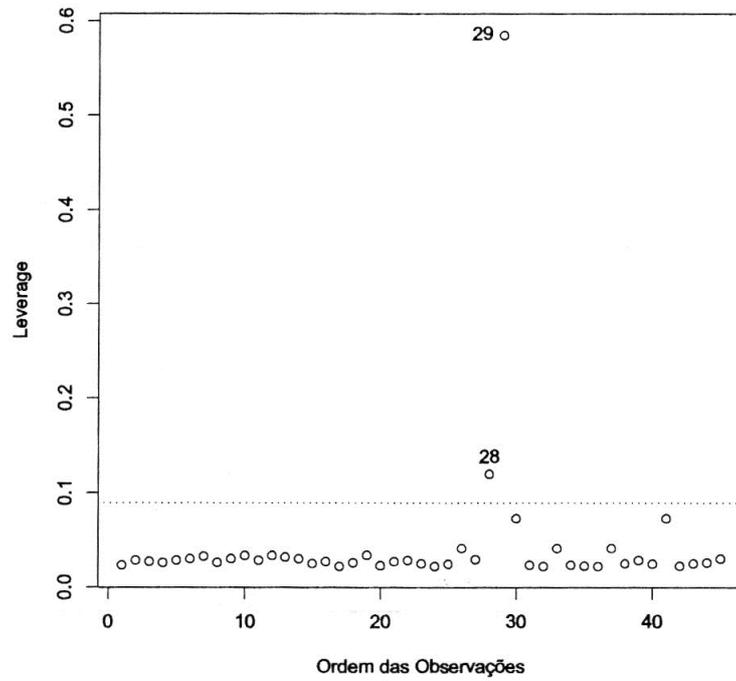


Figura 3 Leverage \hat{h}_{ii} contra a ordem das observações

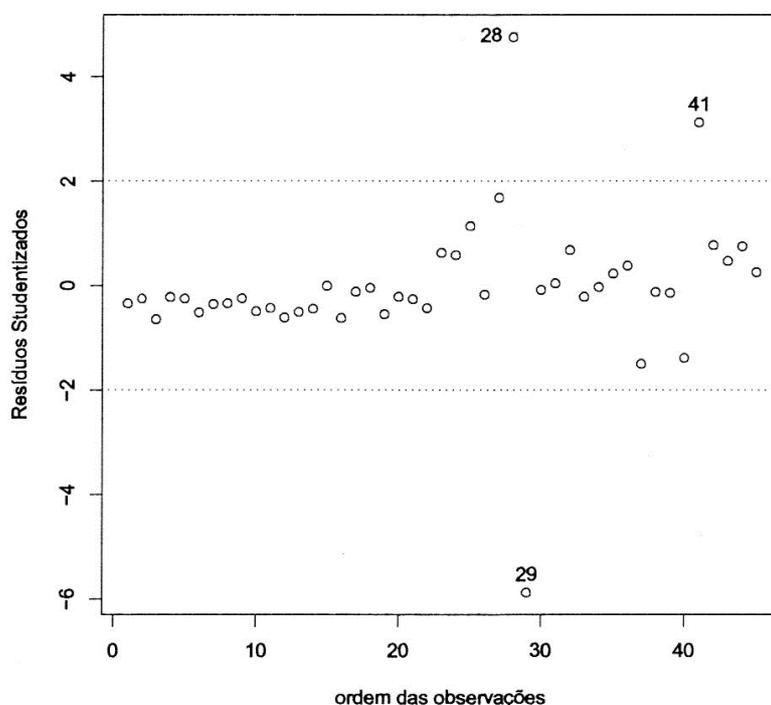


Figura 4 Resíduos studentizados (t_{Si}) contra a ordem das observações

O gráfico dos resíduos studentizados em função das ordens das observações, dado na Figura 4, destaca novamente os pontos 28 e 29, além do ponto 41, sendo estes candidatos a serem pontos aberrantes.

Na Figura 5 observa-se o gráfico dos resíduos studentizados em função dos valores ajustados, verificando-se, novamente em destaque, os pontos 28, 29 e 41. O gráfico renova a indicação de heterocedasticidade já percebida pelo diagrama de dispersão (Figura 5).

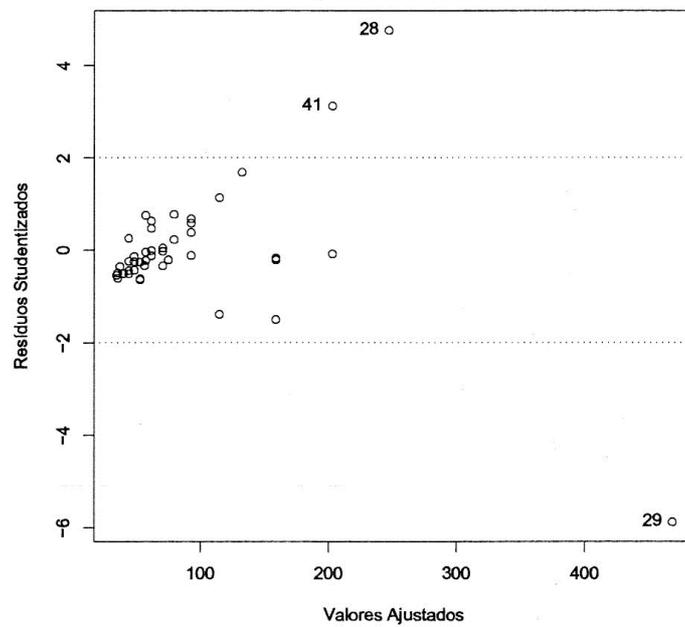


Figura 5 Resíduos studentizados t_{Si} contra os valores ajustados

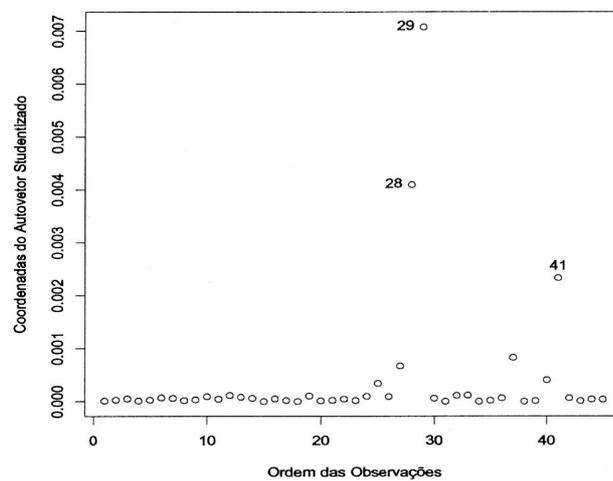


Figura 6 Coordenadas do autovetor d_{max} contra a ordem das observações

Os mesmos pontos já destacados anteriormente (28,29 e 41) também aparecem em destaque no gráfico da Figura 6, das coordenadas do autovetor d_{max} em função das ordens das observações.

Dentre os pontos destacados pelos gráficos como candidatos a *outliers*, o que aparece sempre mais remoto é o correspondente à observação 29, daí a escolha deste ponto para sofrer uma pequena perturbação. Espera-se, com isso, que o modelo seja sensível a esta alteração e o modo de medir esta sensibilidade será por meio da **curvatura normal** de Cook para modelo normal, dada em (3.45). Para introduzir a perturbação nas observações, considerou-se o vetor \mathbf{e} da referida fórmula, com 1 em todas as posições, com exceção da posição 29, que recebeu o valor 0,9. O valor obtido foi de 17.766, o que indica extrema sensibilidade local.

Utilizando o esquema de perturbação de observação descrito na seção (4.3), aplicou-se o resultado da inclinação máxima (\mathbf{e}_{max}^*) dada na equação (4.28). Com isso, obteve-se o resultado 16,573. Considerando como referência o valor sugerido pela fórmula em (4.33), que é de 7,381, tem-se, aqui, forte indicação de sensibilidade local.

Para o esquema de perturbação da matriz de variância abordado na seção (4.1), o valor da inclinação máxima é de 17.766, obtido por meio da fórmula dada em (4.20). Utilizando o valor de referência para \mathbf{e}_{max}^* , sugerido na seção (4.2), que é de 13,798, também aqui detectou-se forte sensibilidade local.

A abordagem de Billor e Loynes (1993) permite que se tenham os valores individuais das coordenadas ($\mathbf{e}_{j_s}^*$) do gradiente $\mathbf{e}_{max} = \sqrt{LD^*}(\omega_0)$, que é o vetor que indica a direção onde se obtém a inclinação máxima. Assim, pode-se avaliar quais observações contribuem com a maior parcela para o valor de \mathbf{e}_{max}^* . Os $\mathbf{e}_{j_s}^*$ individuais, obtidos pela perturbação da matriz de variância, são

apresentados na Tabela 1, na qual nota-se que as observações que contribuem com os maiores valores são as de números 28, 29 e 41.

Tabela 5 Valores dos β_{js}^* para os dados de gansos.

Observação (i)	β_{js}^*	Observação (i)	β_{js}^*	Observação (i)	β_{js}^*
1	0,8892	16	0,6268	31	0,9975
2	0,9417	17	0,9881	32	0,5387
3	0,5988	18	0,9988	33	0,9586
4	0,9539	19	0,7177	34	0,9997
5	0,9418	20	0,9572	35	0,9452
6	0,7478	21	0,9372	36	0,8507
7	0,8822	22	0,8224	37	1,0793
8	0,8911	23	0,6010	38	0,9875
9	0,9461	24	0,6530	39	0,9834
10	0,7661	25	0,2726	40	0,8199
11	0,8224	26	0,9749	41	6,5283
12	0,6445	27	1,6615	42	0,3999
13	0,7570	28	12,2618	43	0,7753
14	0,8111	29	7,0421	44	0,4322
15	0,9999	30	0,9949	45	0,9306

3.3 Preferência por um novo detergente em relação a um detergente padrão

Os dados utilizados nesta seção encontram-se em Cordeiro (1986) e referem-se a um experimento comparando a preferência por um novo detergente em relação a um detergente padrão. Considera-se que os fatores dureza, com três níveis (pesada, média e mole) e temperatura da água, com dois níveis (baixa e alta), possam influenciar na escolha de um detergente.

Adotou-se, para os dados, o modelo logístico linear múltiplo com variável resposta Y = proporção de indivíduos que preferem o novo detergente e variáveis explicativas $temp_j$, em que $temp_1$ = alta e $temp_2$ = baixa, e $dure_k$, em que $dure_1$ = média, $dure_2$ = mole e $dure_3$ = pesada, definidas como fatores. O modelo é dado por

$$\eta_i = \beta_1 + \sum_{j=1}^2 \beta_{2j} temp_j + \sum_{k=1}^3 \beta_{3k} dure_k, \quad i = 1, \dots, 6,$$

com erro binomial; é assumido que $\beta_{21} = \beta_{31} = 0$. O desvio obtido para este modelo é $D(y, p) = 0,618$ (2 graus de liberdade), indicando um ajuste adequado, pois o p -value associado a esse valor pelo teste qui-quadrado unicaudal é de 73,42%. O número de iterações para estimar os coeficientes do modelo proposto foi três. Na Tabela 2, a seguir, têm-se as estimativas dos coeficientes do modelo adotado e seus respectivos desvios padrões. Todos se apresentaram significativos ao modelo.

Tabela 2 Estimativas dos coeficientes referentes ao modelo em (5.3), para explicar a proporção de pessoas que preferem o novo detergente.

Coeficiente	Estimativas	Desvio padrão
Constante	-0,652	0,197
$temp_2$	0,457	0,194
$dure_2$	0,311	0,226
$dure_3$	-0,057	0,227

Mostraremos agora os gráficos usados para detectar possíveis indicações de transgressão de alguma suposição adotada para o modelo, bem como da

ocorrência de algum ponto influente. Pode-se verificar, por meio do gráfico de probabilidade com envelopes, na Figura 7, que não há indicação forte de que a variável resposta não siga uma binomial (m_i, p_i) .

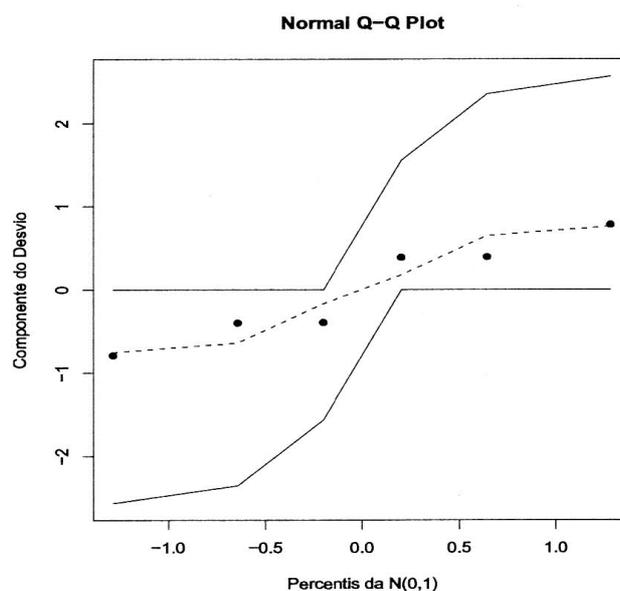


Figura 7 Gráfico de probabilidade com envelopes para o modelo binomial.

No gráfico dos \hat{h}_{ii} em função da ordem das observações, dado na Figura 8, nenhum dos pontos aparece acima da linha $\frac{2 * p}{n}$, não havendo, portanto, pontos suspeitos de serem aberrantes, contudo, os pontos mais afastados correspondem às observações 1, 3 e 5.

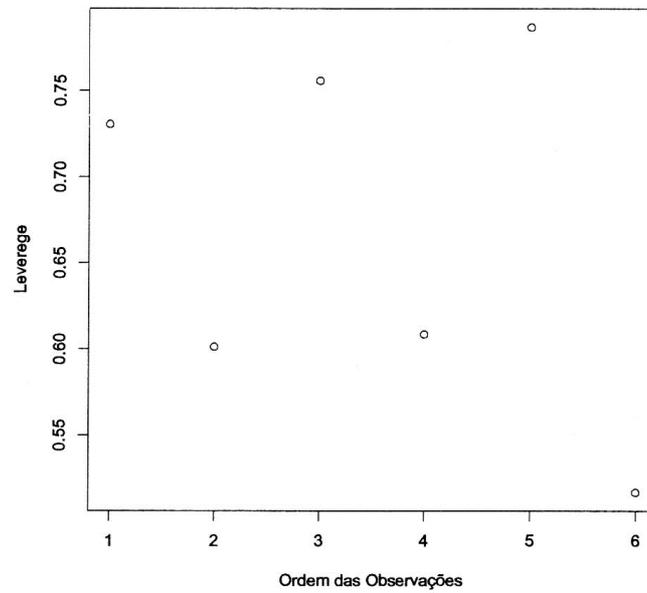


Figura 8 Leverage \hat{h}_{ii} contra a ordem das observações

O gráfico de \hat{h}_{ii} em função dos valores ajustados, mostrado na Figura 9, destaca os pontos correspondentes às observações 1, 3 e 5 como estando mais afastados.

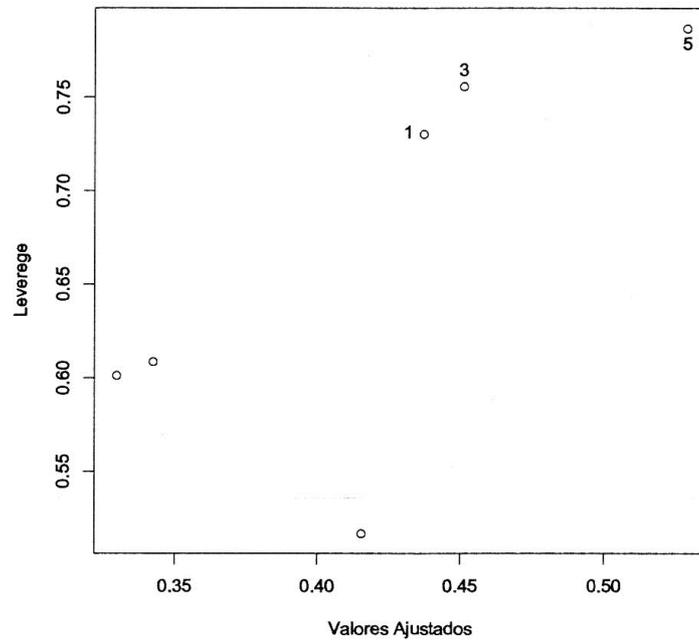


Figura 9 Leverage \hat{h}_{ii} contra valores ajustados

No gráfico do resíduo t_{Di} em função dos valores ajustados (Figura 10), os pontos que se encontram mais afastados correspondem às observações 2, 3 e 5.

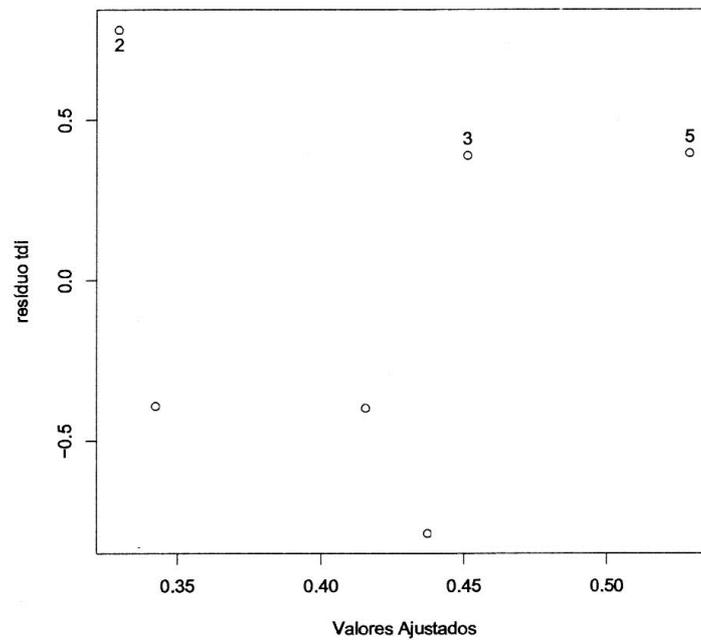


Figura 10 Resíduo t_{Di} contra valores ajustados

Na Figura 11 observa-se o gráfico da medida afastamento pela verossimilhança, LD_i , em função da ordem das observações. Nele pode-se perceber que o ponto correspondente à observação 1 parece mais afastado.

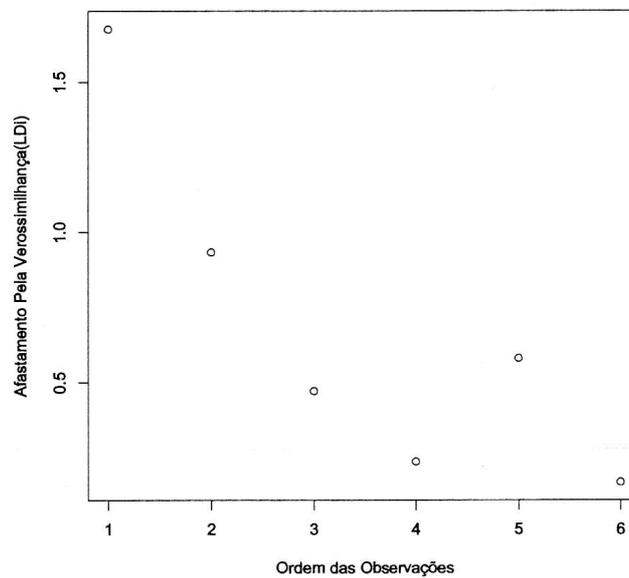


Figura 11 Gráfico de LD_i contra a ordem das observações

No gráfico das coordenadas do autovetor \mathbf{e}_{max} em função da ordem das observações, mostrado na Figura 12, pode-se observar que os pontos correspondentes às observações 1 e 2 encontram-se mais afastados, em especial a de número 2.

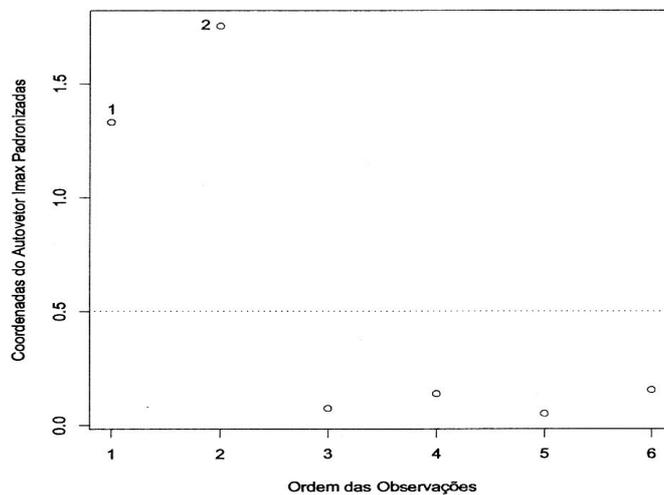


Figura 12 Coordenadas do autovetor \mathbf{e}_{max} contra a ordem das observações

Avaliando a situação dos pontos destacados pelos gráficos não se percebe forte evidência de ocorrência de *outliers*, porém, os pontos que se apresentaram mais afastados foram aqueles correspondentes às observações 1, 2, 3 e 5. Para calcular a curvatura normal, optou-se por perturbar a observação 1. Para isso, adotou-se um procedimento análogo ao realizado na aplicação da seção (5.2), utilizando o vetor de perturbação \mathbf{e} com 1, em todas as posições, menos na observação 1, que recebe o valor 0,9. O resultado obtido para a curvatura normal foi de 1.298,88. O valor da inclinação máxima obtido para o esquema de perturbação abordado na seção (5.3) foi de 171,75.

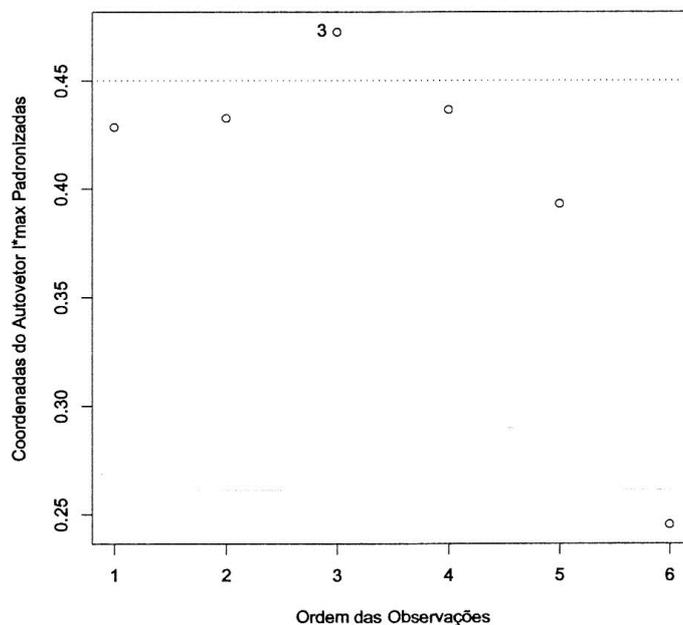


Figura 13 Coordenadas do autovetor l^*_{max} contra a ordem das observações

O gráfico mostrado na Figura 5.13, das coordenadas do vetor gradiente l^*_{max} em função da ordem das observações, destaca o ponto correspondente à observação 3.

Na Tabela estão dispostas as coordenadas do gradiente l^*_{max} e do vetor l_{max} por observação; a observação 3 apresenta a maior coordenada do vetor l^*_{max} e a observação 2 é a maior coordenada do vetor l_{max} . As técnicas utilizadas nesta aplicação não coincidem na identificação de possíveis pontos influentes. Mesmos os pontos mais afastados não apresentam característica de proporcionar alguma interferência no resultado do ajuste.

Tabela 3 Coordenadas padronizadas dos vetores \hat{e}_{max} e \hat{e}_{min}

Observação(i)	$ \hat{e}_{max} $	$ \hat{e}_{min} $
1	0,4287	1,3312
2	0,4328	1,7522
3	0,4725	0,0749
4	0,4367	0,1386
5	0,3931	0,0516
6	0,2454	0,1157

4 CONCLUSÃO

Tanto a proposta de Cook (1986) como a de Billor e Loynes (1993) foram bastante inovadoras na análise de diagnóstico em modelos de regressão e trouxeram contribuições significativas, dentre as quais se destaca o fato de permitirem a análise de diagnósticos de outros modelos além do normal, como, por exemplo, o modelo logístico.

Um dos pontos fracos da proposta de Cook (1986), destacado no capítulo 4, é justamente a falta de um valor de referência para comparar com o valor obtido para a curvatura normal. Em resposta a essa e a outras dificuldades da versão de Cook (ver capítulo 4), Billor e Loynes (1993) propuseram a inclinação máxima como instrumento da análise de diagnóstico em modelos de regressão e apresentaram sugestões de valores de referência para comparar com o valor da inclinação máxima, sob alguns esquemas de perturbação, do modelo de regressão normal.

Neste trabalho apresentou-se o cálculo da inclinação máxima no caso de esquema de perturbação de observação do modelo logístico, porém, não foi sugerido nenhum valor de referência, para efeito de comparação, que leve à indicação da sensibilidade do modelo. Destaca-se que a natureza de cada problema e o auxílio de outras técnicas de diagnósticos podem ajudar na hora de decidir pela sensibilidade do modelo.

Na aplicação da seção (5.1), em que foi utilizado o modelo normal linear simples, a curvatura normal e a inclinação máxima apresentaram valores que coincidiram na indicação da sensibilidade do modelo adotado. Vale ressaltar que os gráficos indicaram claramente as observações 28, 29 e 41 como as mais afastadas.

Na aplicação correspondente à seção (5.2), em que foi aplicado o modelo logístico múltiplo, não houve observações que evidenciassem possíveis

outliers, pois houve certa variação dos gráficos utilizados na indicação desses pontos. Quanto ao valor da curvatura normal e da inclinação máxima, a falta de um valor de referência, combinado à inexatidão dos gráficos utilizados na detecção de *outliers*, não permitiu decidir pela sensibilidade dos modelos. Espera-se ter contribuído, com este trabalho, para o enriquecimento da análise de diagnóstico de modelos de regressão normal e logístico.

REFERÊNCIAS

- BELSLEY, D. A.; KUH, E.; WELSCH, R. E. **Regression diagnostics**. New York: J. Wiley, 1980. 292 p.
- BILLOR, N.; LOYNES, R. M. Local Influence: a new approach. **Communications in Statistics: Theory and Methods**, Toronto, v. 22, n. 6, p. 1595-1661, 1993.
- BOLFARINE, H.; SANDOVAL, M. C. **Introdução à inferência estatística**. São Paulo: SBM, 2001. 120 p. (Coleção Matemática).
- COOK, R. D. Assessment of local influence: with discussion. **Journal of the Royal Statistical Society: Series B**, London, v. 48, n. 1/2, p. 133-169, 1986.
- COOK, R. D. Detection of influential observations in linear regressions. **Technometrics**, Washington, v. 15, p. 15-18, 1977.
- COOK, R. D.; WEISBERG, S. **Residuals and influence in regression**: school of statistics. New York: Chapman and Hall, 1982. 230 p.
- CORDEIRO, G. M. Modelos of influential observations in linear regressions. **Technometrics**, Washington, v. 15, n. 1, p. 15-18, 1986.
- DAVISON, A. C.; GIGLI, A. Deviance bresiduals and normal score plots. **Biometrika**, London, v. 76, n. 2, p. 211-221, 1989.
- DEMÉTRIO, C. G. B. **Modelos lineares generalizados em experimentação agrônômica**. Piracicaba: ESALQ/USP, 2002. 121 p.
- GRAYBILL, F. A. **Matrices with applications in statistics**. 2nd ed. Belmont: Wadsworth International Group, 1983. 480 p.
- HOFFMAN, K.; KUNZE, R. **Álgebra linear**. 2. ed. Rio de Janeiro: LTC, 1979. 407 p.
- JAMES, B. R. **Probabilidade**: um curso em nível intermediário. Rio de Janeiro: Livro Técnico, 1981. 299 p.
- LIMA, E. **Projeto Euclides**: curso de análise. Rio de Janeiro: IMPA, 1981. v. 2, 546 p.

MAGALHÃES, M. **Probabilidade e variáveis aleatórias**. 2. ed. São Paulo: EDUSP, 2006. 428 p.

MCCULLAGH, P. **Tensor methods in statistics**. London: Chapman and Hall, 1987. 489 p.

MCCULLAGH, P.; NELDER, J. A. Generalized linear models. **Journal of the Royal Statistical Society A**, London, v. 135, p. 370-384, 1989.

NELDER, J. A.; WENDDERBURN, R. W. M. Generalized linear models. **Journal of the Royal Statistical Society A**, London, v. 135, p. 370-384, 1972.

PAULA, G. A. **Modelos de regressão com apoio computacional**. São Paulo: USP, 2004. 324 p.

PREGIBON, D. Logistic regression diagnostics. **Annals of Statistics**, Hayward, v. 9, p. 705-724, 1981.

SCHALL, R.; DUNNE, T. T. A note on the relationship between parameter collinearity and local influence. **Biometrika**, London, v. 79, n. 2, p. 399-404, 1992.

STOKER, J. J. **Differential geometry**. New York: Wiley, 1969. 404 p.

WEISBERG, S. **Applied linear regression**. 2nd ed. New York: J. Wiley, 1985. 336 p.

APÊNDICE

Apêndice A - Definições, teoremas, proposições e outros resultados importantes

Neste apêndice faremos uma síntese de algumas definições, teoremas, proposições e alguns resultados importantes que foram mencionados no transcorrer deste trabalho.

1. Definições

Definição A.1.1 Seja Y uma variável aleatória. Dizemos que a distribuição de Y pertence à **família exponencial de distribuições**, se pudermos escrever sua função densidade (ou de probabilidade) como:

$$f(y; \theta, \phi) = \exp[\phi \{y\theta - b(\theta)\} + c(y, \phi)], \quad (A.1)$$

em que y pertence ao suporte da variável aleatória Y , $\theta_i, i = 1, \dots, q$ são os parâmetros de interesse e ϕ é o parâmetro de dispersão.

Definição A.1.2 Seja Y_1, Y_2, \dots, Y_n uma amostra aleatória de uma variável aleatória Y . A função de probabilidade (f.p) ou de densidade probabilidade (f.d.p) conjunta para essa amostra, dada por

$$f(y_1, \dots, y_n | \theta) = \prod_{i=1}^n f_y(y_i | \theta) \quad (A.2)$$

é denominada de **função verossimilhança** do parâmetro θ , correspondente à amostra observada $\mathbf{y} = (y_1, \dots, y_n)^T$ e é denotada por $L(\theta; \mathbf{y})$.

Definição A.1.3 Definimos a **função log-verossimilhança** em relação a θ e ϕ dado \mathbf{y} , como sendo

$$l(y; \theta, \phi) = \ln f(y; \theta, \phi). \quad (A.3)$$

2. Teoremas

Teorema A.2.1 (Desigualdade de Cauchy-Schwarz) (HOFFMAN; KUNZE, 1979, p. 354) Se V é um espaço com produto interno, então, para quaisquer vetores α e β em V vale a desigualdade

$$|\langle \alpha, \beta \rangle| \leq \|\alpha\| \|\beta\|. \quad (A.4)$$

em que $|\langle \cdot, \cdot \rangle|$ denota o valor absoluto do produto interno de dois vetores quaisquer em V e $\|\cdot\|$ indica a norma de um vetor em V .

Teorema A.2.2 (GRAYBILL, 1983, p. 183) Seja B uma matriz $n \times n$ particionada na seguinte forma:

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

em que B_{ij} tem dimensão $n_i \times n_j, i, j = 1, 2$. Nestas condições, valem os seguintes resultados:

- (i) existem as inversas das matrizes B^{11} e B^{22} ;
- (ii) as matrizes $[B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1}$ e $[B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1}$ existem;
- (iii) a matriz B^{-1} pode ser escrita como

$$B^{-1} = \begin{bmatrix} [B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1} & -B_{11}^{-1}B_{12}[B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1} \\ -B_{22}^{-1}B_{21}[B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1} & [B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1} \end{bmatrix},$$

- (iv) $B^{11} = [B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1} = B_{11}^{-1} + B_{11}^{-1}B_{12}B_{22}^{-1}B_{21}B_{11}^{-1}$.
- (v) $B^{12} = -B_{11}^{-1}B_{12}[B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1} = -B_{11}^{-1}B_{12}B^{22}$.
- (vi) $B^{22} = [B_{22} - B_{21}B_{11}^{-1}B_{12}]^{-1} = B_{22}^{-1} + B_{22}^{-1}B_{21}B_{11}^{-1}B_{12}B_{22}^{-1}$.

$$(vii) \quad B^{21} = -B_{22}^{-1}B_{21}[B_{11} - B_{12}B_{22}^{-1}B_{21}]^{-1} = -B_{22}^{-1}B_{21}B^{11}.$$

$$(viii) \quad |B| = \frac{B^{11}}{B_{22}} = \frac{B_{22}}{B^{11}} \text{ e } |B_{11}B^{11}| = |B_{22}B^{22}|.$$

Teorema A.2.3 (GRAYBILL, 1983, p. 189) Seja C uma matriz $k \times k$ dada por $C = D + \alpha ab^t$,

em que D é uma matriz não singular, a e b são vetores $k \times 1$ e α é um escalar, tal que

$$\alpha \neq - \left[\frac{\sum_{i=1}^k a_i b_i}{d_{ii}} \right]^{-1}.$$

Nessas condições, a inversa da matriz C dada por

$$C^{-1} = D^{-1} + \gamma \alpha^{-1} b^t a,$$

em que $\gamma = -\alpha \left(1 + \alpha \sum_{i=1}^n a_i b_i d_{ii}^{-1} \right)^{-1}$; $a_i^* = \frac{a_i}{d_{ii}}$; $b_i^* = \frac{b_i}{d_{ii}}$ e d_{ii} é o i -

ésimo elemento da diagonal da matriz D .

Teorema A.2.4 (COOK; WEISBERG, 1982, p. 210) Seja A uma matriz simétrica $p' \times p'$; considere a e b matrizes $q \times q'$. Então, as inversas abaixo existem e vale a relação:

$$(A + a^t b)^{-1} = A^{-1} - A^{-1} a^t (I + b A^{-1} a^t)^{-1} b A^{-1},$$

em que I é a matriz identidade $q \times q$.

A.3 Proposições

Proposição A.3.1 Seja F o vetor $q \times 1$ definido na seção (2.3.3), então, pode-se escrever

$$[I_q + FF^t]^{-1} = I_q - \frac{1}{1 + \|F\|^2} FF^t,$$

em que I_q é a matriz identidade de ordem q .

De fato, observe que I_q é uma matriz diagonal e não singular. Fazendo $D = I_q$, $\alpha = b = F$ e $\alpha = 1$, segue que $d_{ii} = 1$ e $\alpha_i = b_i = F_i$. Assim,

$$-\left[\sum_{i=1}^q \frac{\alpha_i b_i}{d_{ii}}\right]^{-1} = -\left[\sum_{i=1}^q \frac{F_i^2}{1}\right]^{-1} = -\frac{1}{\sum_{i=1}^q F_i^2} \neq 1 = \alpha.$$

Logo, todas as condições do teorema (A.2.3) estão satisfeitas. Aplicando o resultado deste teorema, tem-se que

$$\gamma = -1 \left[1 + 1 \sum_{i=1}^q F_i F_i 1^{-1} \right]^{-1} = -\frac{1}{1 + \|F\|^2}.$$

Portando, conclui-se que

$$[I_q + FF^t]^{-1} = I_q - \frac{1}{1 + \|F\|^2} FF^t.$$

Proposição A.3.2 Sejam $B = -F - \lambda FF^t F$ e $C = 1 - F^t F - \lambda F^t FF^t F$,

em que $\lambda = -\frac{1}{1 + \|F\|^2}$ e F é definido na seção (2.3.3) Segue daí que

$$B^t B = \frac{\|F\|^2}{(1 + \|F\|^2)^2}$$

e

$$C^2 = \frac{1}{1 + \|F\|^2}.$$

De fato,

$$\begin{aligned} \text{(i)} \quad B^t B &= F^t F + \lambda F^t FF^t F + \lambda F^t FF^t F + \lambda^2 F^t FF^t FF^t F \\ &= \|F\|^2 + 2\lambda(\|F\|^2)^2 + \lambda^2(\|F\|^2)^3 \\ &= \|F\|^2 + 2\lambda\|F\|^4 + \lambda^2\|F\|^6 \\ &= \|F\|^2 + 2\left(-\frac{1}{1 + \|F\|^2}\right)\|F\|^4 + \left(-\frac{1}{1 + \|F\|^2}\right)^2\|F\|^6 \end{aligned}$$

$$\begin{aligned}
&= \|F\|^2 \left(1 - \frac{2}{1 + \|F\|^2} \|F\|^2 + \frac{1}{(1 + \|F\|^2)^2} \|F\|^4 \right) \\
&= \frac{\|F\|^2}{(1 + \|F\|^2)^2} [(1 + \|F\|^2)^2 - 2(1 + \|F\|^2)\|F\|^2 + 1 + \|F\|^4] \\
&= \frac{\|F\|^2}{(1 + \|F\|^2)^2} (1 + 2\|F\|^2 + \|F\|^4 - 2\|F\|^2 - 2\|F\|^4 + \|F\|^4) \\
&= \frac{\|F\|^2}{(1 + \|F\|^2)^2}.
\end{aligned}$$

(ii)

$$\begin{aligned}
C^2 &= (1 - F^2 F - \lambda F^2 F F^2 F)^2 \\
&= (1 - \|F\|^2 + \lambda \|F\|^4)^2 \\
&= 1 + \|F\|^4 + \lambda^2 \|F\|^8 - 2\|F\|^4 - 2\lambda \|F\|^4 + 2\lambda \|F\|^6 \\
&= 1 - 2\|F\|^2 - 2 \left(-\frac{1}{1 + \|F\|^2} \right) \|F\|^4 + \|F\|^4 + 2 \left(-\frac{1}{1 + \|F\|^2} \right) \|F\|^6 + \left(-\frac{1}{1 + \|F\|^2} \right) \|F\|^8 \\
&= \frac{1}{(1 + \|F\|^2)^2} [(1 + \|F\|^2)^2 - 2\|F\|^2(1 + \|F\|^2)^2 + 2(1 + \|F\|^2)\|F\|^4 + (1 + \|F\|^2)^2 \|F\|^4 - 2(1 + \|F\|^2)\|F\|^6 + \|F\|^8] \\
&= \frac{1}{(1 + \|F\|^2)^2} [1 + 2\|F\|^2 + 3\|F\|^4 - \|F\|^8 - 2\|F\|^2 - 4\|F\|^4 - 2\|F\|^6 + \|F\|^4 + 2\|F\|^6 + \|F\|^8] \\
&= \frac{1}{(1 + \|F\|^2)^2} [1 + 4\|F\|^4 - 4\|F\|^4] \\
&= \frac{1}{(1 + \|F\|^2)^2}.
\end{aligned}$$

Proposição A.3.3 Conforme visto na seção (4.3) do capítulo (4), tem-se que

$$A = \sum_{j=1}^n \left[\left(\frac{\varepsilon_j}{\sigma} \right)^2 - 1 \right].$$

Seja $\left(\frac{\varepsilon_j}{\sigma} \right)^2 = Z_j$, logo $Z_j \sim \chi^2(1)$. Segue daí que

$$A = \sum_{j=1}^n [Z_j - 1]^2$$

$$= \sum_{j=1}^n (Z_j^2 - 2Z_j + 1)^2.$$

Então,

$$\begin{aligned} A^2 &= \{(Z_1^2 - 2Z_1 + 1) + \dots + (Z_n^2 - 2Z_n + 1)\}^2 \\ &= \sum_{j=1}^n (Z_j^2 - 2Z_j + 1)^2 + 2 \sum_{i < j} (Z_i^2 - 2Z_i + 1)(Z_j^2 - 2Z_j + 1) \\ &= \sum_{j=1}^n (Z_j^4 - 4Z_j^3 + 6Z_j^2 - 4Z_j + 1) + 2 \sum_{i < j} (Z_i^2 Z_j^2 - 2Z_i^2 Z_j - 2Z_i Z_j^2 + Z_i^2 + Z_j^2 + 4Z_i Z_j - 2Z_i - 2Z_j) \end{aligned}$$

Utilizando a independência entre os Z_i 's, segue que

$$E[A^2] = \sum_{j=1}^n \{E[Z_j^4] - 4E[Z_j^3] + 6E[Z_j^2] - 4E[Z_j] + 1\} + 2 \sum_{i < j} \{E[Z_i^2]E[Z_j^2] - 2E[Z_i^2 Z_j] - 2E[Z_i]E[Z_j^2]\}$$

A função geradora de momentos (FGM) de uma distribuição $X \stackrel{k}{\sim} (k)$ é dada por $(1 - 2t)^{\frac{k}{2}}$ (JAMES, 1981). Denotando a FGM da variável aleatória

$Z_j = \left(\frac{E_j}{\sigma}\right)^2$, por $M_X(t)$, tem-se que

$$M_X(t) = (1 - 2t)^{\frac{1}{2}}.$$

Sejam $M'_X, M''_X, M_X^{(3)}$ e $M_X^{(4)}$ a primeira, a segunda, a terceira e a quarta derivadas da FGM da variável aleatória Z_j , respectivamente. Segue que

$$M'_X(t) = (1 - 2t)^{-\frac{3}{2}} \Rightarrow E[Z_j] = M'_X(t) \Big|_{t=0} = 1$$

$$M''_X(t) = 3(1 - 2t)^{-\frac{5}{2}} \Rightarrow E[Z_j^2] = M''_X(t) \Big|_{t=0} = 3$$

$$M_X^{(3)}(t) = 15(1 - 2t)^{-\frac{7}{2}} \Rightarrow E[Z_j^3] = M_X^{(3)}(t) \Big|_{t=0} = 15$$

$$M_X^{(4)}(t) = 105(1 - 2t)^{-\frac{9}{2}} \Rightarrow E[Z_j^4] = M_X^{(4)}(t) \Big|_{t=0} = 105.$$

Substituindo-se os resultados acima na equação (A.5) e fazendo algumas simplificações, encontramos

$$\begin{aligned} E[A^2] &= 60n + 8 \sum_{i < j} 1 \\ &= 60n + 8\{(n-1) + (n-2) + (n-3) + \dots + (n - (n-1))\} \\ &= 60n + 8\{n(n-1) - [1 + 2 + 3 + \dots + (n-1)]\} \\ &= 60n + 8\left\{n(n-1) - \frac{n(n-1)}{2}\right\} \\ &= 60n + 4(n^2 - n) \\ &= 56n + 4n^2. \quad (A.6) \end{aligned}$$

4. Outros resultados importantes

A.4.1 Propriedades da função de log-verossimilhança

Seja Y uma variável aleatória pertencente à família exponencial de distribuição, considere a função de log-verossimilhança em relação a θ e ϕ dado \mathcal{Y} , a seguir

$$l(\theta, \phi | \mathcal{Y}) = \ln f(y; \theta, \phi), \quad (A.7)$$

em que y pertence ao suporte da variável aleatória Y .

Dentre as propriedades desta função, pode-se destacar que

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0 \quad (A.8)$$

e

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0. \quad (A.9)$$

Conforme a definição da função de densidade e/ou probabilidade em (A.1), pode-se escrever a equação (a.7), como segue

$$l(\theta, \phi|y) = [y\theta - b(\theta)]\phi + c(y, \phi). \quad (A.10)$$

Derivando-se a equação acima em relação ao parâmetro θ , encontra-se

$$\frac{\partial l}{\partial \theta} = (y - b'(\theta)) \phi, \quad (A.11)$$

em que b' representa a primeira derivada da função $b(\theta)$. Note, agora, que, combinando-se os resultados das equações (A.8) e (A.11), obtém-se

$$0 = E\left(\frac{\partial l}{\partial \theta}\right) = [\mu - b'(\theta)] \phi \Rightarrow \mu = b'(\theta),$$

de onde concluímos que

$$E(Y) = \mu = b'(\theta). \quad (A.12).$$

Derivando-se a equação (A.11) tem-se que

$$\frac{\partial^2 l}{\partial \theta^2} = -b''(\theta) \phi, \quad (A.13)$$

em que b'' é a segunda derivada da função $b(\theta)$. Combinando os resultados das equações (A.9), (A.11) e (A.13), segue que

$$-b''(\theta)\phi + \text{Var}(Y)\phi^2 = 0 \Rightarrow b''(\theta)\phi = \text{Var}(Y)\phi^2,$$

de onde concluímos que

$$\text{Var}(Y) = b''(\theta)\phi^{-1}. \quad (A.14)$$

A.4.2 Forma exponencial da função densidade ou de probabilidade de algumas distribuições

A seguir apresentam-se as principais distribuições pertencentes à família exponencial, escrevendo suas densidades ou funções de probabilidades na forma apresentada na definição (A.1).

- **Normal**

Se $Y \sim N(\mu, \sigma^2)$, então, sua função densidade pode ser escrita como

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right] \\ &= \exp\left\{\log\left[\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}(y - \mu)^2}\right]\right\} \\ &= \exp\left\{\frac{1}{\sigma^2}\left(y\mu - \frac{\mu^2}{2}\right) - \frac{1}{2}\left[\log(2\pi\sigma^2) + \frac{y^2}{\sigma^2}\right]\right\} \end{aligned}$$

em que $-\infty < \mu, y < \infty$ e $\sigma^2 > 0$. Daí obtém-se que, $\phi = \frac{1}{\sigma^2}$, $b(\mu) = \frac{\mu^2}{2}$,

$$c(y, \phi) = -\frac{1}{2}\left[\log\left(\frac{2\pi}{\phi}\right) + y^2\phi\right] \text{ e } V(\mu) = 1.$$

- **Poisson**

Neste caso, tem-se que $Y \sim \text{Poisson}(\lambda)$, a sua função de probabilidade pode ser escrita como

$$\begin{aligned} f(y; \lambda) &= \frac{e^{-\lambda}\lambda^y}{y!} \\ &= \exp\left[\log\left(\frac{e^{-\lambda}\lambda^y}{y!}\right)\right] \\ &= \exp[-\lambda + y \log \lambda - \log y!] \\ &= \exp\left[\mathbf{1}(y\theta - e^\theta) - \log y!\right] \end{aligned}$$

em que $\lambda > 0$ e $y = 0, 1, 2, \dots$. Logo, $\phi = \mathbf{1}$, $b(\theta) = e^\theta$, $c(y, \phi) = -\log y!$ e $V(\lambda) = \lambda$.

- **Binomial**

Seja $Y \sim \text{Binomial}(n, p)$, então, sua função de probabilidade é escrita como

$$f(y, n, p) = \binom{n}{y} p^y (1-p)^{n-y}$$

em que $0 < p < 1$, $y = 0, 1, 2, \dots$. Considere Y^* como a proporção de sucessos em n ensaios independentes, cada um com probabilidade de ocorrência p . Então, $Y = nY^*$. Dessa forma,

$$\begin{aligned} f(ny^*, n, p) &= \binom{n}{ny^*} p^{ny^*} (1-p)^{n-ny^*} \\ &= \exp \left\{ \log \left[\binom{n}{ny^*} p^{ny^*} (1-p)^{n-ny^*} \right] \right\} \\ &= \exp \left\{ \log \binom{n}{ny^*} + ny^* \log \left(\frac{p}{1-p} \right) + n \log(1-p) \right\} \end{aligned}$$

fazendo $\theta = \log \left(\frac{p}{1-p} \right)$ obtém-se que $p = \frac{e^\theta}{1+e^\theta}$. Segue daí que

$$\begin{aligned} f(ny^*; n, p) &= \exp \left\{ \log \binom{n}{ny^*} + ny^* \theta + n \left[\log \left(1 - \frac{e^\theta}{1+e^\theta} \right) \right] \right\} \\ &= \exp \left\{ n \left[y^* \theta - \log(1+e^\theta) \right] + \log \binom{n}{ny^*} \right\} \end{aligned}$$

comparando com forma definida em (A.1), tem-se que, $\phi = n$,

$$b(\theta) = \log(1+e^\theta), \quad c(y, \phi) = \log \binom{n}{ny^*} \quad \text{e} \quad V(p) = p(1-p).$$