

Resumo

Neste trabalho, estudamos a influência local em modelos de regressão. Este método foi proposto inicialmente por Cook (1986) e tem se mostrado como um poderoso instrumento da análise de diagnóstico.

A proposta de Cook consiste em avaliar a curvatura normal de uma superfície, baseada na medida *Likelihood Displacement*, sob uma pequena perturbação no modelo. Em seguida apresentamos a abordagem de Billor e Loynes (1993), que aplicam uma medida *Likelihood Displacement* modificada, cuja primeira derivada não se anula, exceto em casos triviais, com isso a inclinação máxima é usada como medida de influência local.

No final, discutimos três aplicações onde utilizamos algumas técnicas gráficas de diagnóstico e as duas propostas de influência local aqui referidas.

Abstract

In this work, we study the local influence in regression models, this method was offer first of all by Cook (1986) and have show how an important instrument of the diagnostics analysis.

The Cook's approach consist in assessing the normal curvature of the geometric surface, using the measure Likelihood Displacement, under an minor perturbations of statistical model. After we present the approach of the Billor e Loynes (1993), that apply an measure Likelihood Displacement changes whose first derivative does not, except in trivial cases, vanish. Thus the maximum slop is using how measure of the local influence.

In finish, we present three apply where use any technic graphics of the diagnostics and the two proposed of the local influence here reported.

Universidade Federal de Campina Grande
Centro de Ciências e Teconologia
Programa de Pós-Graduação em Matemática
Curso de Mestrado em Matemática

Influência Local em Modelos de Regressão

por

Dorival Lobato Júnior

sob orientação do

Prof. Dr. Francisco Antônio Morais de Souza

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Campina Grande - PB

Agosto/2005

Influência Local em Modelos de Regressão

por

Dorival Lobato Júnior

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de Concentração: Estatística Aplicada

Aprovada por:

Prof. Dr. Antonio José da Silva

Prof. Dra. Claudia Regina Oliveira de Paiva Lima

Prof. Dr. Francisco Antônio Morais de Souza

Orientador

**Universidade Federal de Campina Grande
Centro de Ciências e Tecnologia
Programa de Pós-Graduação em Matemática
Curso de Mestrado em Matemática**

Agosto/2005

Agradecimentos

- A Deus por permitir a minha existência.
- Aos meus irmãos e irmãs que sempre estiveram ao meu lado, seja nos momentos alegres como nos tristes.
- Ao meu orientador Prof. Dr. Francisco Antônio Morais de Souza, pela forma paciente e dedicada que sempre nos recebeu, assim como pelo auxílio nas horas de maior dificuldade durante o transcorrer deste trabalho.
- Aos professores Dr. Antonio José e Dra. Claudia Regina pelas sugestões e importante contribuição para a versão final dessa dissertação.
- A todos os meus professores do mestrado por todos os ensinamentos repassados durante o curso.
- Ao Coordenador do Programa de Pós-graduação em Matemática da Universidade Federal de Campina Grande (UFCG) Prof. Dr. Marco Aurélio, pela forma atenciosa que sempre nos dispensou.
- Ao Prof. Dr. Marivaldo Pereira Matos e a ex-secretária da Pós-graduação em Matemática da Universidade Federal da Paraíba (UFPb) Sônia Maria do Nascimento, por sempre que solicitados terem se colocado a disposição.
- Aos colegas de curso: Gesson, Cícero, Cidade, Haroldo, Fernando, Orlando, Thicyane e Luís Paulo, pela amizade e companheirismo durante esses anos.
- À Secretária da Pós-graduação em Matemática da UFCG, Salete e a todos os outros funcionários do DME-UFCG, 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.

Dedicatória

Aos meus pais **Dorival de
Oliveira Lobato** (in memoriam)
e **Maria Jardelina Mendes
Lobato**.

Conteúdo

Introdução	1
1 Análise de Diagnóstico em Modelos de Regressão	4
1.1 Análise de Regressão	4
1.1.1 Regressão Normal Linear	4
1.1.2 Modelos Lineares Generalizados	8
1.1.3 Regressão Logística	11
1.2 Avaliação do Modelo Ajustado	14
1.2.1 Resíduos	15
1.2.2 Esquemas de Perturbação	17
1.2.3 Influência	18
1.2.4 Influência Local	19
1.2.5 Técnicas Gráficas	20
2 Avaliando a Influência Local através da Curvatura Normal	22
2.1 Introdução	22
2.2 Gráficos de Influência	23
2.3 A Curvatura Normal em um Gráfico de Influência	26
2.3.1 A Curvatura Normal de uma <i>lifted line</i>	28
2.3.2 Uma Forma Operacional para a Curvatura Normal	30
2.3.3 Avaliando a Curvatura Normal em um Ponto Diferente de ω_0	32
2.4 Perturbação no Modelo Normal Linear	37
2.4.1 Avaliando Coeficientes Individuais	42
2.5 Perturbação de Observações na Família Exponencial	49

2.5.1	Observações Individuais em Regressão Logística	53
3	Avaliando a Influência Local Através da Inclinação Máxima de uma <i>lifted line</i>	57
3.1	Introdução	57
3.2	Medida de Afastamento Pela Verossimilhança Modificada	58
3.3	Avaliando a Inclinação Máxima d_{max}^*	67
3.4	Inclinação Máxima no caso de Perturbação de Observação no Modelo Normal Linear	69
3.5	Inclinação Máxima no caso de Perturbação de Observação no Modelo Logístico	71
4	Aplicações	72
4.1	Estimativa do Número de Gansos por Bando	73
4.2	Distúrbio Coronário versus Pressão Sanguínea e Nível de Colesterol	79
4.3	Preferência por um Novo Detergente em Relação a um Detergente Padrão	86
5	Conclusões	91
A	Definições, Teoremas, Proposições e Outros Resultados Importantes	93
A.1	Definições	93
A.2	Teoremas	94
A.3	Proposições	95
A.4	Outros Resultados Importantes	99
A.4.1	Propriedades da Função de Log-verossimilhança	99
A.4.2	Forma Exponencial da Função Densidade ou de Probabilidade de Algumas Distribuições	100
B	Aplicativo R	102
	Bibliografia	104

Introdução

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

Cook (1977), Belsley, Kuh e Welsch (1980) 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 idéias 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 **influên-**

cia local teve grande aceitação entre os pesquisadores e usuários de regressão, Paula (2002).

Billor e Loynes (1993) descrevem algumas dificuldades práticas associadas com a metodologia apresentada por Cook e apresentam uma proposta alternativa para avaliar a influência local, onde procuram superar os problemas encontrados com a aplicação da metodologia sugerida por Cook.

Tanto a metodologia de Cook como a de Billor e Loynes se utilizam da medida *Likelihood Displacement* (afastamento pela verossimilhança) como função objetivo, no entanto outras medidas têm sido usadas para análise de influência local, Fung e Kwan (1997). Como por exemplo, as estimativas dos parâmetros (Lawrence, 1988), a predição dos erros quadrados, Thomas e Cook (1990), a estatística de Pearson da bondade do ajuste, Lee e Zhao (1996) e outros.

A proposta deste trabalho é apresentar o método de diagnóstico da influência local através das abordagens de Cook (1986) e de Billor e Loynes (1993), mostrando o desenvolvimento teórico de ambas, assim como apresentando aplicações práticas das mesmas.

No **primeiro capítulo** tratamos dos modelos de regressão e das técnicas de diagnóstico. Na primeira seção enfatizamos o modelo clássico (normal linear), os modelos lineares generalizados (MLGs) e, em particular, o modelo logístico linear, expomos as suposições básicas para cada um dos modelos e os métodos de estimação dos parâmetros. Na segunda seção definimos os principais resíduos utilizados e discutimos suas aplicações no modelo normal linear e nos MLGs, descrevemos os esquemas de perturbação, avaliamos a influência de uma observação através da medida *Likelihood Displacement* (LD) e da influência local proposta por Cook. Finalizamos o capítulo descrevendo os principais tipos de gráficos de diagnóstico usados para o modelo normal linear e para os MLGs em geral.

No **segundo capítulo** apresentamos a formulação da influência local avaliada através da curvatura normal, definimos e construímos a noção de gráfico de influência como uma superfície geométrica, fundamentamos a idéia de como utilizar a curvatura normal para medir a influência causada por uma perturbação no modelo. Avaliamos o modelo normal linear sob um esquema de perturbação de observação individual, mostrando a forma da curvatura normal para este caso. De forma geral, encontramos

a curvatura normal para os MLGs, sob o esquema de perturbação de uma observação individual e procedemos do mesmo modo para o caso do modelo logístico linear.

No **terceiro capítulo** apresentamos a formulação da influência local avaliada através da primeira derivada. Definimos uma versão da medida *Likelihood Displacement* modificada (LD^*), de modo que sua primeira derivada não se anule.

Considerando a matriz de perturbação geral

$$\mathbf{W}_g = \text{diag} \{1 + \omega_1, 1 + \omega_2, \dots, 1 + \omega_n\}$$

para o modelo normal linear

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

onde $\text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{W}_g^{-1}$, encontramos o gradiente da superfície $(\omega_1, \omega_2, \dots, \omega_n, LD^*(\boldsymbol{\omega}))$ e conseqüentemente a inclinação máxima. Em seguida obtemos um valor de referência para a inclinação máxima. Supondo o esquema de perturbação de uma observação individual no modelo normal linear, encontramos a inclinação máxima e um valor de referência para ela. Finalizamos este capítulo, obtendo o gradiente e, conseqüentemente, a inclinação máxima para o modelo logístico linear, sob o esquema de perturbação de uma observação individual.

No **quarto capítulo** apresentamos três aplicações. Na primeira utilizamos o modelo normal linear acompanhado da utilização de algumas técnicas gráficas para detectar pontos influentes e verificar a validade da suposição de normalidade para a variável resposta, também avaliamos a influência local através da curvatura normal e da inclinação máxima, sob o esquema de perturbação de uma observação individual. Obtemos ainda a inclinação máxima sob o esquema de perturbação da variância, exposto no capítulo 3. Nas duas aplicações restantes adotamos o modelo logístico linear e algumas técnicas gráficas de diagnóstico, avaliamos a influência local através da curvatura normal e da inclinação máxima, obtidas sob um esquema de perturbação de observação individual para o modelo logístico.

No **apêndice A** apresentamos definições, teoremas, proposições e alguns resultados importantes, os quais fizemos uso no transcorrer deste trabalho.

No **apêndice B** descrevemos os comandos para se obter, no aplicativo **R**, os resultados mais importantes que foram utilizados no capítulo 4.

Capítulo 1

Análise de Diagnóstico em Modelos de Regressão

1.1 Análise de Regressão

O objetivo principal da análise de regressão é estudar o efeito que algumas variáveis exercem sobre outras, o que consiste na obtenção e análise de uma relação funcional entre uma das variáveis e as restantes. Deste modo, podemos identificar dois tipos de variáveis, uma variável dependente (ou variável resposta) e as outras independentes (ou explicativas, explanatórias ou covariáveis). A variável dependente é aquela cujo comportamento deseja-se avaliar e as independentes são usadas para explicar a sua variabilidade, Draper & Smith (1981).

1.1.1 Regressão Normal Linear

Consideremos inicialmente o caso mais simples em que temos uma variável resposta Y e uma única variável independente X , admitiremos que a relação entre a média condicional de Y dado X , ou seja, $\mu = \mu(x) = E[Y|X = x]$, seja da forma

$$\mu = \mu(x) = \alpha + \beta x. \quad (1.1)$$

Esta equação 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, denomi-

nada **componente aleatória** e denotada por uma variável aleatória ε . Dessa forma, o modelo completo fica estabelecido pela seguinte equação

$$Y = \alpha + \beta X + \varepsilon. \quad (1.2)$$

O **modelo de regressão normal linear simples** é obtido a partir das seguintes suposições assumidas para a componente aleatória ε :

$$\begin{aligned} E(\varepsilon) &= 0 \\ \text{Var}(\varepsilon) &= \sigma^2 \\ \varepsilon &\sim N(0, \sigma^2). \end{aligned}$$

Como consequência das suposições acima, feitas sobre ε , temos que $Y \sim N(\alpha + \beta x, \sigma^2)$.

Para a obtenção dos **estimadores** dos parâmetros α e β , vamos considerar uma amostra de pares de observações $(X = x_i, Y = y_i)$, $i = 1, 2, \dots, n$, logo por (1.2), podemos escrever para cada par

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

onde os ε_{is} 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, utilizando o método dos mínimos quadrados, minimizamos a função

$$S = S(\alpha, \beta) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n [y_i - \alpha - \beta x_i]^2. \quad (1.3)$$

Assim, precisamos obter as derivadas parciais $\frac{\partial S}{\partial \alpha}$ e $\frac{\partial S}{\partial \beta}$ e em seguida, as igualamos a zero, com isso encontramos o sistema de equações

$$\begin{cases} -2 \sum_{i=1}^n [y_i - (\hat{\alpha} + \hat{\beta} x_i)] &= 0 \\ -2 \sum_{i=1}^n [y_i - (\hat{\alpha} + \hat{\beta} x_i)] x_i &= 0 \end{cases}$$

onde $\hat{\alpha}$ e $\hat{\beta}$ são os valores que minimizam a função (1.3), ou seja, são os **estimadores de mínimos quadrados** para os parâmetros α e β , respectivamente, Draper & Smith (1981). Simplificando as equações do sistema acima, obtemos o sistema de **equações normais** dado a seguir

$$\begin{cases} n\hat{\alpha} + \hat{\beta} \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ \hat{\alpha} \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i. \end{cases} \quad (1.4)$$

Resolvendo o sistema acima achamos a solução

$$\begin{aligned}\hat{\alpha} &= \frac{1}{n} \sum_{i=1}^n y_i - \hat{\beta} \frac{1}{n} \sum_{i=1}^n x_i \\ \hat{\beta} &= \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2}\end{aligned}$$

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

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

É comum encontrarmos na literatura esta equação na forma

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

onde os \hat{y}_i 's são chamados de **valores preditos**.

No **modelo de regressão normal linear múltipla**, onde utilizamos mais de uma variável independente, consideramos que a média condicional da variável resposta Y pode ser expressa em função das variáveis explicativas X_1, X_2, \dots, X_p através da relação linear,

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

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 $(y_i, x_{1i}, x_{2i}, \dots, x_{pi})$, temos o seguinte modelo

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_p x_{pi} + \epsilon_i, \text{ para } i = 1, 2, \dots, n, \quad p < n, \quad (1.5)$$

onde $\epsilon_i \sim N(0, \sigma^2)$, para todo i .

Considere a matriz de variáveis independentes,

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

e os vetores

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \quad (1.6)$$

Dessa forma, podemos escrever o modelo (1.5) matricialmente, por:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (1.7)$$

Para o modelo múltiplo a função S definida em (1.3) pode ser escrita como

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

Derivando a função acima em relação a $\beta_0, \beta_1, \dots, \beta_p$ e igualando a zero essas derivadas, obtemos 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:

$$(\mathbf{X}^t \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^t \mathbf{y}, \quad (1.8)$$

como $(\mathbf{X}^t \mathbf{X})$ é uma matriz quadrada de posto completo, então existe a sua matriz inversa. Assim, o estimador de mínimos quadrados do vetor de parâmetros $\boldsymbol{\beta}$, é dado por,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}. \quad (1.9)$$

O vetor com os valores preditos pode ser escrito como,

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

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

1.1.2 Modelos Lineares Generalizados

Durante muitos anos os modelos normais lineares 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 procurada, Paula (2004).

Com o desenvolvimento computacional, novos modelos têm sido propostos, dentre os quais destacam-se os modelos lineares generalizados (MLGs), apresentado por Nelder e Wedderburn (1972), Paula (2004). 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 \mathbf{Y} tem a função densidade ou função de probabilidade na forma

$$f(y, \theta_i, \phi) = \exp \{ \phi [y\theta_i - b(\theta_i)] + c(y, \phi) \}, \quad (1.10)$$

onde os $\theta_{i's}$ são os parâmetros de interesse e ϕ é o parâmetro de dispersão. Nessa classe são válidas as seguintes propriedades

$$\begin{aligned} E[y_i] &= \mu_i = b'(\theta_i) \\ \text{Var}[y_i] &= \phi^{-1}V_i \\ V_i &= b''(\theta_i) = \frac{d\mu_i}{d\theta_i} \end{aligned}$$

onde b' e b'' representam a primeira e segunda derivadas da função $b(\theta)$, respectivamente e V_i é definida como a função de variância, Paula (2004). A função de variância desempenha um papel importante na família exponencial uma vez que ela 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 $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)^t$ como **preditor linear**, que no caso da regressão normal linear múltipla é definido por,

$$\begin{aligned} \eta_i &= \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} \\ &= \sum_{j=0}^p x_{ji} \beta_j \\ &= \mathbf{x}_i^t \boldsymbol{\beta}, \text{ com } i = 1, 2, \dots, n \end{aligned} \quad (1.11)$$

onde o vetor $\mathbf{x}_i^t = (1, x_{1i}, x_{2i}, \dots, x_{pi})$ é a i -ésima linha da matriz \mathbf{X} (ver Paula (2004)). Além disso, a ligação entre a parte sistemática e a média condicional da variável resposta se dá através da seguinte relação,

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

onde $g(\cdot)$ é denominada de função de ligação, que deve ser monótona e diferenciável. No caso da regressão normal linear é a função identidade.

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

$$l(\boldsymbol{\beta}; \mathbf{y}) = \sum_{i=1}^n \phi \{y_i \theta_i - b(\theta_i)\} + \sum_{i=1}^n c(y_i, \phi). \quad (1.12)$$

Um caso particular ocorre quando

$$\theta_i = \eta_i = \sum_{j=0}^p x_{ij} \beta_j. \quad (1.13)$$

Nesse caso, temos que

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

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

Pelo **Teorema da Fatoração**, a estatística $\mathbf{S} = (s_0, s_1, s_2, \dots, s_p)^t$ é suficiente minimal para o vetor $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^t$, (ver Bolfarine & Sandoval (2001)). As funções de ligação tais que

$$g(\mu_i) = \sum_{j=0}^p s_j \beta_j, \quad (1.14)$$

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

$$\begin{aligned} \mu &= \eta, \\ \log \mu &= \eta, \\ \log \left\{ \frac{\mu}{1 - \mu} \right\} &= \eta. \end{aligned}$$

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

$$\begin{aligned} \mathbf{U}(\boldsymbol{\beta}) &= \frac{\partial l(\boldsymbol{\beta}, \mathbf{y})}{\partial \boldsymbol{\beta}} \\ &= \phi \mathbf{X}^t \mathbf{W}^{1/2} \mathbf{V}^{-1/2} (\mathbf{y} - \boldsymbol{\mu}) \end{aligned} \quad (1.15)$$

e

$$\begin{aligned}\mathbf{K}(\boldsymbol{\beta}) &= E \left\{ -\frac{\partial^2 l(\boldsymbol{\beta}, \mathbf{y})}{\partial \boldsymbol{\beta}} \right\} \\ &= \phi \mathbf{X}^t \mathbf{W} \mathbf{X}\end{aligned}\tag{1.16}$$

onde $\mathbf{W} = \text{diag}\{w_1, \dots, w_n\}$ é a matriz de pesos com componentes $w_i = \frac{(d\mu_i/d\eta_i)^2}{V_i}$, $\mathbf{V} = \text{diag}\{V_1, \dots, V_n\}$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^t$ e V_i é a função de variância definida anteriormente. Em particular, para a ligação canônica, essas quantidades tomam as formas simplificadas

$$\mathbf{U}(\boldsymbol{\beta}) = \phi \mathbf{X}^t (\mathbf{y} - \boldsymbol{\mu})$$

e

$$\mathbf{K}(\boldsymbol{\beta}) = \phi \mathbf{X}^t \mathbf{V} \mathbf{X},$$

respectivamente, Paula (2004).

Para obtermos o estimador de máxima verossimilhança para o vetor de parâmetros $\boldsymbol{\beta}$, devemos encontrar a solução do sistema com $p + 1$ equações não lineares dado por

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}.$$

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

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

onde \mathbf{U}' denota a primeira derivada de $\mathbf{U}(\boldsymbol{\beta})$ em relação a $\boldsymbol{\beta}$. Logo, temos que

$$\mathbf{U}(\boldsymbol{\beta}^{(0)}) + \mathbf{U}'(\boldsymbol{\beta}^{(0)})\hat{\boldsymbol{\beta}} - \mathbf{U}'(\boldsymbol{\beta}^{(0)})\boldsymbol{\beta}^{(0)} \approx \mathbf{0} \Rightarrow$$

$$\mathbf{U}'(\boldsymbol{\beta}^{(0)})\hat{\boldsymbol{\beta}} \approx \mathbf{U}'(\boldsymbol{\beta}^{(0)})\boldsymbol{\beta}^{(0)} - \mathbf{U}(\boldsymbol{\beta}^{(0)}).$$

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

Segue que

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

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

A partir de um valor inicial para $\widehat{\boldsymbol{\beta}}$, a equação acima pode ser usada para obter um processo iterativo na seguinte forma

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} + \left[-\mathbf{U}'(\boldsymbol{\beta}^{(m)}) \right]^{-1} \mathbf{U}(\boldsymbol{\beta}^{(m)}), m = 0, 1, 2, \dots$$

um dos critérios utilizados para que esse processo pare, é dado por

$$\max_j \left\{ \left| \widehat{\beta}_j^{(m+1)} - \widehat{\beta}_j^{(m)} \right| \right\} < \epsilon,$$

onde ϵ é um valor pré-determinado. O procedimento desenvolvido acima é conhecido como **processo iterativo de Newton-Raphson**.

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

$$\widehat{\boldsymbol{\beta}}^{(m+1)} = \widehat{\boldsymbol{\beta}}^{(m)} + \left[\mathbf{K}(\widehat{\boldsymbol{\beta}}^{(m)}) \right]^{-1} \mathbf{U}(\boldsymbol{\beta}^{(m)}), m = 0, 1, 2, \dots$$

substituindo os resultados (1.15) e (1.16) na equação acima, tem-se

$$\begin{aligned} \widehat{\boldsymbol{\beta}}^{(m+1)} &= \widehat{\boldsymbol{\beta}}^{(m)} + \frac{1}{\phi} (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \phi \mathbf{X}^t \mathbf{W}^{1/2} \mathbf{V}^{1/2} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \\ &= (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} (\mathbf{X}^t \mathbf{W} \mathbf{X}) \widehat{\boldsymbol{\beta}}^{(m)} + (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}^{1/2} \mathbf{W}^{1/2} \mathbf{W}^{-1/2} \mathbf{V}^{1/2} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \\ &= (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \{ \boldsymbol{\eta}^{(m)} + \mathbf{W}^{-1/2} \mathbf{V}^{1/2} (\mathbf{y} - \boldsymbol{\mu}^{(m)}) \} \\ &= (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} \mathbf{z}^{(m)} \end{aligned} \tag{1.17}$$

onde $\mathbf{z}^{(m)} = \boldsymbol{\eta}^{(m)} + \mathbf{W}^{-1/2} \mathbf{V}^{1/2} (\mathbf{y} - \boldsymbol{\mu}^{(m)})$ e $\boldsymbol{\eta}^{(m)} = \mathbf{X} \widehat{\boldsymbol{\beta}}^{(m)}$. 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 (1.17) ocorre em um número finito de passos, independente dos valores iniciais utilizados, Paula (2004).

1.1.3 Regressão Logística

A regressão logística se caracteriza pelo fato da variável resposta assumir apenas dois valores possíveis: sucesso ou fracasso. Este é um caso particular dos MLGs, onde a distribuição dessa variável resposta é assumida binomial de parâmetro igual à probabilidade de ocorrer sucesso.

Primeiro vamos tratar da **regressão logística simples**. Seja Y uma variável resposta que assuma o valor 0 para o fracasso e 1 para o sucesso, indicando a presença ou ausência de uma determinada característica de interesse, respectivamente. Consideremos que x seja um valor de uma variável explicativa X e $p(x)$ a probabilidade condicional de sucesso dado x , isto é,

$$p(x) = P(Y = 1|X = x) .$$

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

$$\begin{aligned} \text{logit} : (0, 1) &\longmapsto \mathbb{R} \\ t &\longmapsto \text{logit}(t) = \ln \frac{t}{1-t} . \end{aligned} \quad (1.18)$$

Dessa forma, obtemos o modelo logístico linear simples:

$$\text{logit } p(x) = \eta, \quad (1.19)$$

onde $\eta = \beta_0 + \beta_1 x$ é o preditor linear, β_0 e β_1 são parâmetros desconhecidos. Segue de (1.18) e (1.19) que,

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

daí obtemos que,

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

ou mais precisamente,

$$p(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} . \quad (1.20)$$

Os parâmetros β_0 e β_1 podem ser estimados pelo método de **máxima verossimilhança**, (ver Bolfarine & Sandoval (2001)). Para isso, considere uma amostra aleatória de n observações independentes (y_i, x_i) , $i = 1, 2, \dots, n$; onde x_i é um valor de uma variável independente, discreta ou contínua. Temos 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, obtemos a função de verossimilhança para observações independentes,

$$L(\beta_0, \beta_1; y_1, y_2, \dots, y_n) = \prod_{i=1}^n p(x_i)^{y_i} [1 - p(x_i)]^{1-y_i}. \quad (1.21)$$

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

$$l(\beta_0, \beta_1; y_1, y_2, \dots, y_n) = \sum_{i=1}^n \{y_i \ln [p(x_i)] + (1 - y_i) \ln [1 - p(x_i)]\}. \quad (1.22)$$

Diferenciando (1.22) com relação a β_0 e β_1 e igualando cada um desses resultados a zero, encontramos o sistema com as equações,

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

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

A seguir abordaremos o caso da **regressão logística múltipla**. Consideremos o modelo geral de regressão logística múltipla,

$$\ln \left[\frac{p(\mathbf{x})}{1 - p(\mathbf{x})} \right] = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, \quad (1.23)$$

onde $\mathbf{x} = (1, x_1, \dots, x_p)^t$ contém os valores observados de p variáveis explicativas. A ligação *logit*, neste caso, é dada por,

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

onde $p(\mathbf{x}) = P[Y = 1|\mathbf{x}]$ 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 um procedimento análogo ao utilizado no modelo simples, obtemos

$$p(\mathbf{x}) = \frac{e^\eta}{1 + e^\eta} = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}. \quad (1.24)$$

Para obtermos o estimador de máxima verossimilhança do vetor de parâmetros $\boldsymbol{\beta}$, no modelo logístico múltiplo, desenvolvemos 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:

$$L(\boldsymbol{\beta}; \mathbf{y}) = \prod_{i=1}^n p(\mathbf{x}_i)^{y_i} [1 - p(\mathbf{x}_i)]^{1-y_i}. \quad (1.25)$$

onde $\mathbf{x}_i = (1, x_{1i}, x_{2i}, \dots, x_{pi})^t$, $i = 1, 2, \dots, n$. Segue que, diferenciando a função de log-verossimilhança $l(\boldsymbol{\beta}; \mathbf{y})$, em relação a cada componente do vetor de parâmetros $\boldsymbol{\beta}$ e após igualando essas derivadas a zero, obtemos o sistema:

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

O sistema acima também não admite uma solução explícita, sendo utilizado o processo iterativo de Newton-Raphson para obter uma solução aproximada.

1.2 Avaliação do Modelo Ajustado

Modelos estatísticos são geralmente descrições aproximadas de processos bastante complexos, conseqüentemente podem levar a resultados imprecisos, daí surge uma importante motivação para o estudo de técnicas que avaliem essa inexatidão, Billor e Loynes (1993).

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

Podemos classificar um *outlier*, conforme a sua característica, do seguinte modo, ver Paula, (2004):

- **Pontos Aberrantes:** são pontos que têm uma certa influência sobre os valores ajustados, embora não estejam muito afastados dos demais pontos.
- **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.

- **Pontos Influentes:** são pontos que ao mesmo tempo são de alto *leverage* e aberrantes.

1.2.1 Resíduos

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

$$r_i = y_i - \hat{\mu}_i, \quad (1.26)$$

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

Para $\mathbf{r} = (r_1, r_2, \dots, r_n)^t$, temos que, para o modelo de regressão normal linear,

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

onde $\mathbf{H} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t$ é a matriz *hat* já definida na seção (1.1). O elemento h_{ii} dado a seguir, pertencente à diagonal da matriz \mathbf{H} ,

$$h_{ii} = \mathbf{x}_i^t(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{x}_i, \quad (1.27)$$

desempenha um importante papel na construção de técnicas de diagnóstico. Em particular, $Var(r_i) = (1 - h_{ii})\sigma^2$. Partindo do pré-suposto que todos os pontos exerçam 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, temos que $r_i \sim N(0, \sigma^2)$. Assim podemos obter os **resíduos studentizados**, denotados por t_i , dividindo-se cada r_i pelo seu respectivo desvio padrão amostral $s(1 - h_{ii})^{1/2}$, onde $s^2 = \sum_{i=1}^n \frac{r_i^2}{n - p}$ é o estimador da variância σ^2 . Logo

$$t_i = \frac{r_i}{s(1 - h_{ii})^{1/2}}, \quad i = 1, \dots, n. \quad (1.28)$$

Como r_i não é independente de s^2 , 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 s^2 por $s_{(i)}^2$ na expressão de t_i anterior, onde $s_{(i)}^2$ 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, temos que o novo resíduo *studentizado*,

$$t_i^* = t_i \left(\frac{n - p - 1}{n - p - t_i^2} \right)^{1/2}, \quad (1.29)$$

segue uma distribuição t_{n-p-1} central.

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

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^t \widehat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^t \widehat{\mathbf{W}} \mathbf{z},$$

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

$$\begin{aligned} \mathbf{r}^* &= \widehat{\mathbf{W}}^{-1/2} (\mathbf{z} - \widehat{\boldsymbol{\eta}}) \\ &= \widehat{\mathbf{V}}^{-1/2} (\mathbf{y} - \widehat{\boldsymbol{\mu}}). \end{aligned} \quad (1.30)$$

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

$$t_{S_i} = \frac{\phi^{1/2} (y_i - \widehat{\mu}_i)}{\sqrt{\widehat{v}_i (1 - \widehat{h}_{ii})}}, \quad (1.31)$$

onde h_{ii} foi definido na equação (1.27).

Os resíduos mais utilizados em MLGs são definidos a partir dos componentes da função desvio. A versão padronizada para esses desvios é dada por (ver Paula (2004))

$$t_{D_i} = \frac{\phi^{1/2} d(y_i; \widehat{\mu}_i)}{\sqrt{1 - \widehat{h}_{ii}}}, \quad (1.32)$$

onde $d(y_i; \widehat{\mu}_i) = \pm \sqrt{2} \left\{ y_i (\widehat{\theta}_i^0 - \widehat{\theta}_i) + [b(\widehat{\theta}_i) - b(\widehat{\theta}_i^0)] \right\}^{1/2}$.

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

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

com $0 < y_i < n_i$. Quando $y_i = 0$ temos que

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

e quando $y_i = n_i$, segue que

$$t_{D_i} = - \frac{\{2n_i |\ln \hat{p}_i|\}^{1/2}}{\sqrt{1 - \hat{h}_{ii}}},$$

onde \hat{h}_{ii} assume a seguinte forma

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

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

O resíduo *studentizado* t_{S_i} , para o modelo logístico linear, é dado pela expressão a seguir, Paula (2004)

$$t_{S_i} = \frac{(y_i - n_i \hat{p}_i)^2}{\sqrt{(1 - \hat{h}_{ii}) [n_i \hat{p}_i (1 - \hat{p}_i)]}}. \quad (1.33)$$

1.2.2 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 e Loynes (1993).

Dessa forma, podemos 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 Loynes (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, teremos a oportunidade de apresentar alguns esquemas de perturbação, como por exemplo: perturbação de observação individual em modelos normais e logísticos, perturbação de coeficientes individuais e da matriz de covariância em modelos normal linear.

1.2.3 Influência

Considere a função de verossimilhança para o parâmetro β a seguir

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

onde $l(\beta, y_j)$ denota 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

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

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

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

$$\begin{aligned} D_i &= \frac{(\hat{\beta} - \hat{\beta}_{(i)})^t (\mathbf{X}^t \mathbf{X}) (\hat{\beta} - \hat{\beta}_{(i)})}{ps^2} \\ &= t_i^2 \frac{h_{ii}}{(1 - h_{ii})} \frac{1}{p}. \end{aligned} \quad (1.34)$$

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

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 $l(\beta, \mathbf{y})$ com a retirada desta observação.

Uma medida utilizada para avaliar este impacto é conhecida como afastamento pela verossimilhança (*Likelihood Displacement*), Cook e Weisberg (1982), a qual é definida por

$$LD_i \cong 2[l(\widehat{\boldsymbol{\beta}}, \mathbf{y}) - l(\widehat{\boldsymbol{\beta}}_{(i)}, \mathbf{y})],$$

onde $l(\widehat{\boldsymbol{\beta}}_{(i)}, \mathbf{y})$ é a log-verossimilhança sem a i -ésima observação. Em Paula (2004) é mostrado que

$$LD_i \cong \left[\frac{\widehat{h}_{ii}}{1 - \widehat{h}_{ii}} \right] t_{S_i}^2. \quad (1.35)$$

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

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

1.2.4 Influência Local

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

Considerando o caso geral da distância de Cook $D_{\boldsymbol{\omega}}$, definida na subseção (1.2.3), queremos estudar as mudanças produzidas nesta medida quando $\omega_i \rightarrow 1, \forall i$. Expandindo $D_{\boldsymbol{\omega}}$ em série de Taylor até segunda ordem, em torno de $\boldsymbol{\omega}_0 = \mathbf{1}$, temos que

$$\begin{aligned} D_{\boldsymbol{\omega}} &\cong D_{\boldsymbol{\omega}_0} + (\boldsymbol{\omega}_0 - \boldsymbol{\omega})^t D'_{\boldsymbol{\omega}_0} + \frac{1}{2} (\boldsymbol{\omega}_0 - \boldsymbol{\omega})^t D''_{\boldsymbol{\omega}_0} (\boldsymbol{\omega}_0 - \boldsymbol{\omega}) \\ &\cong \frac{1}{2} (\boldsymbol{\omega}_0 - \boldsymbol{\omega})^t D''_{\boldsymbol{\omega}_0} (\boldsymbol{\omega}_0 - \boldsymbol{\omega}). \end{aligned}$$

Pode-se mostrar que

$$D''_{\boldsymbol{\omega}_0} = \text{diag}(\mathbf{r}) \mathbf{H} \text{diag}(\mathbf{r}),$$

onde $\text{diag}(\mathbf{r}) = \text{diag}\{r_1, \dots, r_n\}$. Cook estuda a maior variação de $D_{\boldsymbol{\omega}}$ em torno de $\boldsymbol{\omega}_0$, o que equivale a maximizar a forma quadrática

$$\boldsymbol{\ell}^t \mathbf{F} \boldsymbol{\ell},$$

onde $\boldsymbol{\ell} = \boldsymbol{\omega}_0 - \boldsymbol{\omega}$, $\boldsymbol{\ell}^t \boldsymbol{\ell} = 1$ e $\mathbf{F} = \text{diag}(\mathbf{r})\mathbf{H}\text{diag}(\mathbf{r})$.

O máximo da forma quadrática $\boldsymbol{\ell}^t \mathbf{F} \boldsymbol{\ell}$ corresponde ao maior autovalor da matriz \mathbf{F} , denotado por λ_{max} . As coordenadas do autovetor $\boldsymbol{\ell}_{max}$, associado a λ_{max} , contêm as influências locais das observações na direção que cause maior alteração. Assim, o gráfico de $|\boldsymbol{\ell}_{max}|$ contra a ordem das observações pode revelar os pontos com maior influência na vizinhança de $\boldsymbol{\omega}_0$, 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 $\boldsymbol{\ell}_{max}$ para avaliar a influência local das observações nas estimativas dos parâmetros é o autovetor correspondente ao maior autovalor da matriz

$$\mathbf{F} = \text{diag}(\widehat{\mathbf{r}}_P)\widehat{\mathbf{H}}\text{diag}(\widehat{\mathbf{r}}_P),$$

onde $\widehat{\mathbf{r}}_P = (\widehat{r}_{P_1}, \dots, \widehat{r}_{P_1})^t$ e $\widehat{r}_{P_i} = \frac{\phi^{1/2}(y_i - \widehat{\mu}_i)}{\widehat{\mathbf{V}}^{1/2}}$ é o i -ésimo resíduo de Pearson avaliado em $\widehat{\boldsymbol{\beta}}$.

1.2.5 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 t_i^* contra a ordem das observações, usado para detectar pontos aberrantes.
- (ii) Gráfico de t_i^* contra os valores ajustados, indica a ocorrência de heterocedasticidade nos dados.
- (iii) Gráfico de h_{ii} contra a ordem das observações, também é usado para detectar observações suspeitas de serem aberrantes.
- (iv) Gráfico das coordenadas do vetor $|\boldsymbol{\ell}_{max}|$ 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_{S_i} 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_{D_i} contra os valores ajustados ou contra a ordem das observações, detecta pontos influentes.
- (iii) Gráfico de \widehat{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 das coordenadas do vetor $|\ell_{max}|$ contra a ordem das observações, detecta pontos com indícios de serem influentes.
- (vi) Gráfico normal de probabilidade para t_{D_i} com envelopes, fornece indícios de afastamento da suposição sobre a distribuição adotada para a variável resposta.

Capítulo 2

Avaliando a Influência Local através da Curvatura Normal

2.1 Introdução

Neste capítulo centraremos nossa atenção para o método de diagnóstico da influência local avaliado através da **curvatura normal**. Este método, introduzido por Cook (1986), consiste em estudar o comportamento de um **gráfico de influência**, que definiremos na Seção (2.2), em torno de um ponto particular ω_0 . Para isso, utiliza-se o conceito de curvatura normal, que pode ser encontrado em textos básicos de Geometria Diferencial.

Em geral os **métodos de diagnóstico** visam a verificação de possíveis afastamentos das suposições feitas para o modelo bem como identificar a existência de observações extremas com alguma interferência desproporcional nos resultados do ajuste, Paula (2004).

Billor e Loynes (1993) apresentam uma proposta alternativa à idéia de Cook para avaliar a influência local, que será objeto de estudo no Capítulo 3. Eles utilizam uma medida de **afastamento pela verossimilhança modificada**, sendo a principal vantagem desta medida o fato de sua primeira derivada não se anular. Com isso a inclinação da curva é utilizada como função de interesse em vez da curvatura.

Outras funções têm sido usadas com o objetivo de avaliar a influência local como, por exemplo, uma estimativa dos parâmetros ou um teste estatístico. Fung e Kwan

(1997), mostram que quando a primeira derivada dessas funções são diferentes de zero, a curvatura não é invariante por escala e assim, podem levar a conclusões ambíguas.

2.2 Gráficos de Influência

No desenvolvimento desta seção utilizaremos o modelo de regressão linear múltipla dado na equação (1.7). Considere a seguinte forma da estatística de influência proposta por Cook (1977),

$$D_i = \frac{\|\widehat{\mathbf{y}} - \widehat{\mathbf{y}}_{(i)}\|^2}{p \sigma^2}, \quad (2.1)$$

onde $\|\cdot\|$ denota a norma euclidiana de um vetor, $\widehat{\mathbf{y}}$ e $\widehat{\mathbf{y}}_{(i)}$ 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 $\boldsymbol{\beta}$. Vale ressaltar que por i -ésima observação entendemos 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 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 (ver Cook (1986)).

Frequentemente deseja-se investigar outros interesses, como por exemplo a heterocedasticidade. Para isso usamos uma versão mais geral para a medida definida em (2.1), dada por

$$D_i(\omega) = \frac{\|\widehat{\mathbf{y}} - \widehat{\mathbf{y}}_\omega\|^2}{p \sigma^2}, \quad (2.2)$$

onde $\widehat{\mathbf{y}}_\omega$ é o vetor de valores ajustados obtido quando a i -ésima observação possui peso ω , com $0 \leq \omega \leq 1$ e as outras observações possuem 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 idéias, permitindo que um vetor de perturbação $\boldsymbol{\omega}$ seja utilizado para perturbar outros componentes do modelo. Uma motivação para essa extensão segue da relação entre $D_i(\boldsymbol{\omega})$ e a log-verossimilhança $l(\boldsymbol{\beta})$, que veremos a seguir.

Observe que podemos escrever o numerador da equação (2.2), na seguinte forma

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

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

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

Com isso

$$\|\hat{\mathbf{y}} - \hat{\mathbf{y}}_\omega\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}_\omega\|^2 - \|\mathbf{y} - \hat{\mathbf{y}}\|^2 - 2\langle \hat{\mathbf{y}} - \hat{\mathbf{y}}_\omega, \mathbf{y} - \hat{\mathbf{y}} \rangle, \quad (2.3)$$

segue ainda que,

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

Do Capítulo 1 sabemos que $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{y}$ e $\hat{\mathbf{y}}_\omega = \mathbf{X}(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{W}\mathbf{y}$, onde $\mathbf{W} = \text{diag}(w_1, w_2, \dots, w_n)$. Com isso, temos que

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

Assim, concluímos que

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

Portanto, a equação em (2.3) fica resumida a

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

Substituindo-se este último resultado na equação em (2.2) tem-se que

$$p D_i(\boldsymbol{\omega}) = \frac{\|\mathbf{y} - \hat{\mathbf{y}}_\omega\|^2 - \|\mathbf{y} - \hat{\mathbf{y}}\|^2}{\sigma^2}. \quad (2.4)$$

Seja $l(\boldsymbol{\beta}_\omega)$ a log-verossimilhança correspondente ao modelo perturbado por um vetor $\boldsymbol{\omega} \in \boldsymbol{\Omega}$, onde $\boldsymbol{\Omega}$ é um conjunto aberto de pequenas perturbações em \mathbb{R}^q . Assumimos que existe um $\boldsymbol{\omega}_0 \in \boldsymbol{\Omega}$ tal que $l(\boldsymbol{\beta}_{\boldsymbol{\omega}_0}) = l(\boldsymbol{\beta})$, para todo vetor $\boldsymbol{\beta}$.

Assumindo que a log-verossimilhança no modelo perturbado é de classe C^2 e considerando apenas as partes relevantes da função de log-verossimilhança para o modelo postulado $l(\widehat{\boldsymbol{\beta}})$ e para o modelo perturbado $l(\widehat{\boldsymbol{\beta}}_\omega)$, onde $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\beta}}_\omega$ são os estimadores de máxima verossimilhança nos modelos postulado e perturbado, respectivamente, obtemos as seguinte relações:

$$\begin{aligned} l(\widehat{\boldsymbol{\beta}}) &\propto -\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \mathbf{x}_i^t \widehat{\boldsymbol{\beta}} \right)^2 \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \widehat{y}_i)^2 \\ &= -\frac{1}{2\sigma^2} \|\mathbf{y} - \widehat{\mathbf{y}}\|^2 \end{aligned}$$

$$\begin{aligned} l(\widehat{\boldsymbol{\beta}}_\omega) &\propto -\frac{1}{2\sigma^2} \sum_{i=1}^n \left(y_i - \mathbf{x}_i^t \widehat{\boldsymbol{\beta}}_\omega \right)^2 \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \widehat{y}_{i\omega})^2 \\ &= -\frac{1}{2\sigma^2} \|\mathbf{y} - \widehat{\mathbf{y}}_\omega\|^2, \end{aligned}$$

segue daí que

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

ou ainda

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

Finalmente, substituindo o último resultado acima na equação (2.4), encontramos a relação entre a estatística $D_i(\boldsymbol{\omega})$ e a função de log-verossimilhança de $\widehat{\boldsymbol{\beta}}$,

$$p D_i(\boldsymbol{\omega}) = 2 \{ l(\widehat{\boldsymbol{\beta}}) - l(\widehat{\boldsymbol{\beta}}_\omega) \}. \quad (2.5)$$

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

$$LD(\boldsymbol{\omega}) = 2 \{ l(\widehat{\boldsymbol{\beta}}) - l(\widehat{\boldsymbol{\beta}}_\omega) \}, \quad (2.6)$$

quando $\boldsymbol{\omega}$ varia numa vizinhança de $\boldsymbol{\omega}_0$. Observe que a equação em (2.5), na verdade relaciona a estatística $D_i(\boldsymbol{\omega})$ de Cook e a medida afastamento pela verossimilhança definida em (2.6). Decorre daí que um gráfico de $LD(\boldsymbol{\omega})$ versus $\boldsymbol{\omega}$ 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:

$$\alpha(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_q \\ LD(\boldsymbol{\omega}) \end{pmatrix}, \quad (2.7)$$

quando $\boldsymbol{\omega}$ varia em $\boldsymbol{\Omega}$. A superfície definida acima é chamada de **gráfico de influência**.

2.3 A Curvatura Normal em um Gráfico de Influência

Nesta seção avaliaremos o comportamento da curvatura normal em torno de um ponto $\boldsymbol{\omega}_0$ pertencente ao gráfico de influência $\alpha(\boldsymbol{\omega})$ definido na seção anterior, lembramos que $\boldsymbol{\omega}_0$ corresponde ao ponto em que $l(\boldsymbol{\beta}) = l(\boldsymbol{\beta}|\boldsymbol{\omega}_0)$. Para o caso particular de termos $q = 1$, o gráfico de influência se reduz a $\alpha(\boldsymbol{\omega}) = \{\alpha_i(\boldsymbol{\omega})\}$, $i = 1, 2$, o que equivale à curva plana $\alpha(\boldsymbol{\omega}) = (\boldsymbol{\omega}, LD(\boldsymbol{\omega}))$, cuja curvatura é dada por, Stoker (1969, p. 26),

$$C = \frac{|\dot{\alpha}_1 \ddot{\alpha}_2 - \dot{\alpha}_2 \ddot{\alpha}_1|}{(\dot{\alpha}_1^2 + \dot{\alpha}_2^2)^{3/2}}, \quad (2.8)$$

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

- $\dot{\alpha}_1 = \frac{d\boldsymbol{\omega}}{d\boldsymbol{\omega}} = 1$, $\ddot{\alpha}_1 = \frac{d^2\boldsymbol{\omega}}{d\boldsymbol{\omega}^2} = 0$;
- $\dot{\alpha}_2|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{dLD(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = LD(\boldsymbol{\omega}) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = 0$, pois $LD(\boldsymbol{\omega})$ alcança um mínimo local em $\boldsymbol{\omega}_0$, e $\ddot{\alpha}_2 = \ddot{LD}(\boldsymbol{\omega})$.

Portanto, obtemos que

$$C = \ddot{\alpha}_2|_{\omega=\omega_0} = \ddot{L}D(\omega_0). \quad (2.9)$$

Quando $q > 1$, um gráfico de influência é uma superfície em \mathbb{R}^{q+1} , 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 , para isso iremos considerar o comportamento das **seções normais** da superfície próximo desse ponto. As seções normais são obtidas pelas intersecções 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** (ver Stoker (1969)).

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

$$\omega(a) = \omega_0 + a\ell, \text{ com } a \in \mathbb{R}. \quad (2.10)$$

Essa reta gera uma **linha projetada** (*lifted line*) sobre a superfície $\alpha(\omega)$ passando pelo ponto $\alpha(\omega_0)$, essa linha projetada consiste de 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 são $\frac{\partial \alpha_i(\omega)}{\partial \omega_j}$, $i = 1, 2, \dots, q+1$; $j = 1, 2, \dots, q$, 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} possui a forma $\mathbf{V} = (\mathbf{I}_q, \mathbf{0})^t$. Seja \mathbf{b}_{q+1} o vetor da base canônica de \mathbb{R}^{q+1} com 1 na última posição e zeros nas outras. Desta 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 $(\ell^t, 0)_{q+1}$. Como os vetores \mathbf{b}_{q+1} e $(\ell^t, 0)_{q+1}$ 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 A Curvatura Normal de uma *lifted line*

A *lifted line* (linha projetada) em uma dada direção ℓ consiste na curva plana dada por $\boldsymbol{\rho}(a) = (\rho_1, \rho_2)^t = [a, LD\boldsymbol{\omega}(a)]^t$ e pode ser vista como o conjunto dos pontos de $\alpha(\boldsymbol{\omega})$ gerado pelos vetores \mathbf{b}_{q+1} e $(\ell^t, 0)^t$. A curvatura normal da curva $\boldsymbol{\rho}(a)$, na direção ℓ , adaptada de (2.8) fica escrita como

$$C_\ell = \frac{|\dot{\rho}_1\ddot{\rho}_2 - \dot{\rho}_2\ddot{\rho}_1|}{(\dot{\rho}_1^2 + \dot{\rho}_2^2)^{3/2}}, \quad (2.11)$$

avaliada em $\boldsymbol{\omega}_0$.

Para melhor avaliar a expressão acima devemos lembrar da suposição de que a função de log-verossimilhança perturbada $l(\boldsymbol{\beta}, \boldsymbol{\omega})$ é de classe C^2 em $(\boldsymbol{\beta}^t, \boldsymbol{\omega}^t)^t$ e, para a superfície $\boldsymbol{\rho}(a) = (\rho_1, \rho_2)^t = [a, LD\boldsymbol{\omega}(a)]^t$, temos o seguinte:

- $\dot{\rho}_1 = \frac{da}{da} = 1$, $\ddot{\rho}_1 = \frac{d^2a}{da^2} = 0$;
- $\frac{d}{da}\boldsymbol{\omega}(a) = \frac{d}{da}(\boldsymbol{\omega}_0 + a\ell)$, logo $\frac{d^2}{da^2}\boldsymbol{\omega}(a) = 0$;
- Pela regra da cadeia,

$$\begin{aligned} \frac{d}{da}LD\boldsymbol{\omega}(a) &= [\text{gradiente}(LD)]^t \left[\frac{d}{da}\boldsymbol{\omega}(a) \right] \\ &= \left[\frac{\partial}{\partial \boldsymbol{\omega}} LD(\boldsymbol{\omega}) \right]^t \boldsymbol{\ell}; \end{aligned}$$

- $\dot{\rho}_2(0) = \frac{d}{da}LD\{\boldsymbol{\omega}(a)\} \Big|_{a=0} = \left(\left[\frac{\partial}{\partial \boldsymbol{\omega}} LD(\boldsymbol{\omega}) \right]^t \boldsymbol{\ell} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = 0$, pois $\boldsymbol{\omega}_0$ é ponto de mínimo local da função LD .

Substituindo esses resultados em (2.11), obtemos

$$\begin{aligned} C_\ell &= \left| \ddot{\rho}_2 \right| \\ &= \left| \ddot{LD}\{\boldsymbol{\omega}(a)\} \right|, \end{aligned}$$

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

$$\begin{aligned}
\ddot{LD} \{\boldsymbol{\omega}(a)\} &= \frac{d^2}{da^2} LD \{\boldsymbol{\omega}(a)\} \\
&= \frac{d}{da} \left[\frac{\partial}{\partial \boldsymbol{\omega}} LD \{\boldsymbol{\omega}(a)\} \right]^t \boldsymbol{\ell} \\
&= \frac{d}{da} \frac{\partial}{\partial \boldsymbol{\omega}^t} LD \{\boldsymbol{\omega}(a)\} \boldsymbol{\ell} \\
&= \frac{d}{da} \left(\frac{\partial}{\partial \omega_1} LD \{\boldsymbol{\omega}(a)\}, \dots, \frac{\partial}{\partial \omega_q} LD \{\boldsymbol{\omega}(a)\} \right) \boldsymbol{\ell} \\
&= \left(\left[\frac{\partial^2}{\partial \omega_1 \partial \boldsymbol{\omega}} LD \{\boldsymbol{\omega}(a)\} \right]^t \frac{d}{da} \boldsymbol{\omega}(a), \dots, \left[\frac{\partial^2}{\partial \omega_q \partial \boldsymbol{\omega}} LD \{\boldsymbol{\omega}(a)\} \right]^t \frac{d}{da} \boldsymbol{\omega}(a) \right) \boldsymbol{\ell} \\
&= \left[\begin{pmatrix} \left[\frac{\partial^2}{\partial \omega_1 \partial \boldsymbol{\omega}} LD \{\boldsymbol{\omega}(a)\} \right]^t \\ \vdots \\ \left[\frac{\partial^2}{\partial \omega_q \partial \boldsymbol{\omega}} LD \{\boldsymbol{\omega}(a)\} \right]^t \end{pmatrix} \boldsymbol{\ell} \right]^t \boldsymbol{\ell} \\
&= \left[\begin{pmatrix} \frac{\partial^2}{\partial \omega_1 \partial \boldsymbol{\omega}^t} LD \{\boldsymbol{\omega}(a)\} \\ \vdots \\ \frac{\partial^2}{\partial \omega_q \partial \boldsymbol{\omega}^t} LD \{\boldsymbol{\omega}(a)\} \end{pmatrix} \boldsymbol{\ell} \right]^t \boldsymbol{\ell} \\
&= \left[\left(\frac{\partial^2}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^t} LD \{\boldsymbol{\omega}(a)\} \right) \boldsymbol{\ell} \right]^t \boldsymbol{\ell} \\
&= \boldsymbol{\ell}^t \left[\frac{\partial^2}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^t} LD \{\boldsymbol{\omega}(a)\} \right]^t \boldsymbol{\ell} \\
&= \boldsymbol{\ell}^t \frac{\partial^2}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^t} LD \{\boldsymbol{\omega}(a)\} \boldsymbol{\ell},
\end{aligned}$$

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

$$C_{\boldsymbol{\ell}} = \left| \boldsymbol{\ell}^t \ddot{LD}(\boldsymbol{\omega}_0) \boldsymbol{\ell} \right|. \quad (2.12)$$

Observe que, pela definição da afastamento pela verossimilhança, temos que

$$\ddot{LD}(\boldsymbol{\omega}_0) = \frac{\partial^2}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^t} \left[2 \left\{ l(\hat{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}) \right\} \right] \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = -2 \frac{\partial^2}{\partial \boldsymbol{\omega} \partial \boldsymbol{\omega}^t} l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0},$$

portanto, a curvatura normal vista em (2.12), fica resumida por

$$C_{\boldsymbol{\ell}} = 2 \left| \boldsymbol{\ell}^t \ddot{\mathbf{F}} \boldsymbol{\ell} \right|, \quad (2.13)$$

onde $\ddot{\mathbf{F}}$ é uma matriz quadrada de dimensão $q \times q$, cujos elementos são dados pelas derivadas $\frac{\partial^2 l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_r \partial \omega_s}$, avaliadas em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

2.3.2 Uma Forma Operacional para a Curvatura Normal

O elemento (r, s) da matriz $\ddot{\mathbf{F}}$ pode ser obtido através da regra da cadeia,

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

Por hipótese, temos que $l(\boldsymbol{\beta}, \boldsymbol{\omega}_0) = l(\boldsymbol{\beta})$, para todo $\boldsymbol{\beta} \in \mathbb{R}^p$, assim $\hat{\boldsymbol{\beta}}_{\omega_0} = \hat{\boldsymbol{\beta}}$, e portanto,

$$\left. \frac{\partial}{\partial \hat{\boldsymbol{\beta}}_\omega} l(\hat{\boldsymbol{\beta}}_\omega) \right|_{\omega=\omega_0} = \left. \frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \boldsymbol{\omega}_0) \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0,$$

pois, $\hat{\boldsymbol{\beta}}$ é o ponto de máximo da função de log-verossimilhança, tanto no modelo pos-
tulado $l(\boldsymbol{\beta})$, quanto no modelo perturbado $l(\boldsymbol{\beta}, \boldsymbol{\omega}_0)$. Dessa forma, a segunda parcela
da derivada $\ddot{\mathbf{F}}_{rs}$ acima, é nula. Logo, temos que

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

avaliada em $\boldsymbol{\omega} = \boldsymbol{\omega}_0$.

Em notação matricial, podemos escrever

$$\ddot{\mathbf{F}} = \mathbf{J}^t \ddot{\mathbf{L}} \mathbf{J}, \tag{2.14}$$

onde \mathbf{J} é uma matriz $p \times q$ cujos elementos são dados por $J_{ks} = \left. \frac{\partial \hat{\beta}_{k\omega}}{\partial \omega_s} \right|_{\omega=\omega_0}$, sendo $\hat{\beta}_{k\omega}$ a k -ésima componente do vetor $\hat{\boldsymbol{\beta}}_\omega$, enquanto que $(-\ddot{\mathbf{L}})$ é a matriz de informação observada de Fisher.

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

$$\left. \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\omega})}{\partial \beta_i} \right|_{\widehat{\boldsymbol{\omega}}} = 0, \quad i = 1, \dots, p. \quad (2.15)$$

Devemos observar que o primeiro membro da igualdade acima é uma função das $p + q$ variáveis $\beta_{1\boldsymbol{\omega}}, \dots, \beta_{p\boldsymbol{\omega}}$ e $\omega_1, \dots, \omega_q$. Já o segundo membro é uma função constante, a função identicamente nula. Segue que, derivando os dois membros da equação (2.15) em relação a $\boldsymbol{\omega}$, e em seguida avaliando a derivada no ponto $\boldsymbol{\omega}_0$, a derivada do segundo membro é nula e a do primeiro é dada por

$$\begin{aligned} \frac{\partial}{\partial \omega_s} \left(\frac{\partial l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}_0}, \boldsymbol{\omega}_0)}{\partial \beta_i} \right) &= \sum_{k=1}^p \left(\frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{\omega})}{\partial \widehat{\beta}_{k\boldsymbol{\omega}} \partial \beta_i} \frac{\partial \widehat{\beta}_{k\boldsymbol{\omega}}}{\partial \omega_s} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} + \sum_{r=1}^q \left(\frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{\omega})}{\partial \omega_r \partial \beta_i} \frac{\partial \omega_r}{\partial \omega_s} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ &= \sum_{k=1}^p \left(\frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{\omega})}{\partial \widehat{\beta}_{k\boldsymbol{\omega}} \partial \beta_i} \frac{\partial \widehat{\beta}_{k\boldsymbol{\omega}}}{\partial \omega_s} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} + \left(\frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{\omega})}{\partial \omega_s \partial \beta_i} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ &= \sum_{k=1}^p \left(\frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{\omega})}{\partial \widehat{\beta}_{k\boldsymbol{\omega}} \partial \beta_i} \frac{\partial \widehat{\beta}_{k\boldsymbol{\omega}}}{\partial \omega_s} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} + \frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}_0}, \boldsymbol{\omega}_0)}{\partial \omega_s \partial \beta_i}. \end{aligned}$$

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

$$\sum_{k=1}^p \left(\frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{\omega})}{\partial \widehat{\beta}_{k\boldsymbol{\omega}} \partial \beta_i} \frac{\partial \widehat{\beta}_{k\boldsymbol{\omega}}}{\partial \omega_s} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} + \frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}_0}, \boldsymbol{\omega}_0)}{\partial \omega_s \partial \beta_i} = 0. \quad (2.16)$$

Como as duas funções de log-verossimilhança são iguais em $\boldsymbol{\omega}_0$, para todo vetor $\boldsymbol{\beta}$, temos que

$$\left. \frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}, \boldsymbol{\omega})}{\partial \widehat{\beta}_{k\boldsymbol{\omega}} \partial \beta_i} \right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \left. \frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}_0}, \boldsymbol{\omega}_0)}{\partial \widehat{\beta}_{k\boldsymbol{\omega}_0} \partial \beta_i} \right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \left. \frac{\partial^2 l(\widehat{\boldsymbol{\beta}})}{\partial \beta_k \partial \beta_i} \right|_{\widehat{\boldsymbol{\omega}}} = \left. \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_i} \right|_{\widehat{\boldsymbol{\omega}}}.$$

Segue daí, e pela equação (2.16), que para $i = 1, \dots, p$, e $s = 1, \dots, q$,

$$\sum_{k=1}^p \left\{ \left(\frac{\partial^2 l(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_i} \right) \Big|_{\widehat{\boldsymbol{\omega}}} \left(\frac{\partial \widehat{\beta}_{k\boldsymbol{\omega}}}{\partial \omega_s} \right) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \right\} + \frac{\partial^2 l(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\omega}_0}, \boldsymbol{\omega}_0)}{\partial \omega_s \partial \beta_i} = 0. \quad (2.17)$$

Agora vamos definir uma matriz $\boldsymbol{\Delta}$, de dimensão $p \times q$, dada por

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

avaliada em $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$. Com isto, a equação (2.17) pode ser escrita na forma

$$\sum_{k=1}^p \ddot{\mathbf{L}}_{ki} \mathbf{J}_{ks} + \boldsymbol{\Delta}_{is} = 0,$$

ou ainda, usando a simetria da matriz $\ddot{\mathbf{L}}$,

$$\sum_{k=1}^p \ddot{\mathbf{L}}_{ik} \mathbf{J}_{ks} + \Delta_{is} = 0.$$

Assim, obtemos a forma mais resumida para a matriz \mathbf{J} ,

$$\mathbf{J} = -\ddot{\mathbf{L}}^{-1} \Delta. \quad (2.18)$$

Substituindo \mathbf{J} na equação (2.14) obtém-se

$$\begin{aligned} \ddot{\mathbf{F}} &= \left(-\Delta^t \ddot{\mathbf{L}}^{-1} \right) \ddot{\mathbf{L}} \left(-\ddot{\mathbf{L}}^{-1} \Delta \right) \\ &= \Delta^t \ddot{\mathbf{L}}^{-1} \Delta. \end{aligned} \quad (2.19)$$

Usando o resultado (2.19) na equação (2.13) obtemos a seguinte forma da curvatura normal para o gráfico de influência $\alpha(\omega)$, na direção de um vetor unitário ℓ ,

$$C_\ell = 2 \left| \ell^t \Delta^t \ddot{\mathbf{L}}^{-1} \Delta \ell \right|. \quad (2.20)$$

Com a equação (2.20) podemos avaliar a influência que pequenas perturbações produzem sobre os componentes do modelo. A direção que produz maior influência local, ℓ_{max} , é o autovetor normalizado correspondente ao maior autovalor C_{max} da matriz $\ddot{\mathbf{F}}$, através dele identificamos os fatores mais influentes para o esquema de perturbação em análise. Similarmente, os autovetores associados com os autovalores intermediários, podem também ser usados para investigar o comportamento da superfície em direções relativas à curvaturas menos extremas.

2.3.3 Avaliando a Curvatura Normal em um Ponto Diferente de ω_0

Quando direcionamos o estudo do gráfico de influência $\alpha(\omega)$ para um ponto diferente de ω_0 , vamos denotar este ponto por ω^* , em geral as linhas projetadas geradas pela reta $\omega(a) = \omega^* + a\ell$ não correspondem mais a seções normais. Porém, curvaturas de seções normais podem ser obtidas, como veremos adiante. Até o final desta subseção nós iremos considerar todas as derivadas avaliadas em ω^* .

Se

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

com $j = 1, \dots, q$, então podemos escrever a matriz \mathbf{V} como

$$\mathbf{V}^t = (\mathbf{I}_q, \dot{\mathbf{F}}).$$

Seja \mathbf{w}_{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 vamos obter a primeira e segunda derivadas da superfície $\alpha\{\boldsymbol{\omega}(a)\}$ em $\boldsymbol{\omega}^*$, que denotaremos por $\dot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}}$ e $\ddot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}}$, respectivamente. Assim, temos que

$$\dot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}} = \left(\frac{\partial(\omega_1^* + a \ell_1)}{\partial a}, \dots, \frac{\partial(\omega_q^* + a \ell_q)}{\partial a}, \frac{\partial LD(\boldsymbol{\omega}(a))}{\partial a} \right)^t. \quad (2.21)$$

Mas,

$$\begin{aligned} \frac{\partial LD\{\boldsymbol{\omega}(a)\}}{\partial a} &= \frac{\partial LD\{\omega_1\}}{\partial \omega_1} \frac{\partial \omega_1(a)}{\partial a} + \dots + \frac{\partial LD\{\omega_q\}}{\partial \omega_q} \frac{\partial \omega_q(a)}{\partial a} \\ &= (-2) \frac{\partial l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_1} \ell_1 + \dots + (-2) \frac{\partial l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_q} \ell_q \\ &= \left((-2) \frac{\partial l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_1}, \dots, (-2) \frac{\partial l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_q} \right) \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_q \end{pmatrix} \\ &= \dot{\mathbf{F}}^t \boldsymbol{\ell}. \end{aligned} \quad (2.22)$$

Segue que substituindo o resultado acima na expressão em (2.21), tem-se

$$\dot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}} = \left(\ell_1, \dots, \ell_q, \dot{\mathbf{F}}^t \boldsymbol{\ell} \right)^t.$$

Podemos ainda escrever este vetor, utilizando a notação de matriz em bloco a seguir

$$\dot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}} = \begin{bmatrix} \boldsymbol{\ell} \\ \dot{\mathbf{F}}^t \boldsymbol{\ell} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_q \\ \dot{\mathbf{F}}^t \end{bmatrix} \boldsymbol{\ell} = \mathbf{V} \boldsymbol{\ell}. \quad (2.23)$$

Para obter a segunda derivada observe que

$$\begin{aligned} \ddot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}} &= \frac{\partial \dot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}}}{\partial a} \\ &= \left(\frac{\partial \ell_1}{\partial a}, \dots, \frac{\partial \ell_q}{\partial a}, \frac{\partial \dot{\mathbf{F}}^t \boldsymbol{\ell}}{\partial a} \right)^t \\ &= \left(0, \dots, 0, \frac{\partial \dot{\mathbf{F}}^t \boldsymbol{\ell}}{\partial a} \right)^t. \end{aligned} \quad (2.24)$$

Note que, podemos escrever $\dot{\mathbf{F}}^t \boldsymbol{\ell} = \sum_{r=1}^q \dot{F}_r \ell_r$, onde $\dot{F}_r = -2 \frac{\partial l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial a}$. Assim,

$$\begin{aligned}
\frac{\partial \dot{\mathbf{F}}^t \boldsymbol{\ell}}{\partial a} &= \sum_{r=1}^q \frac{d \dot{F}_r \ell_r}{da} \\
&= \sum_{r=1}^q \ell_r \frac{d}{da} \left[-2 \frac{\partial l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_r} \right] \\
&= -2 \sum_{r=1}^q \ell_r \left[\frac{\partial^2 l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_r \partial \omega_1} \frac{\partial \omega_1(a)}{\partial a} + \dots + \frac{\partial^2 l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_r \partial \omega_q} \frac{\partial \omega_q(a)}{\partial a} \right] \\
&= -2 \sum_{r=1}^q \ell_r \left[\frac{\partial^2 l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_r \partial \omega_1} \ell_1 + \dots + \frac{\partial^2 l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_r \partial \omega_q} \ell_q \right] \\
&= -2 \sum_{r=1}^q \ell_r \sum_{j=1}^q \frac{\partial^2 l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}})}{\partial \omega_r \partial \omega_j} \ell_j \\
&= -2 \sum_{r=1}^q \sum_{j=1}^q \ell_r \ddot{F}_{rj} \ell_j \\
&= -2 \boldsymbol{\ell}^t \ddot{\mathbf{F}} \boldsymbol{\ell}.
\end{aligned} \tag{2.25}$$

Substituindo o resultado acima no vetor $\ddot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}}$, em (2.24), obtemos a segunda derivada,

$$\ddot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -2 \boldsymbol{\ell}^t \ddot{\mathbf{F}} \boldsymbol{\ell} \end{pmatrix} = -2 \boldsymbol{\ell}^t \ddot{\mathbf{F}} \boldsymbol{\ell} \mathbf{b}_{q+1}, \tag{2.26}$$

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

A curvatura normal $C_{\boldsymbol{\ell}}^*$ associada à direção $\boldsymbol{\ell}$ pode ser escrita como (ver Cook, 1986)

$$C_{\boldsymbol{\ell}}^* = \frac{\|\overline{P}_{\mathbf{V}} \ddot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}}\|}{\|\ddot{\boldsymbol{\alpha}}_{\boldsymbol{\ell}}\|^2}, \tag{2.27}$$

onde $P_{\mathbf{V}}$ é o operador de projeção ortogonal do espaço coluna da matriz \mathbf{V} e $\overline{P}_{\mathbf{V}} = \mathbf{I} - P_{\mathbf{V}}$. A curvatura normal $C_{\boldsymbol{\ell}}^*$ será avaliada a seguir através das formas obtidas para a primeira e segunda derivadas calculadas anteriormente.

Desenvolvendo o denominador da equação (2.27), temos que

$$\begin{aligned}
\|\dot{\boldsymbol{\alpha}}_\ell\|^2 &= \dot{\boldsymbol{\alpha}}_\ell^t \dot{\boldsymbol{\alpha}}_\ell \\
&= (\mathbf{V}\boldsymbol{\ell})^t (\mathbf{V}\boldsymbol{\ell}) \\
&= \boldsymbol{\ell}^t \mathbf{V}^t \mathbf{V} \boldsymbol{\ell} \\
&= \boldsymbol{\ell}^t \begin{bmatrix} \mathbf{I}_q & \dot{\mathbf{F}} \\ & \dot{\mathbf{F}}^t \end{bmatrix} \boldsymbol{\ell} \\
&= \boldsymbol{\ell}^t \left[\mathbf{I}_q + \dot{\mathbf{F}} \dot{\mathbf{F}}^t \right] \boldsymbol{\ell}.
\end{aligned} \tag{2.28}$$

Para obter o numerador de (2.27), observe que

$$\begin{aligned}
P_{\mathbf{V}} &= \mathbf{V}(\mathbf{V}^t \mathbf{V})^{-1} \mathbf{V}^t \\
&= \begin{bmatrix} \mathbf{I}_q \\ \dot{\mathbf{F}}^t \end{bmatrix} \left[\mathbf{I}_q + \dot{\mathbf{F}} \dot{\mathbf{F}}^t \right]^{-1} \begin{bmatrix} \mathbf{I}_q & \dot{\mathbf{F}} \end{bmatrix}.
\end{aligned}$$

Pela proposição (A.3.1), temos que

$$\left[\mathbf{I}_q + \dot{\mathbf{F}} \dot{\mathbf{F}}^t \right]^{-1} = \mathbf{I}_q - \frac{1}{1 + \|\dot{\mathbf{F}}\|^2} \dot{\mathbf{F}} \dot{\mathbf{F}}^t.$$

Segue daí que podemos escrever

$$P_{\mathbf{V}} = \begin{bmatrix} \mathbf{I}_q \\ \dot{\mathbf{F}}^t \end{bmatrix} \left[\mathbf{I}_q + \lambda \dot{\mathbf{F}} \dot{\mathbf{F}}^t \right] \begin{bmatrix} \mathbf{I}_q & \dot{\mathbf{F}} \end{bmatrix},$$

onde $\lambda = -\frac{1}{1 + \|\dot{\mathbf{F}}\|^2}$. Dessa forma, obtemos

$$P_{\mathbf{V}} = \begin{bmatrix} \mathbf{I}_q + \lambda \dot{\mathbf{F}} \dot{\mathbf{F}}^t & \dot{\mathbf{F}} + \lambda \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}} \\ \dot{\mathbf{F}}^t + \lambda \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t & \dot{\mathbf{F}}^t \dot{\mathbf{F}} + \lambda \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}} \end{bmatrix}$$

Logo,

$$\overline{P}_{\mathbf{V}} = \mathbf{I}_{q+1} - P_{\mathbf{V}} = \begin{bmatrix} -\lambda \dot{\mathbf{F}} \dot{\mathbf{F}}^t & -\dot{\mathbf{F}} - \lambda \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}} \\ -\dot{\mathbf{F}}^t - \lambda \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t & 1 - \dot{\mathbf{F}}^t \dot{\mathbf{F}} - \lambda \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}} \end{bmatrix}.$$

Podemos escrever

$$\overline{P}_{\mathbf{V}} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^t & \mathbf{C} \end{bmatrix}, \tag{2.29}$$

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

Note ainda que

$$\begin{aligned}
\|\bar{P}_{\mathbf{V}} \ddot{\alpha}_{\ell}\|^2 &= (\bar{P}_{\mathbf{V}} \ddot{\alpha}_{\ell})^t (\bar{P}_{\mathbf{V}} \ddot{\alpha}_{\ell}) \\
&= \ddot{\alpha}_{\ell}^t \bar{P}_{\mathbf{V}}^t \bar{P}_{\mathbf{V}} \ddot{\alpha}_{\ell} \\
&= [\mathbf{0} \ (-2)\ell^t \ddot{\mathbf{F}} \ell] \begin{bmatrix} \mathbf{A}^t & \mathbf{B} \\ \mathbf{B}^t & C \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^t & C \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ (-2)\ell^t \ddot{\mathbf{F}} \ell \end{bmatrix} \\
&= [\mathbf{0} \ (-2)\ell^t \ddot{\mathbf{F}} \ell] \begin{bmatrix} \mathbf{A}^t\mathbf{A} + \mathbf{B}\mathbf{B}^t & \mathbf{A}^t\mathbf{B} + \mathbf{B}C \\ \mathbf{B}^t\mathbf{A} + C\mathbf{B}^t & \mathbf{B}^t\mathbf{B} + C^2 \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ (-2)\ell^t \ddot{\mathbf{F}} \ell \end{bmatrix} \\
&= (-2\ell^t \ddot{\mathbf{F}} \ell)(\mathbf{B}^t\mathbf{B} + C^2)(-2\ell^t \ddot{\mathbf{F}} \ell) \\
&= (2\ell^t \ddot{\mathbf{F}} \ell)^2(\mathbf{B}^t\mathbf{B} + C^2). \tag{2.30}
\end{aligned}$$

Pela proposição (A.3.2) no apêndice A, temos os seguintes resultados

$$\mathbf{B}^t\mathbf{B} = \frac{\|\dot{\mathbf{F}}\|^2}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2}$$

e

$$C^2 = \frac{1}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2}.$$

Substituindo os resultados acima em (2.30), temos que

$$\begin{aligned}
\|\bar{P}_{\mathbf{V}} \ddot{\alpha}_{\ell}\|^2 &= (2\ell^t \ddot{\mathbf{F}} \ell)^2 \left[\frac{\|\dot{\mathbf{F}}\|^2}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2} + \frac{1}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2} \right] \\
&= (2\ell^t \ddot{\mathbf{F}} \ell)^2 \left[\frac{1 + \|\dot{\mathbf{F}}\|^2}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2} \right] \\
&= (2\ell^t \ddot{\mathbf{F}} \ell)^2 \frac{1}{1 + \|\dot{\mathbf{F}}\|^2}, \tag{2.31}
\end{aligned}$$

segue daí que

$$\|\bar{P}_{\mathbf{V}} \ddot{\alpha}_{\ell}\| = \left| 2\ell^t \ddot{\mathbf{F}} \ell \right| \frac{1}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^{1/2}}. \tag{2.32}$$

Finalmente, substituindo os resultados de (2.28) e (2.32) em (2.27), obtemos a expressão a seguir para a curvatura da seção normal de uma superfície a partir de um ponto $\boldsymbol{\omega}^*$,

$$C_{\boldsymbol{\ell}}^* = \frac{|2\boldsymbol{\ell}^t \ddot{\mathbf{F}} \boldsymbol{\ell}|}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^{1/2} \boldsymbol{\ell}^t \left[\mathbf{I}_q + \dot{\mathbf{F}}^t \dot{\mathbf{F}}\right] \boldsymbol{\ell}}. \quad (2.33)$$

2.4 Perturbação no Modelo Normal Linear

Considere uma amostra y_1, y_2, \dots, y_n onde $\mathbf{Y}_i \sim N(\mu_i, \sigma^2)$, com σ^2 conhecido. Seja $\boldsymbol{\omega}$ um vetor de perturbação de ordem $n \times 1$ para o modelo normal linear múltiplo dado em (1.7), temos que a função de verossimilhança fica estabelecida por

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

consequentemente, a função de log-verossimilhança perturbada é

$$l(\boldsymbol{\beta}, \boldsymbol{\omega}) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log 2\pi - \sum_{i=1}^n \frac{1}{2\sigma^2} \omega_i (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2, \quad (2.34)$$

onde ω_i e y_i são os elementos de ordem i dos vetores $\boldsymbol{\omega}$ e \mathbf{y} , respectivamente e \mathbf{x}_i^t é a i -ésima linha da matriz de variáveis explicativas \mathbf{X} .

Diferenciando a função de log-verossimilhança em relação ao vetor $\boldsymbol{\beta}$, temos que

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\beta}} &= -\frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i \frac{\partial}{\partial \boldsymbol{\beta}} (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2 \\ &= -\frac{1}{2\sigma^2} \sum_{i=1}^n \omega_i 2 (y_i - \mathbf{x}_i^t \boldsymbol{\beta}) \frac{\partial}{\partial \boldsymbol{\beta}} (y_i - \mathbf{x}_i^t \boldsymbol{\beta}) \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n \omega_i (y_i - \mathbf{x}_i^t \boldsymbol{\beta}) \mathbf{x}_i \\ &= \frac{1}{\sigma^2} \sum_{i=1}^n \mathbf{x}_i^t r_i \omega_i \\ &= \frac{1}{\sigma^2} \mathbf{X}^t \text{diag}(r_i) \boldsymbol{\omega}, \end{aligned} \quad (2.35)$$

onde $r_i = y_i - \mathbf{x}_i^t \boldsymbol{\beta}$, com $i = 1, \dots, n$, são os resíduos. Diferenciando agora, o resultado acima, em relação ao vetor $\boldsymbol{\omega}^t$, segue que

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

Observe que pela definição dada na seção (2.3), temos que

$$\Delta = \left. \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^t} \right|_{\hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_0},$$

segue daí que

$$\Delta = \frac{1}{\sigma^2} \mathbf{X}^t \mathbf{D}(\mathbf{e}), \quad (2.36)$$

onde $\mathbf{D}(\mathbf{e}) = \text{diag}(e_i)$ e $e_i = r_i|_{\hat{\boldsymbol{\omega}}} = y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}}$.

Diferenciando a equação (2.35) em relação ao vetor $\boldsymbol{\beta}^t$, e avaliando o resultado em $\hat{\boldsymbol{\beta}}$ e $\boldsymbol{\omega}_0$ obtemos a matriz $\ddot{\mathbf{L}}(\hat{\boldsymbol{\beta}})$ definida na subseção (2.3.2). Segue que

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

onde $\mathbf{W} = \text{diag}(\omega_i)$. Dessa forma,

$$\ddot{\mathbf{L}}(\hat{\boldsymbol{\beta}}) = \left. \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \right|_{\hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_0} = -\frac{1}{\sigma^2} \mathbf{X}^t \mathbf{I}_n \mathbf{X} = -\frac{1}{\sigma^2} \mathbf{X}^t \mathbf{X},$$

consequentemente temos que

$$\ddot{\mathbf{L}}^{-1} = -\sigma^2 (\mathbf{X}^t \mathbf{X})^{-1}. \quad (2.37)$$

Substituindo os resultados dados em (2.36) e (2.37) na equação da curvatura em (2.20), tem-se

$$\begin{aligned} C_{\boldsymbol{\ell}} &= 2 \left| \boldsymbol{\ell}^t \frac{1}{\sigma^2} \mathbf{D}(\mathbf{e}) \mathbf{X} (-\sigma^2) (\mathbf{X}^t \mathbf{X})^{-1} \frac{1}{\sigma^2} \mathbf{X}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell} \right| \\ &= 2 |-1| \left| \frac{1}{\sigma^2} \boldsymbol{\ell}^t \mathbf{D}(\mathbf{e}) \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell} \right| \\ &= \frac{2 \boldsymbol{\ell}^t \mathbf{D}(\mathbf{e}) \mathbf{P}_{\mathbf{X}} \mathbf{D}(\mathbf{e}) \boldsymbol{\ell}}{\sigma^2}, \end{aligned} \quad (2.38)$$

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

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

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

Temos ainda que

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

e assim,

$$\begin{aligned} \Delta &= \left[\frac{\partial}{\partial \boldsymbol{\omega}^t} \left(\frac{\partial l}{\partial \boldsymbol{\theta}} \right) \right] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ &= \begin{bmatrix} \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^t} \Big|_{\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ \frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\omega}^t} \Big|_{\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} \end{bmatrix}, \end{aligned} \quad (2.39)$$

note que o bloco superior da matriz acima já foi calculado na equação (2.37), precisamos ainda de uma expressão útil para o bloco inferior, assim

$$\begin{aligned} \frac{\partial l}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} - \sum_{i=1}^n \left(-\frac{1}{2\sigma^4} \right) \omega_i (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2 \\ &= -\frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{1}{2\sigma^4} \omega_i r_i^2 \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n \omega_i r_i^2 \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \boldsymbol{\omega}^t \mathbf{r}_{sq}, \end{aligned}$$

onde $\mathbf{r}_{sq} = (r_1^2, r_2^2, \dots, r_n^2)^t$. Diferenciando a derivada acima em relação ao vetor $\boldsymbol{\omega}^t$, tem-se

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

Logo,

$$\frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\omega}^t} \Big|_{\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0} = \frac{1}{2\hat{\sigma}^4} \mathbf{e}_{sq}^t, \quad (2.40)$$

onde $\mathbf{e}_{sq} = (e^2, \dots, e^2)^t$ e $e_i^2 = r_i^2 |_{\hat{=}}$.

Substituindo os resultados obtidos em (2.36) e (2.40) na equação (2.39), tem-se

$$\Delta = \begin{bmatrix} \frac{1}{\hat{\sigma}^2} \mathbf{X}^t \mathbf{D}(\mathbf{e}) \\ \frac{1}{2\hat{\sigma}^4} \mathbf{e}_{sq}^t \end{bmatrix}, \quad (2.41)$$

Observe agora que,

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

Assim,

$$\ddot{\mathbf{L}}(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \left. \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \right|_{\hat{=}, \omega = \omega_0} & \left. \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \sigma^2} \right|_{\hat{=}, \omega = \omega_0} \\ \left. \frac{\partial^2 l}{\partial \sigma^2 \partial \boldsymbol{\beta}^t} \right|_{\hat{=}, \omega = \omega_0} & \left. \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \right|_{\hat{=}, \omega = \omega_0} \end{bmatrix}. \quad (2.42)$$

Dessa forma, segue que

$$(i) \left. \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \right|_{\hat{=}, \omega = \omega_0} = -\frac{1}{\hat{\sigma}^2} \mathbf{X}^t \mathbf{X}.$$

(ii) Diferenciando a equação (2.35) temos que

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

Assim,

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

mas, para o modelo de regressão linear normal múltipla dado na equação (1.7),

temos que $\hat{\boldsymbol{\beta}} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}$, logo

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

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

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

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n \omega_i (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2,$$

logo,

$$\frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n \omega_i (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2.$$

Segue daí que

$$\begin{aligned} \left. \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \right|_{=\hat{\boldsymbol{\beta}}, \omega = \omega_0} &= \frac{n}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6} \sum_{i=1}^n (y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}})^2 \\ &= \frac{n}{2\hat{\sigma}^4} - \frac{n}{\hat{\sigma}^6} \sum_{i=1}^n \frac{(y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}})^2}{n}, \end{aligned}$$

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

$$\begin{aligned} \left. \frac{\partial^2 l}{\partial \sigma^2 \partial \sigma^2} \right|_{=\hat{\boldsymbol{\beta}}, \omega = \omega_0} &= \frac{n}{2\hat{\sigma}^4} - \frac{n}{\hat{\sigma}^6} \hat{\sigma}^2 \\ &= -\frac{n}{2\hat{\sigma}^4}. \end{aligned}$$

Substituindo os resultados obtidos nos itens (i), (ii), (iii) e (iv), na equação dada em (2.42), tem-se

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

Logo, temos que

$$\ddot{\mathbf{L}}^{-1}(\hat{\boldsymbol{\theta}}) = - \begin{bmatrix} (\mathbf{X}^t \mathbf{X})^{-1} \hat{\sigma}^2 & \mathbf{0}_{p \times 1} \\ \mathbf{0}_{1 \times p} & \frac{2\hat{\sigma}^4}{n} \end{bmatrix}. \quad (2.43)$$

Finalmente substituindo os resultados obtidos em (2.41) e (2.43) na fórmula da curvatura normal dada em (2.20), tem-se

$$\begin{aligned}
C_{\ell} &= 2 \left| \ell^t \begin{bmatrix} \frac{1}{\hat{\sigma}^2} \mathbf{D}(\mathbf{e})\mathbf{X} & \frac{1}{2\hat{\sigma}^4} \mathbf{e}_{sq} \end{bmatrix} \begin{bmatrix} -(\mathbf{X}^t\mathbf{X})^{-1}\hat{\sigma}^2 & \mathbf{0}_{p \times 1} \\ \mathbf{0}_{1 \times p} & -\frac{2\hat{\sigma}^4}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\sigma}^2} \mathbf{X}^t \mathbf{D}(\mathbf{e}) \\ \frac{1}{2\hat{\sigma}^4} \mathbf{e}_{sq}^t \end{bmatrix} \ell \right| \\
&= 2 \left| \ell^t \begin{bmatrix} \frac{-\mathbf{D}(\mathbf{e})\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \mathbf{D}(\mathbf{e})}{\hat{\sigma}^2} & \frac{\mathbf{e}_{sq}\mathbf{e}_{sq}^t}{2n\hat{\sigma}^4} \end{bmatrix} \ell \right| \\
&= \frac{2}{\hat{\sigma}^2} \ell^t \left[\mathbf{D}(\mathbf{e}) \mathbf{P}_{\mathbf{X}} \mathbf{D}(\mathbf{e}) + \frac{\mathbf{e}_{sq}\mathbf{e}_{sq}^t}{2n\hat{\sigma}^2} \right] \ell. \tag{2.44}
\end{aligned}$$

Vale ressaltar que não são conhecidas expressões analíticas para o autovetor ℓ_{max} da matriz $\ddot{\mathbf{F}}$, quando a curvatura normal for dada pelas fórmulas estabelecidas nas equações (2.38) e (2.44), 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 partição para a matriz de variáveis explicativas $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2]$, onde a variável de interesse é representada pela primeira coluna \mathbf{X}_1 e o seu respectivo coeficiente é o β_1 . Seja $\mathbf{r}_{\mathbf{X}_1}$ o vetor de resíduos da regressão da coluna \mathbf{X}_1 sobre as colunas restantes, ou seja, a matriz \mathbf{X}_2 . Assim,

$$\mathbf{r}_{\mathbf{X}_1} = \mathbf{X}_1 - \hat{\mathbf{X}}_1,$$

onde $\hat{\mathbf{X}}_1 = \mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{X}_1$, logo podemos escrever

$$\begin{aligned}
\mathbf{r}_{\mathbf{X}_1} &= \mathbf{X}_1 - \mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{X}_1 \\
&= [\mathbf{I} - \mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t]\mathbf{X}_1 \\
&= [\mathbf{I} - \mathbf{P}_{\mathbf{X}_2}]\mathbf{X}_1 \\
&= \bar{\mathbf{P}}_{\mathbf{X}_2}\mathbf{X}_1, \tag{2.45}
\end{aligned}$$

onde $\mathbf{P}_{\mathbf{X}_2} = \mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t$ é a matriz de projeção ortogonal sobre as colunas da matriz \mathbf{X}_2 e $\bar{\mathbf{P}}_{\mathbf{X}_2} = \mathbf{I} - \mathbf{P}_{\mathbf{X}_2}$.

Consideraremos a equação da curvatura normal a seguir, (ver Cook (1986))

$$C_{\ell}(\beta_1) = 2 \left| \ell^t [\Delta^t \ddot{\mathbf{L}}^{-1} \Delta - \Delta^t \mathbf{B}_{22} \Delta] \ell \right|, \tag{2.46}$$

onde $\mathbf{B}_{22} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{22}^{-1} \end{bmatrix}$, e as matrizes $\mathbf{\Delta}$ e $\ddot{\mathbf{L}}^{-1}$ foram definidas na seção anterior.

Pelos cálculos feitos para obter a fórmula dada na equação (2.44), temos que

$$\mathbf{\Delta}^t \ddot{\mathbf{L}}^{-1} \mathbf{\Delta} = \frac{-\mathbf{D}(\mathbf{e})\mathbf{X}(\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t \mathbf{D}(\mathbf{e})}{\widehat{\sigma}^2} - \frac{\mathbf{e}_{sq}\mathbf{e}_{sq}^t}{2n\widehat{\sigma}^4}. \quad (2.47)$$

Note que podemos escrever a matriz $\mathbf{\Delta}$ e a matriz $\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}})$, por meio das seguintes partições

$$\mathbf{\Delta} = \begin{bmatrix} \frac{1}{\widehat{\sigma}^2} \mathbf{X}_1^t \mathbf{D}(\mathbf{e}) \\ \frac{1}{\widehat{\sigma}^2} \mathbf{X}_2^t \mathbf{D}(\mathbf{e}) \\ \frac{1}{2\widehat{\sigma}^4} \mathbf{e}_{sq}^t \end{bmatrix},$$

e

$$\ddot{\mathbf{L}}(\widehat{\boldsymbol{\theta}}) = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{L}_{12} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix},$$

onde

$$\mathbf{L}_{11} = -\frac{1}{\widehat{\sigma}^2} \mathbf{X}_1^t \mathbf{X}_1, \mathbf{L}_{12} = \mathbf{0}_{1 \times p}, \mathbf{L}_{21} = \mathbf{0}_{p \times 1}, \mathbf{L}_{22} = \begin{bmatrix} -\frac{1}{\widehat{\sigma}^2} \mathbf{X}_2^t \mathbf{X}_2 & \mathbf{0}_{(p-1) \times 1} \\ \mathbf{0}_{1 \times (p-1)} & -\frac{n}{2\widehat{\sigma}^4} \end{bmatrix},$$

de onde obtemos que

$$\mathbf{L}_{22}^{-1} = \begin{bmatrix} (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \widehat{\sigma}^2 & \mathbf{0}_{(p-1) \times 1} \\ \mathbf{0}_{1 \times (p-1)} & -\frac{2\widehat{\sigma}^4}{n} \end{bmatrix}.$$

Quando substituímos a matriz \mathbf{B}_{22} , juntamente com a partição apresentada para a matriz $\mathbf{\Delta}$ anteriormente, na expressão $\mathbf{\Delta}^t \mathbf{B}_{22} \mathbf{\Delta}$, obtemos que

$$\mathbf{\Delta}^t \mathbf{B}_{22} \mathbf{\Delta} = \frac{-\mathbf{D}(\mathbf{e})\mathbf{X}_2(\mathbf{X}_2^t \mathbf{X}_2)^{-1}\mathbf{X}_2^t \mathbf{D}(\mathbf{e})}{\widehat{\sigma}^2} - \frac{\mathbf{e}_{sq}\mathbf{e}_{sq}^t}{2n\widehat{\sigma}^4}. \quad (2.48)$$

Substituindo os resultados obtidos em (2.47) e (2.48) na equação (2.46), temos que

$$\begin{aligned} C_{\boldsymbol{\ell}}(\beta_1) &= 2 \left| \boldsymbol{\ell}^t \frac{-\mathbf{D}(\mathbf{e})}{\widehat{\sigma}^2} [\mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t - \mathbf{X}_2(\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t] \mathbf{D}(\mathbf{e}) \boldsymbol{\ell} \right| \\ &= \frac{2}{\widehat{\sigma}^2} \left| \boldsymbol{\ell}^t \mathbf{D}(\mathbf{e}) \left[\mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t - (\mathbf{X}_1, \mathbf{X}_2) \begin{pmatrix} \mathbf{0}_{1 \times 1} & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{X}_1^t \\ \mathbf{X}_2^t \end{pmatrix} \right] \mathbf{D}(\mathbf{e}) \boldsymbol{\ell} \right| \\ &= \frac{2}{\widehat{\sigma}^2} \left| \boldsymbol{\ell}^t \mathbf{D}(\mathbf{e}) [\mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t - \mathbf{X} \mathbf{A}_{22} \mathbf{X}^t] \mathbf{D}(\mathbf{e}) \boldsymbol{\ell} \right| \\ &= \frac{2}{\widehat{\sigma}^2} \left| \boldsymbol{\ell}^t \mathbf{D}(\mathbf{e}) \mathbf{X} [(\mathbf{X}^t \mathbf{X})^{-1} - \mathbf{A}_{22}] \mathbf{X}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell} \right| \\ &= \frac{2}{\widehat{\sigma}^2} \left| \boldsymbol{\ell}^t \mathbf{D}(\mathbf{e}) \mathbf{N} \mathbf{D}(\mathbf{e}) \boldsymbol{\ell} \right| \end{aligned} \quad (2.49)$$

onde $\mathbf{N} = \mathbf{X} [(\mathbf{X}^t \mathbf{X})^{-1} - \mathbf{A}_{22}] \mathbf{X}^t$ e $\mathbf{A}_{22} = \begin{pmatrix} \mathbf{0}_{1 \times 1} & \mathbf{0}_{1 \times (p-1)} \\ \mathbf{0}_{(p-1) \times 1} & (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \end{pmatrix}$.

Usando a partição da matriz \mathbf{X} referida no início desta sub-seção, temos que

$$\mathbf{N} = [\mathbf{X}_1 \mathbf{X}_2] [(\mathbf{X}^t \mathbf{X})^{-1} - \mathbf{A}_{22}] [\mathbf{X}_1^t \mathbf{X}_2^t]^t$$

e

$$\mathbf{X}^t \mathbf{X} = \begin{bmatrix} \mathbf{X}_1^t \mathbf{X}_1 & \mathbf{X}_1^t \mathbf{X}_2 \\ \mathbf{X}_2^t \mathbf{X}_1 & \mathbf{X}_2^t \mathbf{X}_2 \end{bmatrix}.$$

Seja

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

então $\mathbf{R}_{11} = \mathbf{X}_1^t \mathbf{X}_1$, $\mathbf{R}_{12} = \mathbf{X}_1^t \mathbf{X}_2$, $\mathbf{R}_{21} = \mathbf{X}_2^t \mathbf{X}_1$ e $\mathbf{R}_{22} = \mathbf{X}_2^t \mathbf{X}_2$. Denotando a inversa da matriz \mathbf{R} por

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

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

(i)

$$\begin{aligned} \mathbf{R}^{11} &= [\mathbf{R}_{11} - \mathbf{R}_{12} \mathbf{R}_{22}^{-1} \mathbf{R}_{21}]^{-1} \\ &= [\mathbf{X}_1^t \mathbf{X}_1 - \mathbf{X}_1^t \mathbf{X}_2 (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1]^{-1} \\ &= [\mathbf{X}_1^t \mathbf{X}_1 - \mathbf{X}_1^t \mathbf{P}_{\mathbf{X}_2} \mathbf{X}_1]^{-1} \\ &= [\mathbf{X}_1^t (\mathbf{I} - \mathbf{P}_{\mathbf{X}_2}) \mathbf{X}_1]^{-1} \\ &= [\mathbf{X}_1^t \bar{\mathbf{P}}_{\mathbf{X}_2} \mathbf{X}_1]^{-1}, \end{aligned}$$

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

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

Logo, usando o resultado (2.45), tem-se

$$\begin{aligned} \mathbf{R}^{11} &= [\mathbf{r}_{\mathbf{X}_1}^t \mathbf{r}_{\mathbf{X}_1}]^{-1} \\ &= [\|\mathbf{r}_{\mathbf{X}_1}\|^2]^{-1} \\ &= \frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2}. \end{aligned}$$

(ii)

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

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

$$\mathbf{R}^{22} = \mathbf{R}_{22}^{-1} - \mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{E}\mathbf{R}_{12}(\mathbf{R}_{22}^t)^{-1} + \mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{E}[\mathbf{E} + \mathbf{R}_{11}^{-1}]^{-1}\mathbf{E}^t\mathbf{R}_{12}(\mathbf{R}_{22}^t)^{-1}$$

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

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

usando o fato de $\mathbf{P}_{\mathbf{X}_2}$ ser idempotente, podemos escrever

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

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

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

Substituindo o resultado acima e o restante dos elementos na última expressão para a matriz \mathbf{R}^{22} acima, tem-se

$$\begin{aligned}\mathbf{R}^{22} &= (\mathbf{X}_2^t\mathbf{X}_2)^{-1} - (\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{X}_1\left(\frac{1}{\|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2}\right)\mathbf{X}_1^t\mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1} \\ &\quad + (\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{X}_1\left(\frac{1}{\|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2}\right)\left[\frac{1}{\|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} - (\mathbf{X}_1^t\mathbf{X}_1)^{-1}\right]^{-1}\frac{1}{\|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2}\mathbf{X}_1^t\mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1},\end{aligned}$$

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

$$\begin{aligned}\mathbf{R}^{22} &= (\mathbf{X}_2^t\mathbf{X}_2)^{-1} - [(\mathbf{X}_2^t\mathbf{X}_2)^{-1}\mathbf{X}_2^t\mathbf{X}_1][\mathbf{X}_1^t\mathbf{X}_2(\mathbf{X}_2^t\mathbf{X}_2)^{-1}] \\ &\quad \left\{ \frac{1}{\|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} - \frac{1}{\|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} \left[\frac{1}{\|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} - \frac{1}{\|\mathbf{X}_1\|^2} \right]^{-1} \frac{1}{\|\hat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} \right\}.\end{aligned}$$

Seja $\widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}$ o vetor de parâmetros estimados da regressão de \mathbf{X}_1 por \mathbf{X}_2 , ou seja, $\widehat{\boldsymbol{\beta}}_{\mathbf{X}_1} = (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1$. Logo,

$$\begin{aligned} \mathbf{R}^{22} &= (\mathbf{X}_2^t \mathbf{X}_2)^{-1} - \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1} \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \left\{ \frac{1}{\|\widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} - \frac{1}{\|\widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} \frac{\|\mathbf{X}_1\|^2 \|\widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2}{\|\mathbf{X}_1\|^2 - \|\widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} \frac{1}{\|\widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} \right\} \\ &= (\mathbf{X}_2^t \mathbf{X}_2)^{-1} + \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1} \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \frac{1}{\|\mathbf{X}_1\|^2 - \|\widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2} \end{aligned}$$

Observe ainda que pelo ítem (i) anterior, temos que

$$\frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} = [\mathbf{X}_1^t \mathbf{X}_1 - \mathbf{X}_1^t \mathbf{P}_{\mathbf{X}_1} \mathbf{X}_1]^{-1}$$

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

$$\begin{aligned} \frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} &= [\|\mathbf{X}_1\|^2 - \mathbf{X}_1^t \mathbf{P}_{\mathbf{X}_1} \mathbf{P}_{\mathbf{X}_1} \mathbf{X}_1]^{-1} \\ &= [\|\mathbf{X}_1\|^2 - \widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}^t \widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}]^{-1} \\ &= \frac{1}{\|\mathbf{X}_1\|^2 - \|\widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}\|^2}. \end{aligned}$$

Daí concluímos que

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

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

$$\mathbf{R}^{22} = (\mathbf{X}_2^t \mathbf{X}_2)^{-1} + \frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1} \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t.$$

(iii)

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

vamos denotar por b o numerador da expressão entre colchetes no resultado acima.

Logo, temos que

$$b = \mathbf{r}_{\mathbf{X}_1}^t \mathbf{r}_{\mathbf{X}_1} + \mathbf{X}_1^t \mathbf{X}_2 \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1},$$

substituindo $\mathbf{r}_{\mathbf{X}_1}$ pela sua expressão dada na equação (2.45) e $\widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}$ pela forma definida no item (ii), no resultado acima, tem-se

$$\begin{aligned}
 b &= \mathbf{X}_1^t \overline{\mathbf{P}}_{\mathbf{X}_2} \overline{\mathbf{P}}_{\mathbf{X}_2} \mathbf{X}_1 + \mathbf{X}_1^t \mathbf{X}_2 (\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1 \\
 &= \mathbf{X}_1^t \overline{\mathbf{P}}_{\mathbf{X}_2} \mathbf{X}_1 + \mathbf{X}_1^t \mathbf{P}_{\mathbf{X}_2} \mathbf{X}_1 \\
 &= \mathbf{X}_1^t [\overline{\mathbf{P}}_{\mathbf{X}_2} + \mathbf{P}_{\mathbf{X}_2}] \mathbf{X}_1 \\
 &= \mathbf{X}_1^t \mathbf{I} \mathbf{X}_1 \\
 &= \mathbf{X}_1^t \mathbf{X}_1 \\
 &= \|\mathbf{X}_1\|^2,
 \end{aligned}$$

agora substituímos o resultado de b acima, na última expressão encontrada para \mathbf{R}^{12} , segue que

$$\mathbf{R}^{12} = -\frac{\widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t}{\|\mathbf{r}_{\mathbf{X}_1}\|^2}.$$

(iv)

$$\begin{aligned}
 \mathbf{R}^{21} &= -\mathbf{R}_{22}^{-1} \mathbf{R}_{21} \mathbf{R}^{11} \\
 &= -(\mathbf{X}_2^t \mathbf{X}_2)^{-1} \mathbf{X}_2^t \mathbf{X}_1 \left(\frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} \right) \\
 &= -\frac{\widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}}{\|\mathbf{r}_{\mathbf{X}_1}\|^2}.
 \end{aligned}$$

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

$$\mathbf{R}^{-1} = \begin{bmatrix} \frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} & -\frac{\widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} \\ -\frac{\widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} & (\mathbf{X}_2^t \mathbf{X}_2)^{-1} + \frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1} \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \end{bmatrix}.$$

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

$$\mathbf{N} = \frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} \left\{ [\mathbf{X}_1 - \mathbf{X}_2 \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}] [\mathbf{X}_1^t - \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}^t \mathbf{X}_2^t] \right\}.$$

Note que $\mathbf{X}_2 \widehat{\boldsymbol{\beta}}_{\mathbf{X}_1}$ corresponde ao vetor de valores preditos da regressão de \mathbf{X}_1 sobre \mathbf{X}_2 , ou seja, $\widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}$. Assim, podemos escrever

$$\mathbf{N} = \frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} \left\{ [\mathbf{X}_1 - \widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}] [\mathbf{X}_1^t - \widehat{\boldsymbol{\mu}}_{\mathbf{X}_1}^t] \right\},$$

temos ainda que os elementos do vetor $\mathbf{X}_1 - \hat{\boldsymbol{\mu}}_{\mathbf{X}_1}$, na expressão acima, representam os resíduos da regressão. Portanto, segue que

$$\mathbf{N} = \frac{1}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} \mathbf{r}_{\mathbf{X}_1} \mathbf{r}_{\mathbf{X}_1}^t.$$

Finalmente, substituindo a expressão encontrada para \mathbf{N} acima, na equação (2.49), obtemos a fórmula para a curvatura normal a seguir

$$C_{\boldsymbol{\ell}}(\beta_1) = \frac{2}{\hat{\sigma}^2} \left| \boldsymbol{\ell}^t \mathbf{D}(\mathbf{e}) \frac{\mathbf{r}_{\mathbf{X}_1} \mathbf{r}_{\mathbf{X}_1}^t}{\|\mathbf{r}_{\mathbf{X}_1}\|^2} \mathbf{D}(\mathbf{e}) \boldsymbol{\ell} \right|. \quad (2.50)$$

Para obtermos o valor máximo que a curvatura normal $C_{\boldsymbol{\ell}}$ pode assumir, quando está sujeita à condição de que $\|\boldsymbol{\ell}\| = 1$, devemos observar que na verdade, a expressão $\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell}$ da equação (2.50) é um escalar, logo é igual ao seu transposto. Portanto, podemos reescrever a referida equação como

$$C_{\boldsymbol{\ell}}(\beta_1) = \frac{2 (\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell})^2}{\|\mathbf{r}_{\mathbf{X}_1}\|^2 \hat{\sigma}^2}.$$

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 $\|\mathbf{r}_{\mathbf{X}_1}\|^2 = \sum_{i=1}^n r_i^2$. Dessa forma, segue que

$$\begin{aligned} C_{\boldsymbol{\ell}}(\beta_1) &= \frac{2 (\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell})^2}{\sum_{i=1}^n r_i^2 \frac{\sum_{i=1}^n e_i^2}{n}} \\ &= \frac{2n (\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell})^2}{\sum_{i=1}^n r_i^2 \sum_{i=1}^n e_i^2}. \end{aligned}$$

Note que, para obter o valor máximo da curvatura normal acima, basta analisar a expressão $(\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell})^2$, pois o restante da fórmula é constante. Assim, seja a função vetorial

$$f(\boldsymbol{\ell}) = (\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell})^2.$$

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

$$f(\boldsymbol{\ell}) = (\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e}) \boldsymbol{\ell})^2 \leq (\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e}) \mathbf{D}(\mathbf{e}) \mathbf{r}_{\mathbf{X}_1}) (\boldsymbol{\ell}^t \boldsymbol{\ell}),$$

como o vetor $\boldsymbol{\ell}$ está sujeito à condição $\boldsymbol{\ell}^t \boldsymbol{\ell} = 1$, segue que

$$\begin{aligned} f(\boldsymbol{\ell}) &\leq (\mathbf{r}_{\mathbf{X}_1}^t \mathbf{D}(\mathbf{e})) (\mathbf{D}(\mathbf{e}) \mathbf{r}_{\mathbf{X}_1}) \\ &\leq \|\mathbf{D}(\mathbf{e}) \mathbf{r}_{\mathbf{X}_1}\|^2 \\ &\leq \sum_{i=1}^n e_i^2 r_i^2, \end{aligned}$$

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

$$\boldsymbol{\ell} = c \mathbf{D}(\mathbf{e})\mathbf{r}_{\mathbf{X}_1},$$

onde 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 $\mathbf{D}(\mathbf{e})\mathbf{r}_{\mathbf{X}_1}$. Portanto, concluímos daí que

$$C_{max}(\beta_1) = \frac{2n \sum_{i=1}^n e_i^2 r_i^2}{\sum_{i=1}^n r_i^2 \sum_{i=1}^n e_i^2} \quad (2.51)$$

é a curvatura máxima, que obviamente ocorre na direção

$$\boldsymbol{\ell}_{max} = \mathbf{D}(\mathbf{e})\mathbf{r}_{\mathbf{X}_1}. \quad (2.52)$$

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

2.5 Perturbação de Observações na Família Exponencial

Nesta seção, iremos estender 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 (1.10). Logo, podemos 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

$$l(\boldsymbol{\beta}, \phi | \boldsymbol{\omega}) = \phi \sum_{i=1}^n \omega_i (y_i \theta_i - b(\theta_i)) + \sum_{i=1}^n \omega_i c(y_i, \phi), \quad (2.53)$$

onde 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, podemos escrever

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

Logo, temos que

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

Pela equação (2.19) temos a seguinte relação

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

onde o elemento (r,s) da matriz $\boldsymbol{\Delta}$ é dado por

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

e da matriz $\ddot{\mathbf{L}}$ é

$$\ddot{L}_{rs} = \left. \frac{\partial^2 l}{\partial \beta_r \partial \beta_s} \right|_{\hat{\omega} = \omega_0}.$$

Segue que, derivando a função de log-verossimilhança dada na equação (2.53), em relação a ω_s , obtém-se

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

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

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

Portanto, a matriz Δ pode ser escrita da seguinte forma

$$\begin{aligned}\Delta &= \begin{bmatrix} \phi \left(y_1 - \frac{\partial b_1}{\partial \widehat{\theta}_1} \right) \frac{\partial \widehat{\theta}_1}{\partial \widehat{\beta}_1} & \dots & \phi \left(y_n - \frac{\partial b_n}{\partial \widehat{\theta}_n} \right) \frac{\partial \widehat{\theta}_n}{\partial \widehat{\beta}_1} \\ \vdots & \ddots & \vdots \\ \phi \left(y_1 - \frac{\partial b_1}{\partial \widehat{\theta}_1} \right) \frac{\partial \widehat{\theta}_1}{\partial \widehat{\beta}_p} & \dots & \phi \left(y_n - \frac{\partial b_n}{\partial \widehat{\theta}_n} \right) \frac{\partial \widehat{\theta}_n}{\partial \widehat{\beta}_p} \end{bmatrix} \\ &= \phi \begin{bmatrix} \frac{\partial \widehat{\theta}_1}{\partial \widehat{\beta}_1} & \dots & \frac{\partial \widehat{\theta}_n}{\partial \widehat{\beta}_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \widehat{\theta}_1}{\partial \widehat{\beta}_p} & \dots & \frac{\partial \widehat{\theta}_n}{\partial \widehat{\beta}_p} \end{bmatrix} \begin{bmatrix} y_1 - \frac{\partial b_1}{\partial \widehat{\theta}_1} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & y_n - \frac{\partial b_n}{\partial \widehat{\theta}_n} \end{bmatrix} \\ &= \phi \widehat{\boldsymbol{\theta}}^t \text{diag} \left(y_i - \frac{\partial b_i}{\partial \widehat{\theta}_i} \right), \text{ com } i = 1, \dots, n;\end{aligned}$$

onde $\widehat{\boldsymbol{\theta}}$ é a matriz $\dot{\boldsymbol{\theta}}$ expressa no item (i), avaliada em $\widehat{\boldsymbol{\beta}}$ e $\boldsymbol{\omega}_0$. Seja $u_i = y_i - \frac{\partial b_i}{\partial \widehat{\theta}_i}$ e $D(\mathbf{u}) = \text{diag}(u_i)$, então podemos escrever

$$\Delta = \phi \widehat{\boldsymbol{\theta}}^t D(\mathbf{u}). \quad (2.54)$$

Quando derivamos a função de log-verossimilhança na equação (2.53) em relação a β_r , obtemos

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

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

$$\ddot{\mathbf{L}}_{rs} = \phi \sum_{i=1}^n \left\{ -\frac{\partial^2 b_i}{\partial \widehat{\theta}_i^2} \frac{\partial \widehat{\theta}_i}{\partial \beta_r} \frac{\partial \widehat{\theta}_i}{\partial \beta_s} + u_i \frac{\partial^2 \widehat{\theta}_i}{\partial \beta_r \partial \beta_s} \right\},$$

note que $\frac{\partial^2 \widehat{\theta}_i}{\partial \widehat{\beta}_r \partial \widehat{\beta}_s}$ corresponde ao elemento (r, s) da matriz $\ddot{\boldsymbol{\theta}}_i \Big|_{\widehat{\boldsymbol{\omega}} = \boldsymbol{\omega}_0} = \widehat{\boldsymbol{\theta}}_i$, $\frac{\partial \widehat{\theta}_i}{\partial \widehat{\beta}_r}$ e $\frac{\partial \widehat{\theta}_i}{\partial \widehat{\beta}_s}$ são respectivamente os elementos (i, r) e (i, s) da matriz $\dot{\boldsymbol{\theta}}_i \Big|_{\widehat{\boldsymbol{\omega}} = \boldsymbol{\omega}_0} = \widehat{\boldsymbol{\theta}}_i$, onde $\ddot{\boldsymbol{\theta}}_i$ e $\dot{\boldsymbol{\theta}}_i$ foram expressas no item (i) desta seção. Observe ainda que $\frac{\partial^2 b_i}{\partial \widehat{\theta}_i^2}$ é o elemento (i, i)

da matriz $\ddot{\mathbf{b}}$, dada no item (ii), no início desta seção. Assim, podemos escrever

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

Pela simetria da matriz $\dot{\theta}$, temos que $\hat{\theta}_{ir} = \hat{\theta}_{ri}^t$. Logo,

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

Portanto, a matriz $\ddot{\mathbf{L}}$ pode ser escrita na seguinte forma matricial

$$\ddot{\mathbf{L}} = \phi \sum_{i=1}^n u_i \hat{\theta}_i - \phi \hat{\theta}^t \ddot{\mathbf{b}} \hat{\theta}. \quad (2.55)$$

Para finalizar, substituímos os resultados das equações (2.54) e (2.55) em (2.19), com isso obtemos

$$\begin{aligned} \ddot{\mathbf{F}} &= \phi D(\mathbf{u}) \hat{\theta} \left[\phi \left(\sum_{i=1}^n u_i \hat{\theta}_i - \hat{\theta}^t \ddot{\mathbf{b}} \hat{\theta} \right) \right]^{-1} \phi \hat{\theta}^t D(\mathbf{u}) \\ &= \phi D(\mathbf{u}) \hat{\theta} \left[\sum_{i=1}^n u_i \hat{\theta}_i - \hat{\theta}^t \ddot{\mathbf{b}} \hat{\theta} \right]^{-1} \hat{\theta}^t D(\mathbf{u}). \end{aligned} \quad (2.56)$$

Conseqüentemente, obtemos a curvatura normal substituindo o resultado acima, na equação (2.13) dada na seção (2.3).

Muitos MLGs são casos especiais da função de log-verossimilhança em (2.53), Cook (1986). Se considerarmos a condição $\theta_i(\boldsymbol{\beta}) = g(\mathbf{x}_i^t \boldsymbol{\beta})$, onde $i = 1, \dots, n$ 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

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

Usando a regra da cadeia, temos que o elemento (r,s) da matriz $\dot{\theta}$ 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, temos que

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

onde \dot{g}_r é a derivada $\frac{\partial g}{\partial \eta_r}$. Logo, podemos expressar a matriz $\dot{\boldsymbol{\theta}}$ como

$$\begin{aligned}\dot{\boldsymbol{\theta}} &= \begin{bmatrix} \dot{g}_1 x_{11} & \dots & \dot{g}_1 x_{1p} \\ \vdots & \ddots & \vdots \\ \dot{g}_n x_{n1} & \dots & \dot{g}_n x_{np} \end{bmatrix} \\ &= \begin{bmatrix} \dot{g}_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dot{g}_n \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{np} \end{bmatrix} \\ &= \text{diag}(\dot{g}_i) \mathbf{X}, \quad i = 1, \dots, n.\end{aligned}\tag{2.57}$$

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

Agora iremos obter a matriz $\ddot{\boldsymbol{\theta}}_i$, note que

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

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

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

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

$$\left(\ddot{\boldsymbol{\theta}}_i \right)_{rs} = \ddot{g}_i x_{is} x_{ir}.$$

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

$$\ddot{\boldsymbol{\theta}}_i = \ddot{g}_i \mathbf{x}_i \mathbf{x}_i^t.\tag{2.58}$$

2.5.1 Observações Individuais em Regressão Logística

Seja y_1, \dots, y_n uma amostra aleatória de variáveis independentes, com $y_i \sim \text{Binomial}(m_i, p_i)$, 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

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

onde $\boldsymbol{\beta}$ é o vetor de parâmetros desconhecidos do modelo e $p_i = P[y_i = 1 | \mathbf{x}_i]$. Pela seção (1.3) do Capítulo (1), sabemos que

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

segue daí que

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

e

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

Portanto, podemos escrever a função de log-verossimilhança como

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

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

$$\begin{aligned} \frac{\partial l}{\partial \beta_s} &= \sum_{i=1}^n \omega_i y_i \frac{\partial \eta_i}{\partial \beta_s} + \sum_{i=1}^n \omega_i m_i \frac{\partial}{\partial \beta_s} \ln \left(\frac{1}{1 + e^{\eta_i}} \right) \\ &= \sum_{i=1}^n \omega_i y_i x_{is} - \sum_{i=1}^n \omega_i m_i x_{is} \left(\frac{e^{\eta_i}}{1 + e^{\eta_i}} \right) \\ &= \sum_{i=1}^n \omega_i y_i x_{is} - \sum_{i=1}^n \omega_i m_i x_{is} p_i. \end{aligned} \tag{2.59}$$

Agora derivamos o resultado acima em relação a ω_r , de onde obtemos 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 $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^t, \hat{p}_i)^t$ e $\boldsymbol{\omega} = \boldsymbol{\omega}_0$, obtemos o elemento (r, s) da matriz $\boldsymbol{\Delta}$, ou seja,

$$\begin{aligned} \Delta_{rs} &= \left. \frac{\partial^2 l}{\partial \beta_s \partial \omega_r} \right|_{\hat{\boldsymbol{\theta}}, \boldsymbol{\omega} = \boldsymbol{\omega}_0} \\ &= x_{rs} (y_r - m_r \hat{p}_r). \end{aligned}$$

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

$$\begin{aligned} \Delta_{rs} &= x_{rs} v_r^{1/2} \left(\frac{y_r - m_r \hat{p}_r}{v_r^{1/2}} \right) \\ &= x_{rs} v_r^{1/2} \chi_r, \end{aligned}$$

onde $\chi_i = \frac{y_i - m_i \hat{p}_i}{v_i^{1/2}}$. Assim, temos que a matriz Δ pode ser escrita matricialmente por

$$\begin{aligned}\Delta &= \mathbf{X}^t \text{diag} \left\{ v_i^{1/2} \right\} \text{diag} \{ \chi_i \} \\ &= \mathbf{W}^t \mathbf{D} \{ \chi \}\end{aligned}\tag{2.60}$$

onde $\mathbf{W} = \mathbf{D} \{ v \}^{1/2} \mathbf{X}$, $D \{ v \}^{1/2} = \text{diag} \left\{ v_i^{1/2} \right\}$ e $\mathbf{D} \{ \chi \} = \text{diag} \{ \chi_i \}$, $i = 1, \dots, n$.

Derivando o resultado da equação (2.59) em relação a β_r , tem-se

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

Avaliando a derivada acima em $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}, \hat{p}_i)^t$ e $\boldsymbol{\omega}_0$, obtemos o elemento (r, s) da matriz $\ddot{\mathbf{L}}$, ou seja,

$$\begin{aligned}\ddot{L}_{rs} &= \left. \frac{\partial^2 l}{\partial \beta_s \partial \beta_r} \right|_{\hat{\boldsymbol{\theta}}, \boldsymbol{\omega} = \boldsymbol{\omega}_0} \\ &= \sum_{i=1}^n x_{is} v_i x_{ir} \\ &= \sum_{i=1}^n x_{is} v_i^{1/2} v_i^{1/2} x_{ir}.\end{aligned}$$

Dessa forma, a matriz $\ddot{\mathbf{L}}$ pode ser escrita como

$$\begin{aligned}\ddot{\mathbf{L}} &= \mathbf{X}^t \text{diag} \left\{ v_i^{1/2} \right\} \text{diag} \left\{ v_i^{1/2} \right\} \mathbf{X} \\ &= \mathbf{W}^t \mathbf{W}.\end{aligned}\tag{2.61}$$

Substituindo os resultados das expressões (2.60) e (2.61) na equação (2.19), tem-se

$$\begin{aligned}\ddot{\mathbf{F}} &= \mathbf{D} \{ \chi \} \mathbf{W} [\mathbf{W}^t \mathbf{W}]^{-1} \mathbf{W}^t \mathbf{D} \{ \chi \} \\ &= \mathbf{D} \{ \chi \} \mathbf{P}_{\mathbf{W}} \mathbf{D} \{ \chi \},\end{aligned}$$

onde $\mathbf{P}_{\mathbf{W}} = \mathbf{W} [\mathbf{W}^t \mathbf{W}]^{-1} \mathbf{W}^t$, ou seja, é a projeção ortogonal do espaço \mathbb{R}^n sobre as colunas da matriz \mathbf{W} . Segue que substituindo o resultado da matriz $\ddot{\mathbf{F}}$ acima na

equação (2.13), obtém-se a curvatura normal para avaliar a perturbação de observações individuais em regressão logística

$$C_{\ell} = 2 |\ell^t \mathbf{D} \{ \chi \} \mathbf{P}_w \mathbf{D} \{ \chi \} \ell|. \quad (2.62)$$

Capítulo 3

Avaliando a Influência Local Através da Inclinação Máxima de uma *lifted line*

3.1 Introdução

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

- (i) A escolha de um referencial.

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 da 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 da 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, através de um exemplo prático, uma dificuldade decorrente da falta de definição dos parâmetros.

Considere o modelo de regressão normal linear

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta} + \sigma_i \varepsilon_i, \quad i = 1, \dots, n$$

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

- Versão 1:

$$\sigma_i = \begin{cases} \sigma(1 + \omega) & , \quad i = 1 \\ \sigma & , \quad i \geq 2 \end{cases} \quad (3.1)$$

- Versão 2:

$$\sigma_i = \begin{cases} \sigma & , \quad i = 1 \\ \frac{\sigma}{1 - \omega} & , \quad i \geq 2 \end{cases} \quad (3.2)$$

Quando σ não é conhecido os modelos (3.1) e (3.2) são idênticos, contudo Billor e Loynes (1993) constataram a ocorrência de certa diferença no cálculo da curvatura máxima para as duas versões apresentadas acima.

3.2 Medida de Afastamento Pela Verossimilhança Modificada

A medida apresentada por Tsai (1986) e usada por Billor e Loynes (1993) para medir a influência local, que definiremos mais adiante, possui de imediato a vantagem

de estar de acordo com a falta de definição dos parâmetros, discutido no início deste capítulo.

Considere a medida afastamento pela verossimilhança modificada, definida por

$$LD^*(\boldsymbol{\omega}) = -2 \left[l(\widehat{\boldsymbol{\theta}}) - l(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\boldsymbol{\omega}) \right], \quad (3.3)$$

onde $\widehat{\boldsymbol{\theta}}$ e $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ são os estimadores de máxima verossimilhança do vetor de parâmetros desconhecidos $\boldsymbol{\theta}$, sob os modelos original e perturbado, respectivamente.

Note que, com a medida LD^* , o problema da escolha dos parâmetros desaparece, visto que $l(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}|\boldsymbol{\omega})$ 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, pode ser melhor entendida como um outro tipo de operação.

Para lidar com essa dificuldade, Billor e Loynes (1993) usam o fato de que as estimativas dos parâmetros $\widehat{\boldsymbol{\beta}}_{(i)}$ e $\widehat{\sigma}_{(i)}^2$, 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 (1.7), 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 (2.1). 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 $\widehat{\boldsymbol{\beta}}$ e $\widehat{\boldsymbol{\beta}}_{(i)}$, onde $\widehat{\boldsymbol{\beta}}$ é o estimador de máxima verossimilhança de $\boldsymbol{\beta}$.

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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{d}_i\phi + \boldsymbol{\varepsilon} \quad (3.4)$$

onde \mathbf{d}_i é um vetor $n \times 1$, com o i -ésimo elemento igual a 1 e os restantes iguais a zero,

defina

$$LD_i^* = -2 \left[l(\hat{\boldsymbol{\beta}}) - l_{ms}(\hat{\boldsymbol{\beta}}_{(i)}, \phi) \right]. \quad (3.5)$$

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

$$l_{ms}(\boldsymbol{\beta}, \phi) = -\frac{n}{2} \log \sigma^2 - \frac{n}{2} \log (2\pi)^{1/2} - \frac{1}{2\sigma^2} (y_i - \mathbf{x}_i^t \boldsymbol{\beta} - \phi)^2 - \frac{1}{2\sigma^2} \sum_{j \neq i} (y_j - \mathbf{x}_j^t \boldsymbol{\beta})^2. \quad (3.6)$$

Derivando l_{ms} em relação ao vetor $\boldsymbol{\beta}$, tem-se

$$\frac{\partial l_{ms}}{\partial \boldsymbol{\beta}} = -\frac{1}{\sigma^2} (y_i - \mathbf{x}_i^t \boldsymbol{\beta} - \phi) (-\mathbf{x}_i) - \frac{1}{\sigma^2} \sum_{j \neq i} (y_j - \mathbf{x}_j^t \boldsymbol{\beta}) (-\mathbf{x}_j).$$

Igualando a zero a derivada acima, temos que

$$\begin{aligned} y_i \mathbf{x}_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}}_{ms} \mathbf{x}_i - \phi \mathbf{x}_i &= - \sum_{j \neq i} y_j \mathbf{x}_j + \sum_{j \neq i} \mathbf{x}_j^t \hat{\boldsymbol{\beta}}_{ms} \mathbf{x}_j \Rightarrow \\ \mathbf{x}_i y_i + \sum_{j \neq i} \mathbf{x}_j y_j - \phi \mathbf{x}_i &= (\mathbf{x}_i \mathbf{x}_i^t) \hat{\boldsymbol{\beta}}_{ms} + \sum_{j \neq i} (\mathbf{x}_j \mathbf{x}_j^t) \hat{\boldsymbol{\beta}}_{ms} \Rightarrow \\ \mathbf{X}^t \mathbf{y} - \phi \mathbf{x}_i &= \left[\mathbf{x}_i \mathbf{x}_i^t + \sum_{j \neq i} \mathbf{x}_j \mathbf{x}_j^t \right] \hat{\boldsymbol{\beta}}_{ms} \Rightarrow \\ \mathbf{X}^t \mathbf{y} - \phi \mathbf{x}_i &= (\mathbf{X}^t \mathbf{X}) \hat{\boldsymbol{\beta}}_{ms} \Rightarrow \\ \hat{\boldsymbol{\beta}}_{ms} &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} - (\mathbf{X}^t \mathbf{X})^{-1} \phi \mathbf{x}_i. \end{aligned}$$

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

$$\hat{\boldsymbol{\beta}}_{ms} = \hat{\boldsymbol{\beta}} - \phi (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i. \quad (3.7)$$

Agora vamos obter o estimador de máxima verossimilhança do parâmetro ϕ .

Segue que, derivando a equação (3.6) em relação a ϕ , tem-se

$$\frac{\partial l_{ms}}{\partial \phi} = -\frac{1}{\sigma^2} (y_i - \mathbf{x}_i^t \boldsymbol{\beta} - \phi) (-1),$$

igualando a zero o resultado acima, obtemos que

$$\hat{\phi} = y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}}_{ms}.$$

Substituindo o valor de $\widehat{\boldsymbol{\beta}}_{ms}$ dado na equação (3.7) no resultado acima, temos que

$$\begin{aligned}\widehat{\phi} &= y_i - \mathbf{x}_i^t \left[\widehat{\boldsymbol{\beta}} - \widehat{\phi} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \right] \\ &= y_i - \mathbf{x}_i^t \widehat{\boldsymbol{\beta}} + \widehat{\phi} \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \\ &= y_i - \widehat{\mu}_i + \widehat{\phi} h_{ii},\end{aligned}$$

onde $\widehat{\mu}_i = \mathbf{x}_i^t \widehat{\boldsymbol{\beta}}$ é o valor predito para y_i da regressão de \mathbf{y} sobre \mathbf{X} e $h_{ii} = \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i$ é o elemento (i, i) da matriz *hat*. Podemos ainda escrever

$$\begin{aligned}\widehat{\phi} &= r_i + \widehat{\phi} h_{ii} \\ &= \frac{r_i}{1 - h_{ii}},\end{aligned}\tag{3.8}$$

onde $r_i = y_i - \widehat{\mu}_i$ é o i -ésimo resíduo ordinário da regressão de \mathbf{y} sobre \mathbf{X} .

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

$$\widehat{\boldsymbol{\beta}}_{(i)} = (\mathbf{X}_{(i)}^t \mathbf{X}_{(i)})^{-1} \mathbf{X}_{(i)}^t \mathbf{y}_{(i)},\tag{3.9}$$

onde $\mathbf{X}_{(i)}$ 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 o resultado do teorema (A.2.4), descrito no Apêndice (A), ao segundo membro da equação acima, temos que (ver Cook e Weisberg (1982), pag. 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 o resultado acima na equação (3.9), tem-se

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

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

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{(i)} &= (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} + (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \left[\frac{\mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}_{(i)}^t \mathbf{y}_{(i)} - y_i + \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i y_i}{1 - h_{ii}} \right] \\
&= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \left[\mathbf{X}_{(i)}^t \mathbf{y}_{(i)} + \mathbf{x}_i y_i \right] - (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i y_i}{1 - h_{ii}} \\
&= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} - (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i y_i}{1 - h_{ii}} \\
&= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \mathbf{x}_i^t \widehat{\boldsymbol{\beta}} - (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i y_i}{1 - h_{ii}} \\
&= \widehat{\boldsymbol{\beta}} + \frac{(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i [\widehat{\mu}_i - y_i]}{1 - h_{ii}} \\
&= \widehat{\boldsymbol{\beta}} - \frac{r_i}{1 - h_{ii}} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i. \tag{3.10}
\end{aligned}$$

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

$$l(\boldsymbol{\beta}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2.$$

Segue daí que,

$$\begin{aligned}
l(\widehat{\boldsymbol{\beta}}) &= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i^t \widehat{\boldsymbol{\beta}})^2 \\
&= -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \widehat{\mu}_i)^2 \\
&= -\frac{1}{2\sigma^2} \sum_{i=1}^n r_i^2. \tag{3.11}
\end{aligned}$$

Do mesmo modo, vamos considerar agora apenas a parte relevante da log-verossimilhança do modelo deslocamento de média em (3.6). Temos que,

$$l_{ms}(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\phi}) = -\frac{1}{2\sigma^2} (y_i - \mathbf{x}_i^t \widehat{\boldsymbol{\beta}}_{(i)} - \widehat{\phi})^2 - \frac{1}{2\sigma^2} \sum_{j \neq i} (y_j - \mathbf{x}_j^t \widehat{\boldsymbol{\beta}}_{(i)})^2$$

substituindo os resultados em (3.8) e (3.10) na equação acima, tem-se

$$\begin{aligned}
l_{ms}(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\phi}) &= -\frac{1}{2\sigma^2} \left\{ y_i - \mathbf{x}_i^t \left[\widehat{\boldsymbol{\beta}} - \frac{r_i}{1-h_{ii}} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \right] - \frac{r_i}{1-h_{ii}} \right\}^2 \\
&\quad - \frac{1}{2\sigma^2} \sum_{j \neq i} \left\{ y_j - \mathbf{x}_j^t \left[\widehat{\boldsymbol{\beta}} - \frac{r_i}{1-h_{ii}} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \right] \right\}^2 \\
&= -\frac{1}{2\sigma^2} \left\{ y_i - \widehat{\mu}_i + \frac{r_i}{1-h_{ii}} \mathbf{x}_i^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i - \frac{r_i}{1-h_{ii}} \right\}^2 \\
&\quad - \frac{1}{2\sigma^2} \sum_{j \neq i} \left\{ y_j - \widehat{\mu}_j + \frac{r_i}{1-h_{ii}} \mathbf{x}_j^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i \right\}^2 \\
&= -\frac{1}{2\sigma^2} \left\{ r_i + \frac{r_i}{1-h_{ii}} h_{ii} - \frac{r_i}{1-h_{ii}} \right\}^2 - \frac{1}{2\sigma^2} \sum_{j \neq i} \left\{ r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right\}^2
\end{aligned}$$

onde $h_{ij} = \mathbf{x}_j^t (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{x}_i$ é o elemento (i, j) da matriz *hat*. Segue que,

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

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

$$\begin{aligned}
l_{ms}(\widehat{\boldsymbol{\beta}}_{(i)}, \widehat{\phi}) &= -\frac{1}{2\sigma^2} \left\{ \left[\sum_{j \neq i} \left(r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right)^2 + \left(r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right)^2 \right] - \left(r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right)^2 \right\} \\
&= -\frac{1}{2\sigma^2} \sum_{j=1}^n \left(r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right)^2 + \frac{1}{2\sigma^2} \left(r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right)^2. \tag{3.12}
\end{aligned}$$

Substituindo os resultados em (3.11) e (3.12) na equação (3.5), segue que

$$\begin{aligned}
LD_i^* &= -2 \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n r_i^2 - \left[-\frac{1}{2\sigma^2} \sum_{j=1}^n \left(r_j + \frac{r_i}{1-h_{ii}} h_{ij} \right)^2 + \frac{1}{2\sigma^2} \left(r_i + \frac{r_i}{1-h_{ii}} h_{ii} \right)^2 \right] \right\} \\
&= -2 \left(-\frac{1}{2\sigma^2} \right) \left\{ \sum_{i=1}^n r_i^2 - \sum_{j=1}^n \left(r_j^2 + 2r_j \frac{r_i}{1-h_{ii}} h_{ij} + \frac{r_i^2}{(1-h_{ii})^2} h_{ij}^2 \right) + \frac{r_i^2}{(1-h_{ii})^2} \right\} \\
&= \frac{1}{\sigma^2} \left\{ -2 \frac{r_i}{1-h_{ii}} \sum_{j=1}^n h_{ij} r_j + \frac{r_i^2}{(1-h_{ii})^2} \sum_{j=1}^n h_{ij}^2 + \frac{r_i^2}{(1-h_{ii})^2} \right\}. \tag{3.13}
\end{aligned}$$

Observe que

$$\begin{aligned}
\mathbf{Hr} &= \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y} - \mathbf{X} (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{X}\boldsymbol{\beta} - \mathbf{X}\boldsymbol{\beta} \\
&= \mathbf{0}.
\end{aligned}$$

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

$$\sum_{j=1}^n h_{ij}r_j = 0, \forall i. \quad (3.14)$$

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

$$\begin{aligned} \sum_{j=1}^n h_{ij}^2 &= \sum_{j=1}^n h_{ij}h_{ij} \\ &= \sum_{j=1}^n h_{ij}h_{ji} \\ &= h_{ii}. \end{aligned} \quad (3.15)$$

Substituindo os resultados em (3.14) e (3.15) na equação (3.13), tem-se

$$\begin{aligned} LD_i^* &= \frac{1}{\sigma^2} \left(0 - \frac{r_i^2}{(1-h_{ii})^2} h_{ii} + \frac{r_i^2}{(1-h_{ii})^2} \right) \\ &= \frac{1}{\sigma^2} \left\{ \frac{r_i^2(1-h_{ii})}{(1-h_{ii})^2} \right\} \\ &= \frac{r_i^2}{\sigma^2(1-h_{ii})}. \end{aligned} \quad (3.16)$$

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 vimos que uma motivação para a influência local de Cook (1986) é a relação em (2.6), 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:

$$pD_i^*(\boldsymbol{\omega}) = -2 \left[l(\hat{\boldsymbol{\beta}}) - l(\hat{\boldsymbol{\beta}}_{\boldsymbol{\omega}}|\boldsymbol{\omega}) \right]. \quad (3.17)$$

Através de um raciocínio semelhante ao utilizado por Cook (1986) para estudar a influência local, temos que $(\boldsymbol{\omega}, LD^*(\boldsymbol{\omega}))$ 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 \mathbb{R}^q definida por $\boldsymbol{\omega}(a) = \boldsymbol{\omega}_0 + a\boldsymbol{\ell}$, onde $\boldsymbol{\omega}_0$ representa um vetor de não-perturbação do modelo, $\boldsymbol{\ell}$ é um vetor fixo em \mathbb{R}^q e a é um escalar, segue que o comportamento de $LD^*(\boldsymbol{\omega}_0 + a\boldsymbol{\ell})$ proporciona informações importantes sobre a influência do esquema de perturbação adotado.

Como a primeira derivada de LD^* avaliada em $\boldsymbol{\omega}_0$, não se anula, exceto para alguns valores particulares do vetor $\boldsymbol{\ell}$, a primeira derivada de LD^* em relação a $\boldsymbol{\ell}$,

produz informações valiosas sobre o comportamento local de $LD^*(\boldsymbol{\omega}_0)$. Em particular, a inclinação máxima e a correspondente direção máxima, denotadas respectivamente por d_{max}^* e $\boldsymbol{\ell}_{max}^*$ são importantes quando $q \neq 1$, onde q é a dimensão do vetor $\boldsymbol{\omega}$, Billor e Loynes (1993).

A direção máxima da função LD^* no ponto $\boldsymbol{\omega}_0$ ocorre na direção do gradiente $\nabla LD^*(\boldsymbol{\omega}_0)$. Logo, a inclinação máxima (d_{max}^*) é medida por (ver Lima (1981))

$$d_{max}^* = \|\nabla LD^*(\boldsymbol{\omega}_0)\|.$$

Pela regra da cadeia,

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

De imediato, temos que $\frac{\partial l(\hat{\boldsymbol{\beta}})}{\partial \boldsymbol{\omega}} = 0$ e que $\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\omega})}{\partial \boldsymbol{\beta}}$, quando avaliada em $\hat{\boldsymbol{\beta}}$ e $\boldsymbol{\omega}_0$, também se anula. Assim, a equação acima fica resumida a

$$\left. \frac{\partial LD^*}{\partial a} \right|_{\hat{\boldsymbol{\omega}}=\boldsymbol{\omega}_0} = 2\boldsymbol{\ell}^t \left. \frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\omega})}{\partial \boldsymbol{\omega}} \right|_{\hat{\boldsymbol{\omega}}=\boldsymbol{\omega}_0}.$$

Portanto, a inclinação máxima pode ser escrita como

$$d_{max}^* = 2 \left\| \nabla l(\hat{\boldsymbol{\beta}}|\boldsymbol{\omega}) \right\|. \quad (3.18)$$

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

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

onde $Var(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{W}^{-1}$ e $\mathbf{W} = diag(1 + w_1, 1, 1, \dots, 1)$, com o modelo não perturbado sendo obtido quando $w_1 = 0$. A inclinação da linha projetada sobre a superfície $(\boldsymbol{\omega}, LD^*(\boldsymbol{\omega}_0))$ é calculada através da primeira derivada de $LD^*(\boldsymbol{\omega})$ em relação a w_1 . A log-verossimilhança para o modelo acima fica estabelecida por

$$l(\boldsymbol{\beta}, \boldsymbol{\omega}) = -\frac{n}{2} \ln \sigma^2 - \frac{n}{2} \ln 2\pi + \frac{1}{2} \ln(1 + w_1) - \frac{(1 + w_1)(y_1 - \mathbf{x}_1^t \boldsymbol{\beta})^2}{2\sigma^2} - \frac{1}{2\sigma^2} \sum_{i=2}^n (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2,$$

segue daí que

$$\begin{aligned}\frac{\partial LD^*}{\partial w_1} &= 2 \frac{\partial l(\boldsymbol{\beta}, \boldsymbol{\omega})}{\partial w_1} \\ &= 2 \left[\frac{1}{2(1+w_1)} - \frac{(y_1 - \mathbf{x}_1^t \boldsymbol{\beta})^2}{2\sigma^2} \right].\end{aligned}$$

Assim,

$$\begin{aligned}\left. \frac{\partial LD^*}{\partial w_1} \right|_{\hat{\boldsymbol{\omega}}, w_1=0} &= 1 - \frac{(y_1 - \mathbf{x}_1^t \hat{\boldsymbol{\beta}})^2}{\sigma^2} \\ &= 1 - \frac{r_1^2}{\sigma^2},\end{aligned}\tag{3.19}$$

onde $r_1 = y_1 - \mathbf{x}_1^t \hat{\boldsymbol{\beta}}$ é o resíduo ordinário relativo à primeira observação da regressão de \mathbf{y} sobre \mathbf{X} . Temos as seguintes considerações para o resultado do resíduo r_1 acima,

- se $|r_1| = \sigma$, não existe razão para preferir o modelo perturbado ao não perturbado e vice-versa,
- se $|r_1| \ll \sigma$ devemos optar pelo modelo perturbado caso $w_1 < 0$, isto é, se $Var(y_1) < \sigma^2$.

Observe ainda, que $1 - \frac{r_1^2}{\sigma^2}$ possui uma cota superior finita quando $|r_1| \rightarrow 0$, enquanto sua cota inferior, quando $|r_1| \rightarrow +\infty$ é $-\infty$, essa assimetria esta inteiramente de acordo com a teoria vista, Billor e Loynes (1993).

Agora vamos considerar a matriz de perturbação geral dada por

$$\mathbf{W}_g = \text{diag}(1 + w_1, 1 + w_2, \dots, 1 + w_n),$$

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

$$\nabla LD^*(\boldsymbol{\omega}_0) = \left(\frac{\partial LD^*}{\partial w_1}, \frac{\partial LD^*}{\partial w_2}, \dots, \frac{\partial LD^*}{\partial w_n} \right),$$

usando um procedimento análogo ao realizado para obter o resultado em (3.19), temos que

$$\left. \frac{\partial LD^*(\boldsymbol{\omega})}{\partial w_i} \right|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = 1 - \frac{r_i^2}{\sigma^2}$$

quando $w_i = 0$, com $i = 1, \dots, n$. Dessa forma, a razão máxima de crescimento de LD^* é na direção do vetor gradiente $\nabla LD^*(\boldsymbol{\omega}_0)$, ou seja,

$$\begin{aligned} d_{max}^* &= \|\nabla LD^*(\boldsymbol{\omega}_0)\| \\ &= \sqrt{\left(\frac{\partial LD^*}{\partial \omega_1}\right)^2 + \left(\frac{\partial LD^*}{\partial \omega_2}\right)^2 + \dots + \left(\frac{\partial LD^*}{\partial \omega_n}\right)^2} \\ &= \sqrt{\sum_{j=1}^n \left(1 - \frac{r_j^2}{\sigma^2}\right)^2}. \end{aligned} \quad (3.20)$$

O mesmo desenvolvimento adotado nesta seção, pode ser aplicado quando estamos interessados em avaliar o comportamento de um gráfico de influência em um ponto $\boldsymbol{\omega}^*$ diferente de $\boldsymbol{\omega}_0$, visto que, para isso, faz-se necessário 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 $\boldsymbol{\omega}^*$.

3.3 Avaliando a Inclinação Máxima d_{max}^*

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 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 uma tal referência. Por exemplo, vamos considerar o modelo de regressão normal linear perturbado pela matriz $\mathbf{W}_g = \text{diag}(1 + \omega_1, 1 + \omega_2, \dots, 1 + \omega_n)$, já mencionado anteriormente. Já mostramos que para esse problema temos

$$d_{max}^* = A^{\frac{1}{2}}$$

onde $A = \sum_{j=1}^n \left(1 - \frac{r_j^2}{\sigma^2}\right)^2$. A expressão acima 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 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 , podemos obter a esperança e a variância de A , como

segue. Temos que

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

Note que,

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

onde $\chi_{(1)}^2$ representa uma distribuição qui-quadrado com um grau de liberdade. Como $E(\chi_{(1)}^2) = 1$, segue que

$$\begin{aligned} E(A) &= E \left\{ \sum_{j=1}^n \left[\left(\frac{\varepsilon_j}{\sigma} \right)^2 - E \left(\frac{\varepsilon_j}{\sigma} \right)^2 \right]^2 \right\} \\ &= \sum_{j=1}^n E \left[\left(\frac{\varepsilon_j}{\sigma} \right)^2 - E \left(\frac{\varepsilon_j}{\sigma} \right)^2 \right]^2 \\ &= \sum_{j=1}^n \text{Var} \left(\frac{\varepsilon_j}{\sigma} \right)^2 \\ &= 2n, \end{aligned} \tag{3.21}$$

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

Pela proposição (A.3.3) do Apêndice (A), temos

$$E[A^2] = 56n + 4n^2. \tag{3.22}$$

Substituindo os resultados em (3.21) e (3.22) na fórmula da variância a seguir, temos que

$$\begin{aligned} \text{Var}[A] &= E[A^2] - (E[A])^2 \\ &= 56n + 4n^2 - (2n)^2 \\ &= 56n. \end{aligned} \tag{3.23}$$

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

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

Para uma melhor aproximação, podemos 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.

3.4 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 (2.35), temos que

$$l(\boldsymbol{\beta}|\boldsymbol{\omega}) = \sum_{i=1}^n \frac{1}{2\sigma^2} \omega_i (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2, \quad (3.24)$$

onde ω_i e y_i são as i -ésimas componentes dos vetores $\boldsymbol{\omega}$ e \mathbf{y} , respectivamente.

Derivando a função (3.24) em relação a ω_i , temos que

$$\frac{\partial l}{\partial \omega_i} = \frac{1}{2\sigma^2} (y_i - \mathbf{x}_i^t \boldsymbol{\beta})^2$$

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

$$\begin{aligned} \left. \frac{\partial l}{\partial \omega_i} \right|_{=\hat{\cdot}, \sigma^2=\hat{\sigma}^2} &= \frac{1}{2\hat{\sigma}^2} (y_i - \mathbf{x}_i^t \hat{\boldsymbol{\beta}})^2 \\ &= \frac{1}{2\hat{\sigma}^2} (y_i - \hat{\mu}_i)^2 \\ &= \frac{1}{2\hat{\sigma}^2} r_i^2. \end{aligned}$$

Da seção (3.2) temos que

$$\begin{aligned} \frac{\partial LD^*}{\partial \omega_i} &= 2 \left. \frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\omega})}{\partial \omega_i} \right|_{=\hat{\cdot}, \sigma^2=\hat{\sigma}^2} \\ &= 2 \left\{ \frac{1}{2\hat{\sigma}^2} r_i^2 \right\} \\ &= \frac{r_i^2}{\hat{\sigma}^2}. \end{aligned} \quad (3.25)$$

Vimos na seção (3.2) que a direção da inclinação máxima $\boldsymbol{\ell}_{max}^*$ é a direção do gradiente, assim podemos escrever

$$\boldsymbol{\ell}_{max}^* = \nabla LD^*(\boldsymbol{\omega}_0) = \left(\frac{r_1^2}{\hat{\sigma}^2}, \dots, \frac{r_n^2}{\hat{\sigma}^2} \right). \quad (3.26)$$

Segue que a inclinação máxima é obtida por

$$\begin{aligned} d_{max}^* &= \|\nabla LD^*(\boldsymbol{\omega}_0)\| \\ &= \sqrt{\sum_{j=1}^n \left(\frac{r_j^2}{\hat{\sigma}^2} \right)^2} \end{aligned} \quad (3.27)$$

Para obter um valor de referência para a inclinação máxima acima podemos desenvolver um raciocínio análogo ao realizado na seção (3.3). Seja $A = \sum_{j=1}^n \frac{r_j^2}{\hat{\sigma}^2}$, substituindo em A os $r_{j's}$ pelos erros $\varepsilon_{j's}$ que são assumidos variáveis aleatórias independentes normalmente distribuídas com média zero e variância σ^2 . Segue que

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

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

$$\begin{aligned} E[A] &= \sum_{j=1}^n E[x_j] \\ &= \sum_{j=1}^n 1 \\ &= n. \end{aligned} \tag{3.28}$$

Observe que

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

Logo

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

pela independência entre x_i e x_j , para todo $i \neq j$, temos que

$$E[A^2] = \sum_{j=1}^n E[x_j^2] + \sum_{j \neq i} E[x_i]E[x_j]. \tag{3.29}$$

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

$$M_{x_j}''(t) = 3(1 - 2t)^{-5/2} \Rightarrow M_{x_j}''(0) = E[x_j^2] = 3. \tag{3.30}$$

Substituindo o resultado acima em (3.29), obtém-se

$$\begin{aligned} E[A^2] &= \sum_{j=1}^n 3 + \sum_{j=1}^n (n-1) \\ &= 3n + n(n-1) \\ &= n^2 + 2n. \end{aligned} \tag{3.31}$$

Substituindo o resultado das equações (3.28) e (3.31) na fórmula da variância a seguir, obtém-se

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

Portanto, podemos considerar como um valor de referência para a inclinação máxima obtida em (3.27), o resultado de

$$E[A] + 2\sqrt{\text{Var}[A]} = n + 2\sqrt{2n}. \quad (3.32)$$

3.5 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 dada na subseção (2.6.1)

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

Derivando a função acima em relação a ω_i , obtemos

$$\frac{\partial l}{\partial \omega_i} = y_i \ln \left[\frac{p_i}{1-p_i} \right] + m_i \ln [1-p_i]$$

avaliando a expressão acima em \hat{p}_i , temos

$$\left. \frac{\partial l}{\partial \omega_i} \right|_{p_i=\hat{p}_i} = y_i \ln \left[\frac{\hat{p}_i}{1-\hat{p}_i} \right] + m_i \ln [1-\hat{p}_i].$$

Dessa forma as coordenadas do vetor gradiente $\boldsymbol{\ell}_{max}^* = \nabla LD^*(\boldsymbol{\omega}_0)$, são dadas por

$$\begin{aligned} \boldsymbol{\ell}_i^* &= \frac{\partial LD^*}{\partial \omega_i} \\ &= 2 \left. \frac{\partial l(\boldsymbol{\beta}|\boldsymbol{\omega})}{\partial \omega_i} \right|_{\hat{p}_i, \sigma^2=\hat{\sigma}^2} \\ &= 2 \left\{ y_i \ln \left[\frac{\hat{p}_i}{1-\hat{p}_i} \right] + m_i \ln [1-\hat{p}_i] \right\} \end{aligned} \quad (3.33)$$

A inclinação máxima é obtida pela aplicação imediata de

$$d_{max}^* = \|\nabla LD^*(\boldsymbol{\omega}_0)\|.$$

Capítulo 4

Aplicações

Neste capítulo são discutidos três exemplos de análise de dados reais via o pacote estatístico **R**. O objetivo principal destas aplicações é colocar em prática as técnicas de diagnóstico, com ênfase para a influência local de Cook e a abordagem de Billor e Loynes, descritas nesta dissertação.

Utilizamos também algumas **técnicas de 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 (4.1)) usamos o modelo de regressão normal linear simples para os dados propostos. A análise foi desenvolvida de quatro formas: técnicas gráficas de diagnóstico, curvatura normal de Cook para regressão normal linear, sob perturbação de observação, inclinação máxima proposta por Billor e Loynes, sob perturbação na variância e sob perturbação de observação.

No segundo e terceiro exemplos (seções (4.2) e (4.3), respectivamente) usamos o modelo de regressão logística múltipla para os dados apresentados. Em cada exemplo analisamos o modelo proposto sob três aspectos: técnicas gráficas de diagnóstico, curvatura normal de Cook para regressão logística e a inclinação máxima de Billor e Loynes, ambos sob esquema de perturbação de observação.

4.1 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 sobrevôa a região e quando avista um bando de gansos, uma pessoa experiente estima a quantidade dessas aves. Para investigar a eficiência deste método um experimento foi conduzido, onde uma aeronave carregando dois observadores sobrevôu 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 através desta fotografia.

Os dados coletados neste experimento encontram-se em Weisberg (1985, pg.102). Aplicamos 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 1 do experimento. O modelo fica portanto dado por

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, 45, \quad (4.1)$$

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 \mathbf{R} , foram $\hat{\alpha} = 26,650$ e $\hat{\beta} = 0,883$, com respectivos desvios 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, através do teste F, mostrou-se significativa.

O **diagrama de dispersão** (*plot* de X versus Y) mostrado na Fig.4.1, sugere forte evidência de heterocedasticidade. Assim, podemos esperar que a introdução de uma pequena perturbação revele alguma espécie de sensibilidade no modelo.

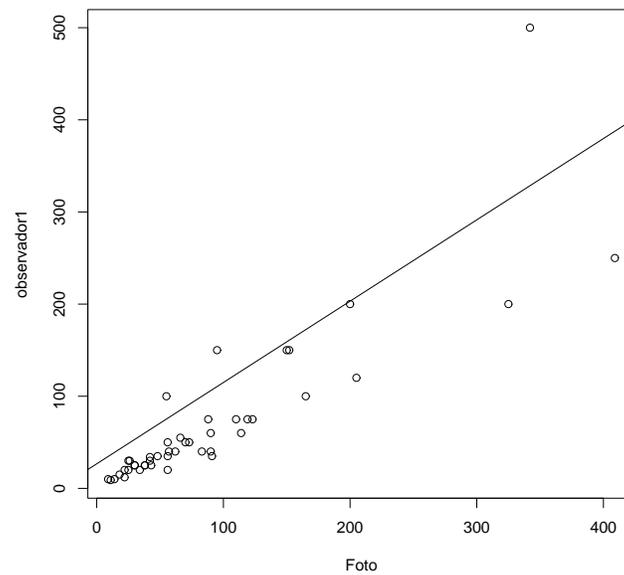


Figura 4.1: Número de gansos contados a partir da foto versus o número de gansos estimado pelo observador 1

A seguir faremos uma análise de alguns gráficos de diagnósticos, com o objetivo de identificar pontos influentes e de verificar a validade das suposições adotadas para o modelo.

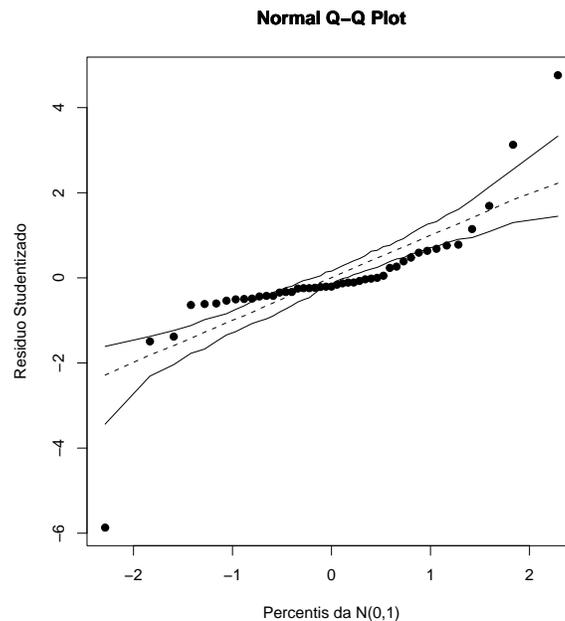


Figura 4.2: Gráfico de probabilidades com envelopes

Na Fig.4.2 temos o gráfico normal de probabilidade com envelopes para os resíduos t_{s_i} , que avalia a validade da suposição de normalidade da variável resposta, para o modelo adotado em (4.1), pode-se notar através 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.

Um outro gráfico importante a ser interpretado é o gráfico dos valores \hat{h}_{ii} em função das ordens das observações dado na Fig.4.3, através dele destacamos dois pontos acima da linha $\frac{2*p}{n}$, que correspondem às observações 28 e 29, sendo portanto candidatos a pontos influentes.

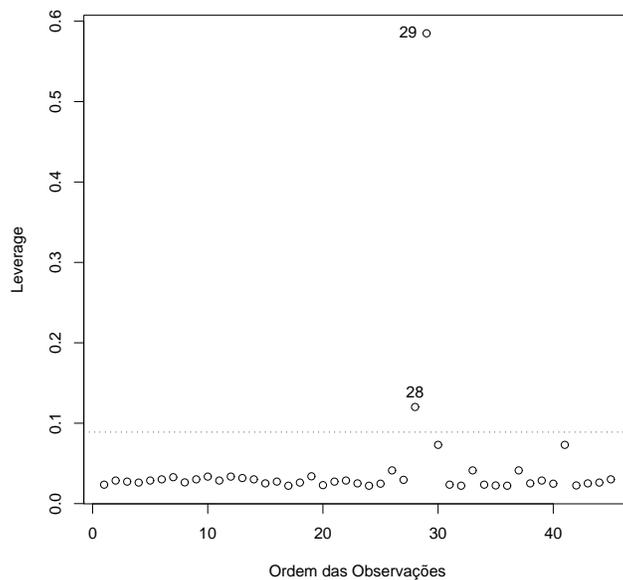


Figura 4.3: Leverage \hat{h}_{ii} contra a ordem das observações

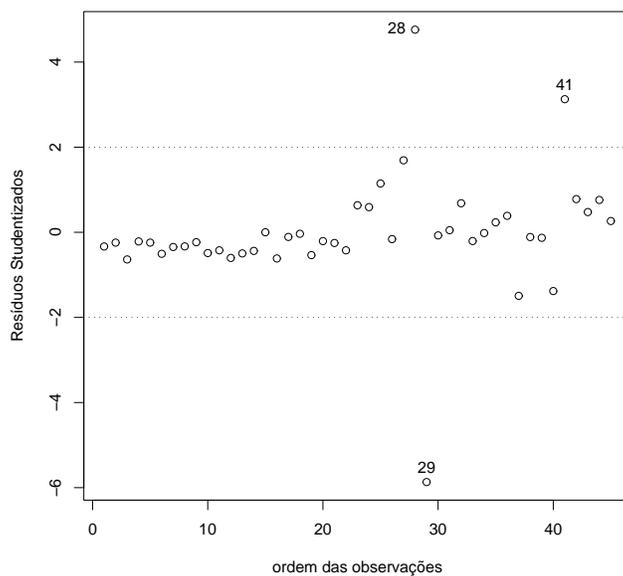


Figura 4.4: Resíduos *studentizados* (t_{S_i}) contra a ordem das observações

O gráfico dos resíduos *studentizados* em função das ordens das observações, dado

na Fig.4.4, destaca novamente os pontos 28 e 29, além do ponto 41, sendo estes candidatos a serem pontos aberrantes.

Na Fig.4.5 temos o gráfico dos resíduos *studentizados* em função dos valores ajustados, verificamos novamente em destaque os pontos 28, 29 e 41, o gráfico renova a indicação de heterocedasticidade já percebida pelo diagrama de dispersão (Fig.4.1).

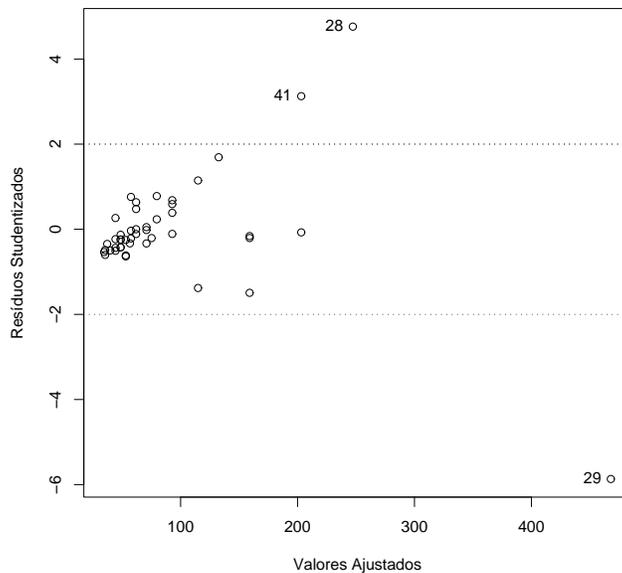


Figura 4.5: Resíduos *studentizados* t_{S_i} contra os valores ajustados

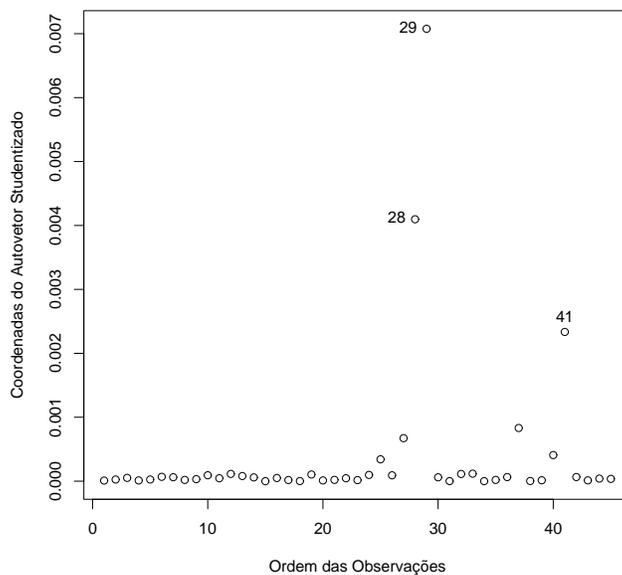


Figura 4.6: Coordenadas do autovetor d_{max} contra a ordem das observações

Os mesmos pontos já destacados anteriormente (28,29 e 41) também aparecem em destaque no gráfico da Fig.4.6, 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 através da **curvatura normal** de Cook para modelo normal, dada em (2.45). Para introduzir a perturbação nas observações consideramos o vetor ℓ 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 17.767, o que indica extrema sensibilidade local.

Usando o esquema de perturbação de observação descrito na seção (3.3), aplicamos o resultado da inclinação máxima (d_{max}^*) dada na equação (3.28), com isso obtemos o resultado 16,573, considerando como referência o valor sugerido pela fórmula em (3.33) que é de 7,381, temos aqui forte indicação de sensibilidade local.

Para o esquema de perturbação da matriz de variância abordado no seção (3.1), o valor da inclinação máxima é de 17,766, obtido através da fórmula dada em (3.20). Usando o valor de referência para d_{max}^* , sugerido na seção (3.2), que é de 13,798, também aqui detectamos forte sensibilidade local.

A abordagem de Billor e Loynes permite que se tenha os valores individuais das coordenadas ($\ell_{j's}^*$) do gradiente $\ell_{max}^* = \nabla LD^*(\omega_0)$, que é o vetor que indica a direção onde se obtém a inclinação máxima. Assim podemos avaliar quais observações contribuem com a maior parcela para o valor de d_{max}^* . Os $\ell_{j's}^*$ individuais, obtidos pela perturbação da matriz de variância, são apresentados na tabela (4.1), na qual nota-se que as observações que contribuem com os maiores valores são as de números 28,29 e 41.

Tabela 4.1: Valores individuais dos $\ell_{j,s}^*$ para os dados de gansos.

observação(i)	$ \ell_j^* $	observação(i)	$ \ell_j^* $	observação(i)	$ \ell_j^* $
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

4.2 Distúrbio Coronário versus Pressão Sanguínea e Nível de Colesterol

Os dados utilizados nesta seção foram obtidos em Cordeiro (1986, pg.113), mostram a classificação de 1330 pacientes segundo a ocorrência ou não de distúrbio coronário pelos seguintes níveis de colesterol (NC_i):

NC_1 =menos de 200

NC_2 =de 200 a 219

NC_3 =de 220 a 260

NC_4 =mais de 260

e de pressão sanguínea(PS_i):

PS_1 =menos de 127

PS_2 =de 127 a 146

PS_3 =de 147 a 166

PS_4 =mais de 166.

Considerando como variável resposta Y =proporção de indivíduos que apresentaram distúrbio coronário e como variáveis explicativas NC e PS , ambas definidas como fator com quatro níveis cada uma, como descritos acima. Adotamos para os dados o modelo logístico linear múltiplo, com preditor linear dado por

$$\eta_i = \beta_1 + \sum_{j=1}^4 \beta_{2j} PS_j + \sum_{k=1}^4 \beta_{3k} NC_k, \quad i = 1, \dots, 16. \quad (4.2)$$

Assumindo que $\beta_{21} = \beta_{31} = 0$ e considerando erro Binomial e ligação logit, o desvio obtido para este modelo foi $D(\mathbf{y}, p) = 4,775$ (9 graus de liberdade), indicando um ajuste adequado, pois o p -value associado a 4,775 pelo teste qui-quadrado unicaudal é de 85,34%. O número de iterações do processo iterativo para estimar os coeficientes desse modelo foi quatro.

Na tabela (4.2) a seguir temos dispostas as estimativas dos parâmetros do modelo proposto, com seus respectivos desvios padrões, todos apresentaram-se significativos ao modelo.

Tabela 4.2: Estimativas dos coeficientes referentes ao modelo em 4.2 para explicar a proporção da ocorrência de distúrbio coronário.

Coeficiente	Estimativas	Desvio Padrão
Constante	-3,495	0,349
PS_2	-0,091	0,451
PS_3	0,562	0,351
PS_4	1,342	0,343
NC_2	-0,038	0,303
NC_3	0,587	0,328
NC_4	1,204	0,327

A Fig.4.7 nos mostra o gráfico de probabilidade com envelopes para os resíduos t_{D_i} , que avalia a validade da suposição de que a variável resposta segue uma Binomial(m_i, p_i), como podemos observar não há indícios de que esta suposição não seja válida para o modelo adotado em (4.2).

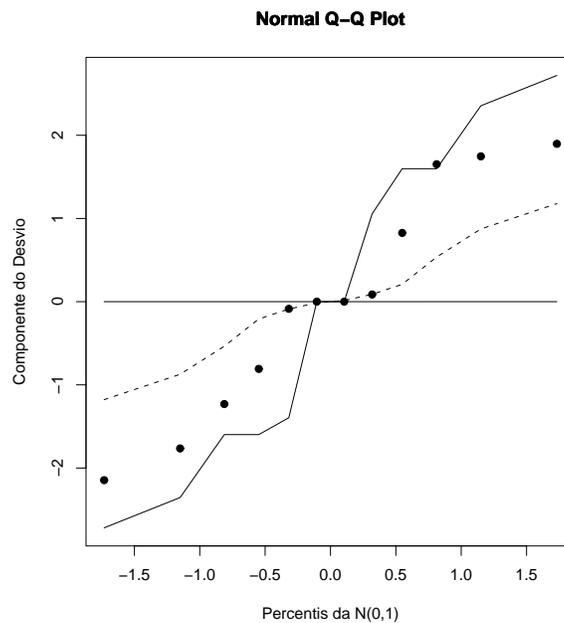


Figura 4.7: Gráfico de probabilidade com envelopes para o modelo binomial

A seguir faremos uma inspeção sobre alguns gráficos utilizados para detectar possíveis pontos candidatos a serem influentes. O gráfico da Fig.4.8 dos h_{ii} em função das ordens das observações não detecta nenhum ponto acima da linha $\frac{2*p}{n}$, assim não temos indicação de pontos aberrantes.

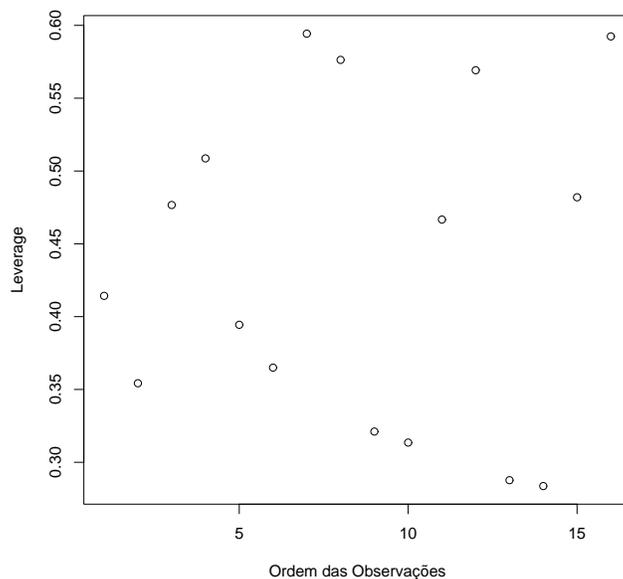


Figura 4.8: Leverage \hat{h}_{ii} contra a ordem das observações

No gráfico da Fig.4.9, dos \hat{h}_{ii} contra os valores ajustados os pontos que aparecem mais distantes são àqueles correspondentes às observações 12,15 e 16.

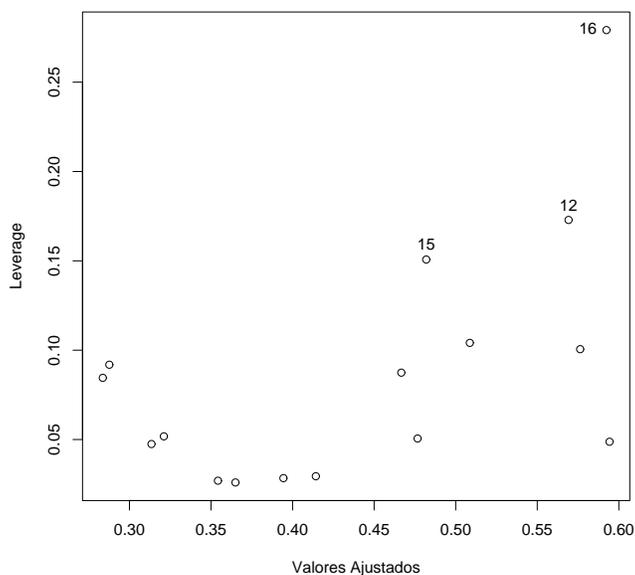


Figura 4.9: Leverage \hat{h}_{ii} contra valores ajustados

O gráfico dos resíduos t_{D_i} contra os valores ajustados, mostrado na Fig.4.10, destaca os pontos correspondentes às observações 1,10 e 13, mas nenhuma com indicação de ser ponto influente.

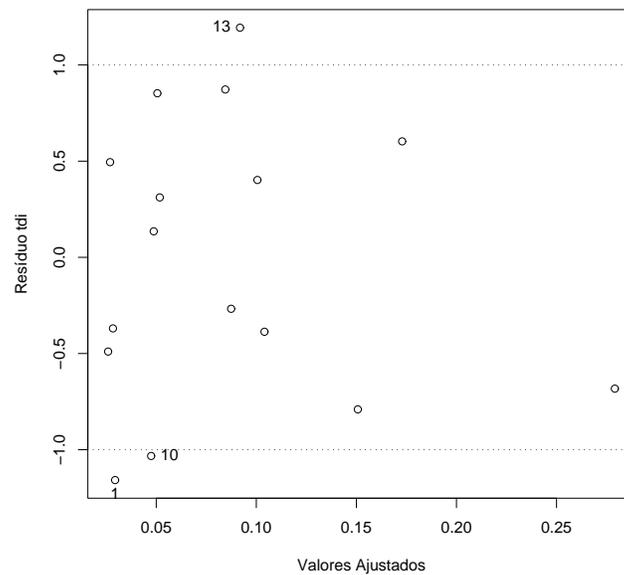


Figura 4.10: Resíduo t_{D_i} contra valores ajustados

Na Fig.4.11 temos o gráfico da medida afastamento pela verossimilhança, LD_i , em função da ordem das observações, nele é possível destacar os pontos correspondentes às observações 1,3 e 13, como mais afastados.

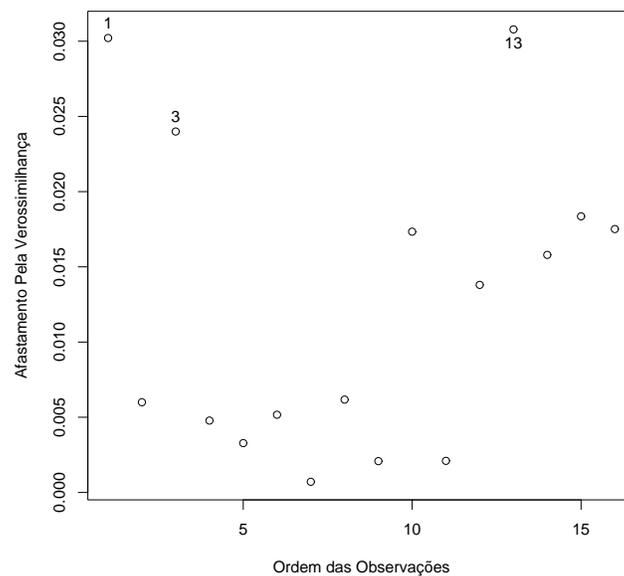


Figura 4.11: LD_i contra a ordem das Observações

No gráfico das coordenadas padronizadas do autovetor ℓ_{max} em função da ordem das observações, mostrado na Fig.4.12, podemos observar os pontos correspondentes às observações 1 e 13, como os mais afastados.

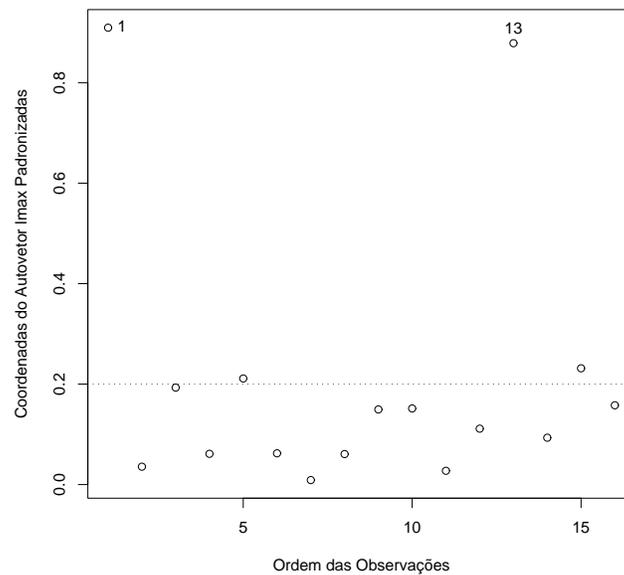


Figura 4.12: Coordenadas do autovetor ℓ_{max} contra a ordem das observações

Avaliando a situação dos pontos destacados pelos gráficos, não temos uma clara definição sobre quais apresentam-se mais discrepantes, podendo influir no resultado do ajuste. Para colocar em prática o cálculo da curvatura normal para perturbação de observação em modelo logístico, usaremos a fórmula dada na equação (2.62), escolhemos a observação 13 para sofrer uma pequena perturbação através do vetor ℓ , com 1 em todas as posições menos na de ordem 13, que recebe 0,9. Aplicando a referida fórmula obtemos o resultado 9,1315, a falta de um valor de referência para comparar com este valor, não permite uma interpretação definitiva a respeito da sensibilidade do modelo.

No cálculo da inclinação máxima (d_{max}^*) para o modelo logístico, usamos o resultado encontrado na seção (3.5), o valor encontrado foi de 89,0611. O gráfico da Fig4.13 mostra o *plot* das coordenadas padronizadas do gradiente ℓ_{max}^* em função da ordem das observações, observa-se que os pontos correspondentes às observações 7 e 8 aparecem mais afastados dos demais.

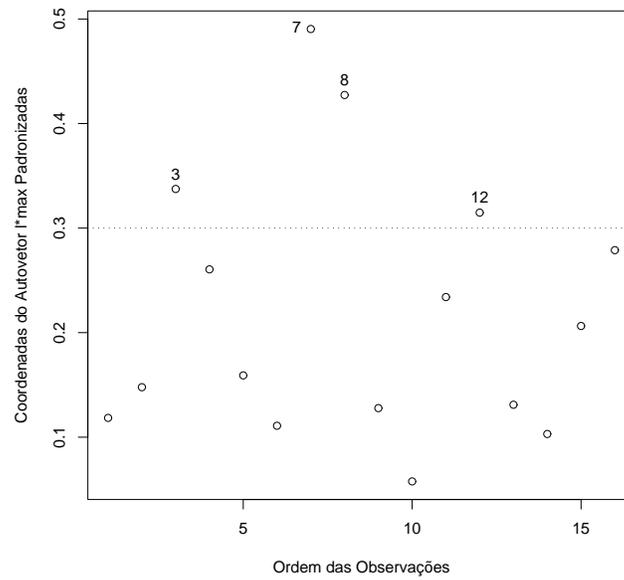


Figura 4.13: Coordenadas do autovetor l_{max}^* contra a ordem das observações

As coordenadas padronizadas em valor absoluto do autovetor l_{max} e do gradiente l_{max}^* são mostradas na tabela a seguir, onde podemos destacar às observações 7 e 8 como as que apresentam as maiores coordenadas do gradiente l_{max}^* , enquanto que as observações 1 e 13 apresentam as maiores coordenadas do vetor l_{max} . As técnicas utilizadas aqui, para detectar *outliers*, não mostraram indicação consistente em relação a algum ponto dos dados, o que talvez indique que para esta aplicação não haja nenhuma observação que possa trazer modificações substanciais ao resultado do ajuste. Em virtude disso, talvez, não tenha havido coincidência entre os pontos destacados.

Tabela 4.3: Coordenadas Padronizadas dos Vetores ℓ_{max} e ℓ_{max}^*

Observação(i)	$ \ell_j^* $	$ \ell_j $
1	0,1184	0,9094
2	0,1478	0,0355
3	0,3374	0,1429
4	0,2605	0,0612
5	0,1591	0,2111
6	0,1109	0,0622
7	0,4904	0,0089
8	0,4273	0,0607
9	0,1278	0,1494
10	0,0577	0,1514
11	0,2340	0,0274
12	0,3148	0,1112
13	0,1310	0,8786
14	0,1030	0,0932
15	0,2063	0,2314
16	0,2788	0,1577

4.3 Preferência por um Novo Detergente em Relação a um Detergente Padrão

Os dados usados nesta seção encontram-se em Cordeiro (1986, pg.117), referen-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.

Adotamos para os dados o modelo logístico linear múltiplo com variável resposta Y =proporção de indivíduos que preferem o novo detergente e variáveis explicativas $temp_j$, onde: $temp_1$ = alta e $temp_2$ = baixa e $dure_k$ onde $dure_1$ = média, $dure_2$ = mole e $dure_3$ = pesada, definidas como fatores. O modelo é dado por

$$\eta_i = \beta_1 + \sum_{j=1}^2 \beta_{2j} temp_j + \sum_{k=1}^3 \beta_{3k} dure_k, \quad i = 1, \dots, 6, \quad (4.3)$$

com erro binomial; é assumido que $\beta_{21} = \beta_{31} = 0$. O desvio obtido para este modelo é $D(\mathbf{y}, p) = 0,618$ (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 (4.4) a seguir temos as estimativas dos coeficientes do modelo adotado e seus respectivos desvios padrões, todos apresentaram-se significativos ao modelo.

Tabela 4.4: Estimativas dos coeficientes referentes ao modelo em (4.3), para explicar a proporção de pessoas que preferem o novo detergente.

Coefficiente	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. Podemos verificar através do gráfico de probabilidade com envelopes, dado na Fig.4.14, que não há indicação forte de que a variável resposta não siga uma Binomial (m_i, p_i) .

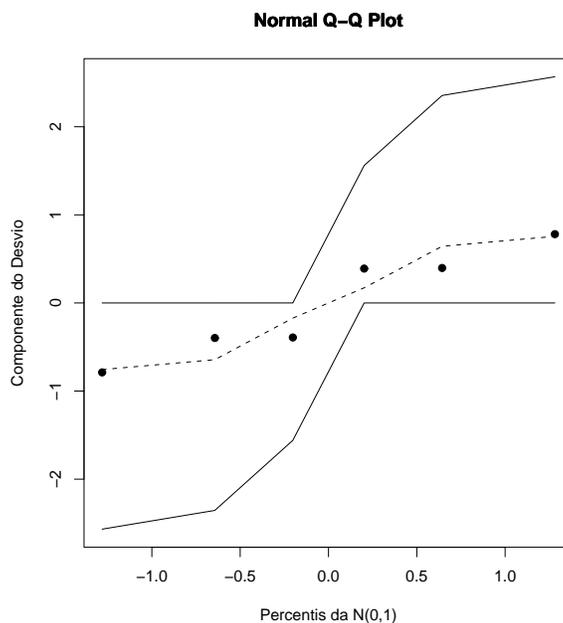


Figura 4.14: Gráfico de probabilidade com envelopes para o modelo binomial.

No gráfico dos \hat{h}_{ii} em função da ordem das observações, dado no Fig.4.15, 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 as observações 1,3 e 5.

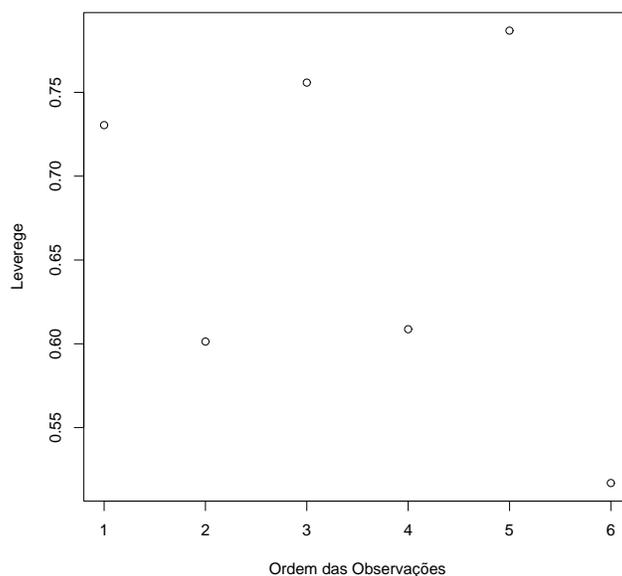


Figura 4.15: Leverage \hat{h}_{ii} contra a ordem das Observações

O gráfico de \hat{h}_{ii} em função dos valores ajustados, mostrado na Fig.4.16, destaca os pontos correspondentes às observações 1,3 e 5 como estando mais afastados.

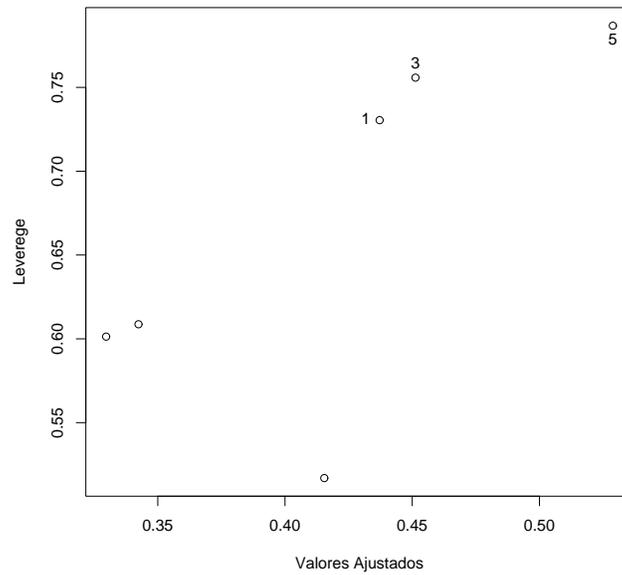


Figura 4.16: Leverage \hat{h}_{ii} contra valores ajustados

No gráfico do resíduo t_{D_i} em função dos valores ajustados, Fig.4.17, os pontos que encontram-se mais afastados correspondem às observações 2,3 e 5.

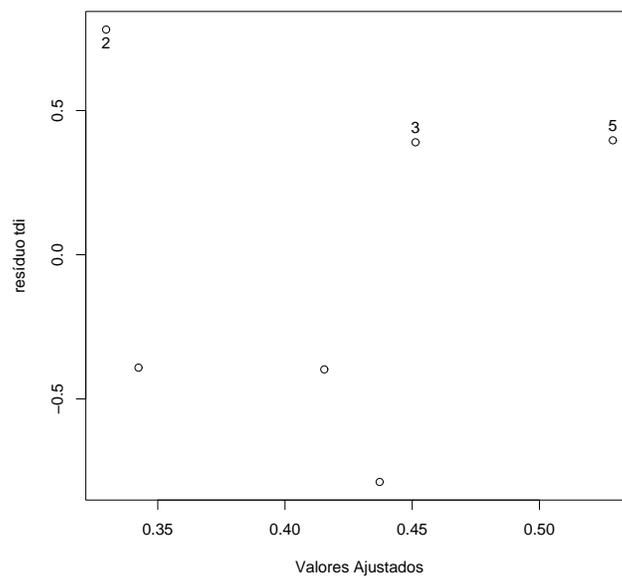


Figura 4.17: Resíduo t_{D_i} contra valores ajustados

Na Fig.4.18 temos o gráfico da medida afastamento pela verossimilhança, LD_i , em função da ordem das observações, nele podemos perceber que o ponto correspondente à observação 1 aparece mais afastado.

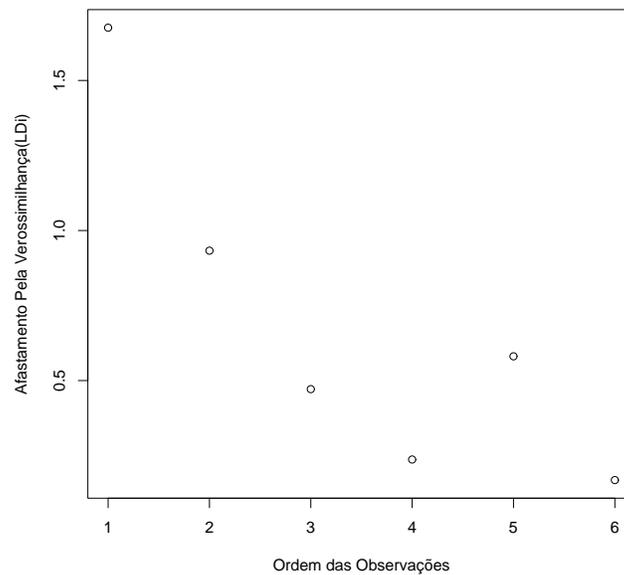


Figura 4.18: Gráfico de LD_i contra a ordem das Observações

No gráfico das coordenadas do autovetor ℓ_{max} em função da ordem das observações, mostrado na Fig.4.19, podemos observar que os pontos correspondentes às observações 1 e 2, encontram-se mais afastados, em especial a de número 2.

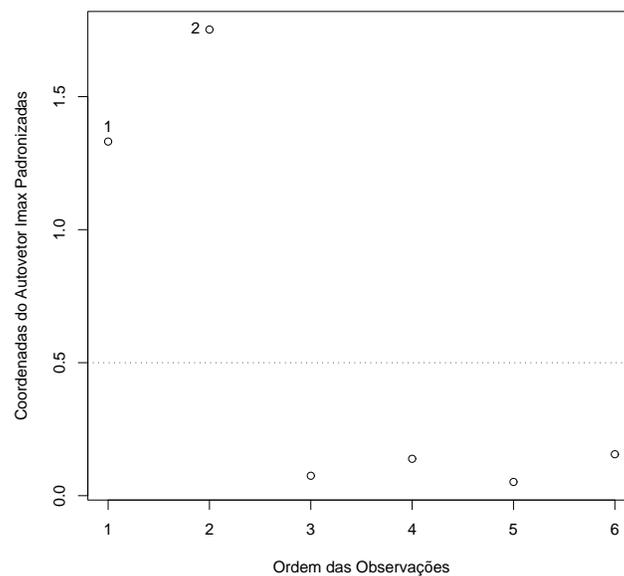


Figura 4.19: Coordenadas do autovetor ℓ_{max} contra a ordem das observações

Avaliando a situação dos pontos destacados pelos gráficos não percebemos uma forte evidência de ocorrência de *outliers*, porém os pontos que apresentaram-se mais afastados foram aqueles correspondentes às observações 1, 2, 3 e 5. Para calcular a curvatura normal, optamos por perturbar a observação 1, para isso, adotamos um

procedimento análogo ao realizado na aplicação da seção (4.2), utilizamos o vetor de perturbação ℓ 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 1298,88. O valor da inclinação máxima obtido para o esquema de perturbação abordado na seção (4.3) foi de 171,75.

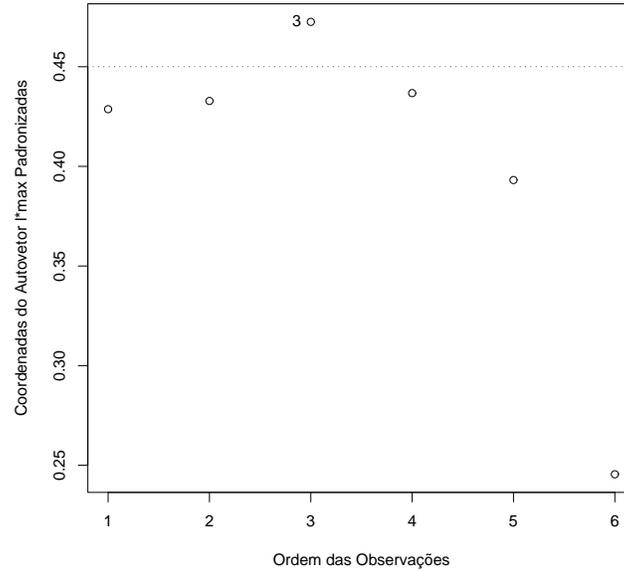


Figura 4.20: Coordenadas do autovetor ℓ_{max}^* contra a ordem das observações

O gráfico mostrado na Fig.4.20, das coordenadas do vetor gradiente ℓ_{max}^* em função da ordem das observações, destaca o ponto correspondente à observação 3.

A tabela (4.4) dispõe 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 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 4.5: Coordenadas Padronizadas dos Vetores ℓ_{max} e ℓ_{max}^*

Observação(i)	$ \ell_j^* $	$ \ell_j $
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

Capítulo 5

Conclusões

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 destacamos 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 3, é justamente a falta de um valor de referência para comparar com o valor obtido para a curvatura normal. Em resposta a essa e outras dificuldades da versão de Cook (ver Capítulo 3) Billor e Loynes (1993) propõem 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 apresentamos o cálculo da inclinação máxima no caso de esquema de perturbação de observação do modelo logístico, porém não sugerimos nenhum valor de referência para efeito de comparação que leve a indicação da sensibilidade do modelo. Destacamos 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 (4.1), onde usamos 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 mais afastadas.

Nas aplicações correspondentes as seções (4.2) e (4.3), onde foram aplicadas o modelo logístico múltiplo, não houveram 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 combinada à inexatidão dos gráficos utilizados na detecção de *outliers*, não nos permitiu decidir pela sensibilidade dos modelos. Embora, os valores tanto da curvatura normal como da inclinação máxima, tenham sido bem maiores no exemplo da aplicação (4.3).

Esperamos ter contribuído com este trabalho para o enriquecimento da análise de diagnóstico de modelos de regressão. Sugerimos que novos trabalhos na área, procurem apresentar valores que sirvam de referência para avaliar a sensibilidade de modelos de regressão não normal.

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 transcórper deste trabalho.

A.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*

$$f(y; \theta_i, \phi) = \exp[\phi \{y\theta_i - b(\theta_i)\} + c(y, \phi)], \quad (\text{A.1})$$

onde 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*

$$f(y_1, \dots, y_n | \theta) = \prod_{i=1}^n f_y(y_i | \theta) \quad (\text{A.2})$$

é denominada de **função de verossimilhança** do parâmetro θ , correspondente a 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*

$$l(\mathbf{y}; \theta, \phi) = \ln f(\mathbf{y}; \theta, \phi). \quad (\text{A.3})$$

A.2 Teoremas

Teorema A.2.1 (Desigualdade de Cauchy-Schwarz) (ver Hoffman & Kunze (1979), pag. 354) Se \mathbf{V} é um espaço com produto interno, então para quaisquer vetores α e β em \mathbf{V} vale a desigualdade

$$|\langle \alpha, \beta \rangle| \leq \|\alpha\| \|\beta\|. \quad (\text{A.4})$$

onde $|\langle \cdot, \cdot \rangle|$ denota o valor absoluto do produto interno de dois vetores quaisquer em \mathbf{V} e $\|\cdot\|$ indica a norma de um vetor em \mathbf{V} .

Teorema A.2.2 (ver Graybill (1983), pag. 183) Seja \mathbf{B} uma matriz $n \times n$ particionada na seguinte forma

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix},$$

onde \mathbf{B}_{ij} possui dimensão $n_i \times n_j$, $i, j = 1, 2$; e $n_1 + n_2 = n$. Suponha que $|\mathbf{B}| \neq \mathbf{0}$, $|\mathbf{B}_{11}| \neq \mathbf{0}$, $|\mathbf{B}_{22}| \neq \mathbf{0}$, onde $|\cdot|$ denota o determinante de uma matriz, além disso, considere a seguinte partição para a inversa da matriz \mathbf{B} , denotada por \mathbf{B}^{-1} :

$$\mathbf{B}^{-1} = \begin{bmatrix} \mathbf{B}^{11} & \mathbf{B}^{12} \\ \mathbf{B}^{21} & \mathbf{B}^{22} \end{bmatrix},$$

onde \mathbf{B}^{ij} possui dimensão $n_i \times n_j$, para $i, j = 1, 2$. Nestas condições, valem os seguintes resultados:

- (i) Existem as inversas das matrizes \mathbf{B}^{11} e \mathbf{B}^{22} .
- (ii) As matrizes $[\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}]^{-1}$ e $[\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}]^{-1}$ existem.
- (iii) A matriz \mathbf{B}^{-1} pode ser escrita como

$$\mathbf{B}^{-1} = \begin{bmatrix} [\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}]^{-1} & -\mathbf{B}_{11}^{-1}\mathbf{B}_{12} [\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}]^{-1} \\ -\mathbf{B}_{22}^{-1}\mathbf{B}_{21} [\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}]^{-1} & [\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}]^{-1} \end{bmatrix},$$

- (iv) $\mathbf{B}^{11} = [\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}]^{-1} = \mathbf{B}_{11}^{-1} + \mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1}$.
- (v) $\mathbf{B}^{12} = -\mathbf{B}_{11}^{-1}\mathbf{B}_{12} [\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}]^{-1} = -\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}^{22}$.
- (vi) $\mathbf{B}^{22} = [\mathbf{B}_{22} - \mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}]^{-1} = \mathbf{B}_{22}^{-1} + \mathbf{B}_{22}^{-1}\mathbf{B}_{21}\mathbf{B}_{11}^{-1}\mathbf{B}_{12}\mathbf{B}_{22}^{-1}$.
- (vii) $\mathbf{B}^{21} = -\mathbf{B}_{22}^{-1}\mathbf{B}_{21} [\mathbf{B}_{11} - \mathbf{B}_{12}\mathbf{B}_{22}^{-1}\mathbf{B}_{21}]^{-1} = -\mathbf{B}_{22}^{-1}\mathbf{B}_{21}\mathbf{B}^{11}$.
- (viii) $|\mathbf{B}| = \frac{\mathbf{B}_{11}}{\mathbf{B}^{11}} = \frac{\mathbf{B}_{22}}{\mathbf{B}^{22}}$ e $|\mathbf{B}_{11}\mathbf{B}^{11}| = |\mathbf{B}_{22}\mathbf{B}^{22}|$.

Teorema A.2.3 (ver Graybill (1983), pag. 189) Seja \mathbf{C} uma matriz $k \times k$ dada por

$$\mathbf{C} = \mathbf{D} + \alpha \mathbf{a} \mathbf{b}^t,$$

onde \mathbf{D} é uma matriz não-singular, \mathbf{a} e \mathbf{b} são vetores $k \times 1$ e α é um escalar tal que

$$\alpha \neq - \left[\frac{\sum_{i=1}^k a_i b_i}{d_{ii}} \right]^{-1}.$$

Nestas condições, a inversa da matriz \mathbf{C} é dada por

$$\mathbf{C}^{-1} = \mathbf{D}^{-1} + \gamma \mathbf{a}^* \mathbf{b}^{*t},$$

onde $\gamma = -\alpha(1 + \alpha \sum_{i=1}^k a_i b_i d_{ii}^{-1})^{-1}$; $a_i^* = \frac{a_i}{d_{ii}}$; $b_i^* = \frac{b_i}{d_{ii}}$ e d_{ii} é o i -ésimo elemento da diagonal da matriz \mathbf{D} .

Teorema A.2.4 (ver Cook & Weisberg (1982), pag. 210) Seja \mathbf{A} uma matriz simétrica $p' \times p'$, considere \mathbf{a} e \mathbf{b} matrizes $q \times p'$. Então as inversas abaixo existem e vale a relação:

$$(\mathbf{A} + \mathbf{a}^t \mathbf{b})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{a}^t (\mathbf{I} + \mathbf{b} \mathbf{A}^{-1} \mathbf{a}^t)^{-1} \mathbf{b} \mathbf{A}^{-1},$$

onde \mathbf{I} é a matriz identidade $q \times q$.

A.3 Proposições

Proposição A.3.1 Seja $\dot{\mathbf{F}}$ o vetor $q \times 1$ definido na seção (2.3.3), então podemos escrever

$$[\mathbf{I}_q + \dot{\mathbf{F}} \dot{\mathbf{F}}^t]^{-1} = \mathbf{I}_q - \frac{1}{1 + \|\dot{\mathbf{F}}\|^2} \dot{\mathbf{F}} \dot{\mathbf{F}}^t,$$

onde \mathbf{I}_q é a matriz identidade de ordem q .

De fato, observe que \mathbf{I}_q é uma matriz diagonal e não-singular, fazendo $\mathbf{D} = \mathbf{I}_q$, $\mathbf{a} = \mathbf{b} = \dot{\mathbf{F}}$ e $\alpha = 1$, segue que $d_{ii} = 1$ e $a_i = b_i = \dot{F}_i$, assim

$$- \left[\sum_{i=1}^q \frac{a_i b_i}{d_{ii}} \right]^{-1} = - \left[\sum_{i=1}^q \frac{\dot{F}_i \dot{F}_i}{1} \right]^{-1} = - \frac{1}{\sum_{i=1}^q \dot{F}_i^2} \neq 1 = \alpha.$$

Logo, todas as condições do teorema (A.2.3) estão satisfeitas. Aplicando o resultado deste teorema, temos que

$$\gamma = -1 \left[1 + 1 \sum_{i=1}^q \dot{F}_i \dot{F}_i 1^{-1} \right]^{-1} = - \frac{1}{1 + \|\dot{\mathbf{F}}\|^2}.$$

Portanto, concluímos que

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

Proposição A.3.2 *Sejam $\mathbf{B} = -\dot{\mathbf{F}} - \lambda \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}}$ e $C = 1 - \dot{\mathbf{F}}^t \dot{\mathbf{F}} - \lambda \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}}$, onde $\lambda = -\frac{1}{1 + \|\dot{\mathbf{F}}\|^2}$ e $\dot{\mathbf{F}}$ é definido na seção (2.3.3) Segue daí que,*

$$\mathbf{B}^t \mathbf{B} = \frac{\|\dot{\mathbf{F}}\|^2}{(1 + \|\dot{\mathbf{F}}\|^2)^2}$$

e

$$C^2 = \frac{1}{(1 + \|\dot{\mathbf{F}}\|^2)^2}.$$

De fato,

(i)

$$\begin{aligned} \mathbf{B}^t \mathbf{B} &= \dot{\mathbf{F}}^t \dot{\mathbf{F}} + \lambda \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}} + \lambda \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}} + \lambda^2 \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}} \\ &= \|\dot{\mathbf{F}}\|^2 + 2\lambda \left(\|\dot{\mathbf{F}}\|^2\right)^2 + \lambda^2 \left(\|\dot{\mathbf{F}}\|^2\right)^3 \\ &= \|\dot{\mathbf{F}}\|^2 + 2\lambda \|\dot{\mathbf{F}}\|^4 + \lambda^2 \|\dot{\mathbf{F}}\|^6 \\ &= \|\dot{\mathbf{F}}\|^2 + 2 \left(-\frac{1}{1 + \|\dot{\mathbf{F}}\|^2} \right) \|\dot{\mathbf{F}}\|^4 + \left(-\frac{1}{1 + \|\dot{\mathbf{F}}\|^2} \right)^2 \|\dot{\mathbf{F}}\|^6 \\ &= \|\dot{\mathbf{F}}\|^2 \left(1 - \frac{2}{1 + \|\dot{\mathbf{F}}\|^2} \|\dot{\mathbf{F}}\|^2 + \frac{1}{(1 + \|\dot{\mathbf{F}}\|^2)^2} \|\dot{\mathbf{F}}\|^4 \right) \\ &= \frac{\|\dot{\mathbf{F}}\|^2}{(1 + \|\dot{\mathbf{F}}\|^2)^2} \left[(1 + \|\dot{\mathbf{F}}\|^2)^2 - 2(1 + \|\dot{\mathbf{F}}\|^2) \|\dot{\mathbf{F}}\|^2 + 1 + \|\dot{\mathbf{F}}\|^4 \right] \\ &= \frac{\|\dot{\mathbf{F}}\|^2}{(1 + \|\dot{\mathbf{F}}\|^2)^2} (1 + 2\|\dot{\mathbf{F}}\|^2 + \|\dot{\mathbf{F}}\|^4 - 2\|\dot{\mathbf{F}}\|^2 - 2\|\dot{\mathbf{F}}\|^4 + \|\dot{\mathbf{F}}\|^4) \\ &= \frac{\|\dot{\mathbf{F}}\|^2}{(1 + \|\dot{\mathbf{F}}\|^2)^2}. \end{aligned}$$

(ii)

$$\begin{aligned}
C^2 &= \left(1 - \dot{\mathbf{F}}^t \dot{\mathbf{F}} - \lambda \dot{\mathbf{F}}^t \dot{\mathbf{F}} \dot{\mathbf{F}}^t \dot{\mathbf{F}}\right)^2 \\
&= \left(1 - \|\dot{\mathbf{F}}\|^2 - \lambda \|\dot{\mathbf{F}}\|^4\right)^2 \\
&= 1 + \|\dot{\mathbf{F}}\|^4 + \lambda^2 \|\dot{\mathbf{F}}\|^8 - 2\|\dot{\mathbf{F}}\|^2 - 2\lambda \|\dot{\mathbf{F}}\|^4 + 2\lambda \|\dot{\mathbf{F}}\|^6 \\
&= 1 - 2\|\dot{\mathbf{F}}\|^2 - 2\left(-\frac{1}{1 + \|\dot{\mathbf{F}}\|^2}\right) \|\dot{\mathbf{F}}\|^4 + \|\dot{\mathbf{F}}\|^4 + 2\left(-\frac{1}{1 + \|\dot{\mathbf{F}}\|^2}\right) \|\dot{\mathbf{F}}\|^6 \\
&\quad + \left(-\frac{1}{1 + \|\dot{\mathbf{F}}\|^2}\right)^2 \|\dot{\mathbf{F}}\|^8 \\
&= \frac{1}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2} \left[\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2 - 2\|\dot{\mathbf{F}}\|^2 \left(1 + \|\dot{\mathbf{F}}\|^2\right)^2 + 2\left(1 + \|\dot{\mathbf{F}}\|^2\right) \|\dot{\mathbf{F}}\|^4 \right. \\
&\quad \left. + \left(1 + \|\dot{\mathbf{F}}\|^2\right)^2 \|\dot{\mathbf{F}}\|^4 - 2\left(1 + \|\dot{\mathbf{F}}\|^2\right) \|\dot{\mathbf{F}}\|^6 + \|\dot{\mathbf{F}}\|^8 \right] \\
&= \frac{1}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2} \left[1 + 2\|\dot{\mathbf{F}}\|^2 + 3\|\dot{\mathbf{F}}\|^4 - \|\dot{\mathbf{F}}\|^8 - 2\|\dot{\mathbf{F}}\|^2 - 4\|\dot{\mathbf{F}}\|^4 - 2\|\dot{\mathbf{F}}\|^6 \right. \\
&\quad \left. + \|\dot{\mathbf{F}}\|^4 + 2\|\dot{\mathbf{F}}\|^6 + \|\dot{\mathbf{F}}\|^8 \right] \\
&= \frac{1}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2} \left[1 + 4\|\dot{\mathbf{F}}\|^4 - 4\|\dot{\mathbf{F}}\|^4 \right] \\
&= \frac{1}{\left(1 + \|\dot{\mathbf{F}}\|^2\right)^2}.
\end{aligned}$$

Proposição A.3.3 *Conforme vimos na seção (3.3) do Capítulo (3), temos 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 $Z_j \sim \chi_{(1)}^2$. Segue daí que

$$\begin{aligned}
A &= \sum_{j=1}^n [Z_j - 1]^2 \\
&= \sum_{j=1}^n (Z_j^2 - 2Z_j + 1).
\end{aligned}$$

Então

$$\begin{aligned}
A^2 &= \{(Z_1^2 - 2Z_1 + 1) + \dots + (Z_n^2 - 2Z_n + 1)\}^2 \\
&= \sum_{j=1}^n (Z_j^2 - 2Z_j + 1)^2 + 2 \sum_{i<j} (Z_i^2 - 2Z_i + 1)(Z_j^2 - 2Z_j + 1) \\
&= \sum_{j=1}^n (Z_j^4 - 4Z_j^3 + 6Z_j^2 - 4Z_j + 1) + 2 \sum_{i<j} (Z_i^2 Z_j^2 - 2Z_i^2 Z_j - 2Z_i Z_j^2 + Z_i^2 + Z_j^2 \\
&\quad + 4Z_i Z_j - 2Z_i - 2Z_j + 1)
\end{aligned}$$

usando a independência entre os Z_i 's, segue que

$$\begin{aligned}
E[A^2] &= \sum_{j=1}^n \{E[Z_j^4] - 4E[Z_j^3] + 6E[Z_j^2] - 4E[Z_j] + 1\} \\
&\quad + 2 \sum_{i \neq j} \{E[Z_i^2]E[Z_j^2] - 2E[Z_i^2 Z_j] - 2E[Z_i]E[Z_j^2] + E[Z_j^2] + 4E[Z_i]E[Z_j]\} \\
&\quad + 2 \sum_{i \neq j} \{-2E[Z_i] - 2E[Z_j] + 1\}. \tag{A.5}
\end{aligned}$$

A função geradora de momentos (FGM) de uma distribuição $\chi_{(k)}^2$ é dada por $(1 - 2t)^{k/2}$ (ver James (1981)). Denotando a FGM da variável aleatória $Z_j = \left(\frac{\varepsilon_j}{\sigma}\right)^2$, por $M_X(t)$, temos que

$$M_X(t) = (1 - 2t)^{1/2}.$$

Sejam $M'_X, M''_X, M_X^{(3)}$ e $M_X^{(4)}$ a primeira, segunda, terceira e quarta derivadas da FGM da variável aleatória Z_j , respectivamente, segue que

$$\begin{aligned}
M'_X(t) &= (1 - 2t)^{-3/2} \Rightarrow E[Z_j] = M'_X(t) \Big|_{t=0} = 1 \\
M''_X(t) &= 3(1 - 2t)^{-5/2} \Rightarrow E[Z_j^2] = M''_X(t) \Big|_{t=0} = 3 \\
M_X^{(3)}(t) &= 15(1 - 2t)^{-7/2} \Rightarrow E[Z_j^3] = M_X^{(3)}(t) \Big|_{t=0} = 15 \\
M_X^{(4)}(t) &= 105(1 - 2t)^{-9/2} \Rightarrow E[Z_j^4] = M_X^{(4)}(t) \Big|_{t=0} = 105.
\end{aligned}$$

Substituindo os resultados acima na equação (A.5) e fazendo algumas simplificações, encontramos

$$\begin{aligned}
E[A^2] &= 60n + 8 \sum_{i<j} 1 \\
&= 60n + 8 \{(n-1) + (n-2) + (n-3) + \dots + (n - (n-1))\} \\
&= 60n + 8 \{n(n-1) - [1 + 2 + 3 + \dots + (n-1)]\} \\
&= 60n + 8 \left\{ n(n-1) - \frac{n(n-1)}{2} \right\} \\
&= 60n + 4(n^2 - n) \\
&= 56n + 4n^2. \tag{A.6}
\end{aligned}$$

A.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 y , a seguir

$$l(\theta, \phi|y) = \ln f(y; \theta, \phi), \quad (\text{A.7})$$

onde y pertence ao suporte da variável aleatória Y .

Dentre as propriedades desta função, podemos destacar que

$$E\left(\frac{\partial l}{\partial \theta}\right) = 0 \quad (\text{A.8})$$

e

$$E\left(\frac{\partial^2 l}{\partial \theta^2}\right) + E\left(\frac{\partial l}{\partial \theta}\right)^2 = 0. \quad (\text{A.9})$$

Conforme a definição da função densidade e/ou probabilidade em (A.1), podemos escrever a equação (A.7), como segue

$$l(\theta, \phi|y) = [y\theta - b(\theta)]\phi + c(y, \phi). \quad (\text{A.10})$$

Derivando a equação acima em relação ao parâmetro θ , encontramos

$$\frac{\partial l}{\partial \theta} = [y - b'(\theta)]\phi, \quad (\text{A.11})$$

onde b' representa a primeira derivada da função $b(\theta)$. Note agora que, combinando 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

$$E(Y) = \mu = b'(\theta). \quad (\text{A.12})$$

Derivando a equação (A.11) temos que

$$\frac{\partial^2 l}{\partial \theta^2} = -b''(\theta)\phi, \quad (\text{A.13})$$

onde 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

$$\text{Var}(Y) = b''(\theta)\phi^{-1}. \quad (\text{A.14})$$

A.4.2 Forma Exponencial da Função Densidade ou de Probabilidade de Algumas Distribuições

A seguir apresentaremos 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,

$$\begin{aligned} f(y; \mu, \sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2}(y - \mu)^2\right] \\ &= \exp\left\{\log\left[\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}e^{-\frac{1}{2\sigma^2}(y-\mu)^2}\right]\right\} \\ &= \exp\left\{\frac{1}{\sigma^2}\left(y\mu - \frac{\mu^2}{2}\right) - \frac{1}{2}\left[\log(2\pi\sigma^2) + \frac{y^2}{\sigma^2}\right]\right\} \end{aligned}$$

onde $-\infty < \mu, y < \infty$ e $\sigma^2 > 0$. Daí obtemos que, $\phi = \frac{1}{\sigma^2}$, $b(\mu) = \frac{\mu^2}{2}$,

$$c(y, \phi) = -\frac{1}{2}\left[\log\left(\frac{2\pi}{\phi}\right) + y^2\sigma^2\right] \text{ e } V(\mu) = 1.$$

- **Poisson**

Neste caso, temos que $Y \sim \text{Poisson}(\lambda)$, a sua função de probabilidade pode ser escrita como,

$$\begin{aligned} f(y; \lambda) &= \frac{e^{-\lambda}\lambda^y}{y!} \\ &= \exp\left[\log\left(\frac{e^{-\lambda}\lambda^y}{y!}\right)\right] \\ &= \exp[-\lambda + y \log \lambda - \log y!] \\ &= \exp[1(y\theta - e^\theta) - \log y!] \end{aligned}$$

onde $\lambda > 0$ e $y = 0, 1, 2, \dots$. Logo, $\phi = 1$, $b(\theta) = e^\theta$, $c(y, \phi) = -\log y!$ e $V(\lambda)\lambda$.

- **Binomial**

Seja $Y \sim \text{Binomial}(n, p)$, então sua função de probabilidade é escrita como

$$f(y; n, p) = \binom{n}{y} p^y (1-p)^{n-y}$$

onde $0 < p < 1$, $y = 0, 1, 2, \dots$. Considere Y^* como a proporção de sucessos em n ensaios independentes, cada um com probabilidade de ocorrência λ . 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 $\theta = \log \left(\frac{p}{1-p} \right)$ obtemos 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) temos que, $\phi = n$, $b(\theta) = \log(1+e^\theta)$, $c(y, \phi) = \log \binom{n}{ny^*}$ e $V(p) = p(1-p)$.

Apêndice B

Aplicativo R

Neste apêndice vamos descrever alguns comandos do aplicativo **R**, que foram utilizadas no Capítulo (4) desta dissertação.

Os resíduos *studentizados* t_i e t_i^* podem ser obtidos no **R**, pela sequência de comandos

```
s<-fit.model$sigma
r<-resid(fit.model)
ti<-r/(s*(1-hii)^0.5)
t*i<-ti*((n-p-1)/(n-p-ti^2))^0,5.
```

Para obter os valores h_{ii} no aplicativo **R**, devemos adotar os procedimentos a seguir, suponha que o resultado do ajuste esteja no objeto `fit.model`, então

```
X<-model.matrix(fit.model)
H<-solve(t(X)%*%X)
H<-X%*%H%*%t(X)
hii<-diag(H).
```

Obtemos os resíduos h_{ii} , t_{S_i} , t_{D_i} , e LD_i para os MLGs, no aplicativo **R**, seguindo os comandos abaixo,

```
X<-model.matrix(fit.model)
n<-nrow(X)
p<-ncol(X)
W<-fit.model$weights
```

```

W<-diag(W) H<-solve(t(X)%*%W%*%X)
H<-sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
hii<-diag(H)
tDi<-resid(fit.model,type="deviance")/sqrt(1-hii)
rp<-resid(fit.model,type="pearson")
tSi<-rp/sqrt(1-hii)
LDi<-hii*(tSi^2)/(1-hii).

```

Para obter o valor da curvatura normal no \mathbf{R} , adotamos a sequência de comandos a seguir, onde o vetor ℓ foi definido na seção (2.3).

```

lms<-summary(fit.model)
s<-lms$sigma
e<-resid(lms)
De<-diag(e)
X<-model.matrix(fit.model)
PX<-X%*%solve(t(X)%*%X)%*%t(X)
p1<-De%*%PX%*%De
esq<-e^2
p2<-(esq%*%t(esq))/2*45*s^2
p3<-p1+p3
Cl<-t(l)%*%p3%*%l
Cl<-2*Cl/s^2.

```

Para obter d_{max}^* no \mathbf{R} , deve-se seguir os comandos:

```

lmax<-esq/s^2
lmax<-lmax^2
dmax<-sum(lmax)
dmax<-sqrt(dmax).

```

Bibliografia

- [1] Billor, N. and Loynes, R. M. (1993). Local Influence: a new approach. *Communs Statist. Theory Meth.*, **22**, 1595-1661. 503-506 (1972).
- [2] Belsley, D.A.; Kuh, E. and Welsch, R.E. (1980). *Regression Diagnostics*. John Wiley, New York.
- [3] Bolfarine, H. e Sandoval, M. C. (2001). *Introdução à Inferência Estatística*. Coleção Matemática Aplicada, SBM.
- [4] Cook, R. D. (1986). Assessment of local influence (with discussion). *J. R. Statist. Soc. B*, **48**, 133-169.
- [5] Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. School of Statistics-University of Minnesota, Chapman and Hall Ltd, New York-NY.
- [6] Cook, R. D. (1977). Detection of influential observations in linear regressions. *Technometrics*, **15**, 15-18.
- [7] Cordeiro, G. M. (1986). *Modelos Lineares Generalizados*. VII Simpósio Nacional de Probabilidade e Estatística, Campinas-SP.
- [8] Davison, A. C. e Gigli, A. (1989). Deviance residuals and normal score plots. *Biometrika* **76**, 211-221.
- [9] Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*. Second Edition, John Wiley and Sons, New York-NY.
- [10] Fung, W. K. and Kwan, C. W. (1997). A Note on Local Influence Based on Normal Curvature. *J. R. Statist. Soc. B*, **59**, 839-843.

- [11] Graybill, F. A. (1983). *Matrices with Applications in Statistics*. Second Edition, Wadsworth International Group, Belmont-California, USA.
- [12] Hoffman, K. e Kunze, R. (1979). *Álgebra Linear*. Segunda Edição, LTC: Livros Técnicos e Científicos, São paulo.
- [13] James, B. R. (1981). *Probabilidade: Um Curso em Nível Intermediário*. Rio de Janeiro: Livro Técnico.
- [14] Lawrence, A. J. (1988). Regression Transformation Diagnostics using Local Influence. *J. Am Statist. Ass.*, **83**,1067-1077.
- [15] Lee, A. H. and Zhao, Y. (1996). Sensitivity of Pearson's goodness-of-fit statistic in generalized linear models. *Commun Statist. Theory Meth.*, **25**, 143-157.
- [16] Lima, E. L. (1981). *Projeto Euclides: Curso de Análise vol 2*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada.
- [17] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Second Edition, Chapman and Hall, London-New York.
- [18] Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society A*, **135**, 370-384.
- [19] O'neill, B. (1997). *Elementary Differential Geometry: Second Edition*. London: Academic Press.
- [20] Paula, G. A. (2004). *Modelos de Regressão com Apoio Computacional*. Instituto de Matemática e Estatística-USP, São Paulo-SP, Brasil.
- [21] Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics*, **9**, 705-724.
- [22] Stoker, J. J. (1969). *Differential Geometry*. New York: Wiley.
- [23] Schall, R. and Dunne, T.T. (1992). A note on the Relationship between parameter collinearity and local influence. *Biometrika*, **79**, 399-404.
- [24] Thomas, W. and Cook, R. D. (1990). Assessing influence on predictions from generalized linear models. *Technometrics*, **32**, 59-65.
- [25] Tsai, C.L. (1986). Discussion of Assessment of local influence by R.D. Cook. *J. R. Statist. Soc. B*, **48**, 165.

- [26] Venables, W. N. and Ripley, D. M. (2003). *An Introduction to R: A Programming Environment for Data Analysis and Graphics*. Version 1.6.2.
- [27] Weisberg, S. (1985). *Applied Linear Regression, 2nd Edition*. John Wiley.