

# Resumo

O método de influência local proposto por Cook (1986) é uma importante ferramenta na análise da influência conjunta das observações nos resultados de um ajuste de regressão. Esta técnica tem como objetivo principal avaliar mudanças nos resultados da análise quando pequenas perturbações são incorporadas ao modelo e/ou aos dados. Se essas perturbações causarem efeitos desproporcionais nas estimativas, pode ser indício de que o modelo está mal ajustado ou que possam existir afastamentos sérios das suposições feitas para o mesmo. Apesar do método baseado na curvatura normal proposto por Cook (1986) ser de grande utilidade, este possui alguns inconvenientes. Por exemplo, a curvatura normal pode tomar qualquer valor real e não é invariante sob uma mudança uniforme de escala. Neste trabalho, estudamos a influência local conforme, proposto por Poon & Poon (1999) em modelos de regressão, que tem como objetivo contornar estes inconvenientes. Mais especificamente, aplicamos esta técnica de diagnóstico para os modelos de regressão log-Birnbaum-Saunders e derivamos as matrizes apropriadas para obter a influência local nos parâmetros estimados de um modelo mais geral, o modelo de regressão senh-normal. Finalmente, ilustramos a teoria desenvolvida em conjuntos de dados reais.

# Abstract

The method of local influence proposed by Cook (1986) is an important tool in the analysis of the joint influence of the observations in the results of a regression fit. This technique has as main goal to evaluate changes in the results of the analysis when small perturbations are incorporated the model and/or to the data. If these perturbations would cause non proportion effect it, can be indication of that the model is badly fitted or that possible departures from the assumptions made for this model can exist. Although the method based on normal curvature proposed by Cook (1986) has been demonstrated to be very useful, this possesses some issues. For example, the normal curvature may take any value and it is not invariant under a uniform change of scale. In this work, we study the conformal local influence, proposed by Poon & Poon (1999) in regression models, that has as objective to address these issues. More specifically, we apply this technique of diagnostic for log-Birnbaum-Saunders regression models and we derive matrices appropriate to obtain the conformal local influence in the estimated parameters of a more general model, the senh-normal regression model. Finally, we consider empirical examples with real data to illustrate the theory developed.

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

Diagnóstico baseado na influência  
local conforme para os modelos de  
regressão Birnbaum-Saunders e  
senh-normal

por

José Iraponil Costa Lima

sob orientação da

Prof<sup>a</sup>. Dr<sup>a</sup>. Michelli Karinne Barros da Silva

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

Setembro/2008

# Diagnóstico baseado na influência local conforme para os modelos de regressão Birnbaum-Saunders e senh-normal

por

**José Iraaponil Costa Lima**

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.

Aprovada por:

---

**Prof. Dr. Víctor Leiva**

---

**Prof. Dr. Francisco Antonio Morais de Souza**

---

**Prof<sup>a</sup>. Dr<sup>a</sup>. Michelli Karinne Barros da Silva**

**Orientadora**

**Universidade Federal de Campina Grande**

**Centro de Ciências e Tecnologia**

**Programa de Pós-Graduação em Matemática**

**Curso de Mestrado em Matemática**

**Setembro/2008**

# Agradecimentos

Primeiramente, agradeço a Deus, pois Ele é quem concede o saber; e à Virgem Santíssima Nossa Senhora do Perpétuo Socorro, pois jamais deixará de interceder por seus filhos.

À Professora Michelli Karinne Barros da Silva, pela orientação que resultou na confecção deste trabalho. E ao seu esposo Severino Horácio da Silva.

Aos Professores Francisco Antônio Morais de Souza e Víctor Leiva, por aceitarem o convite para participarem da banca examinadora.

A todos os professores da Unidade Acadêmica de Matemática e Estatística da UFCG (UAME/UFCG), pois contribuíram efetivamente para minha formação como profissional e ser humano.

Aqueles que compõem a Coordenação do Programa de Pós-Graduação em Matemática da UFCG.

Aos funcionários técnicos administrativos da UAME/UFCG.

Aos meus colegas discentes.

Aos meus familiares, particularmente, à minha irmã e ao seu esposo.

Aos meus colegas de trabalho da Agência dos Correios-Soledade.

Aos meus amigos, pelo companheirismo.

# Dedicatória

À minha esposa Andréia e às nossas filhas, Júlia e Laura.

# Conteúdo

<b>Introdução</b> . . . . .	<b>6</b>
<b>1 Preliminares</b>	<b>9</b>
1.1 Caracterização de dados de sobrevivência . . . . .	10
1.2 Função de sobrevivência, taxa de risco e função de risco acumulativa . .	11
1.2.1 Modelos contínuos . . . . .	11
1.2.2 Modelos discretos . . . . .	18
1.3 Função de verossimilhança, função escore e matriz de informação de Fisher	20
1.3.1 Função de verossimilhança . . . . .	20
1.3.2 Função escore e matriz de informação de Fisher . . . . .	22
<b>2 O modelo de regressão log-Birnbaum-Saunders</b>	<b>23</b>
2.1 A distribuição Birnbaum-Saunders . . . . .	25
2.2 A distribuição senh-normal . . . . .	28
2.3 O modelo de regressão log-Birnbaum-Saunders . . . . .	32
2.3.1 Caracterização . . . . .	32
2.3.2 Estimação . . . . .	33
<b>3 Influência local mediante a curvatura normal conforme</b>	<b>38</b>

	ii
3.1	Influência local, gráfico de influência e curvatura normal . . . . . 39
3.2	A curvatura normal conforme . . . . . 42
3.3	Propriedades da curvatura normal conforme . . . . . 43
3.4	Influência local por meio da curvatura normal conforme . . . . . 48
3.4.1	Influência de um único autovetor . . . . . 49
3.4.2	Contribuição agregada de vetores básicos de perturbação . . . . . 50
<b>4</b>	<b>Influência local no modelo de regressão log-Birnbaum-Saunders</b> . . . . . <b>53</b>
4.1	Esquemas de perturbações . . . . . 54
4.1.1	Ponderação de casos . . . . . 54
4.1.2	Perturbação na variável resposta . . . . . 55
4.1.3	Perturbação em uma variável explicativa . . . . . 56
4.2	Aplicação . . . . . 58
<b>5</b>	<b>Influência local em modelos de regressão lineares com erro senh-normal</b> . . . . . <b>64</b>
5.1	Caracterização e estimação . . . . . 65
5.2	Influência local . . . . . 67
5.2.1	Ponderação de casos . . . . . 67
5.2.2	Perturbação na variável resposta . . . . . 69
5.3	Aplicação . . . . . 69
	<b>Conclusão</b> . . . . . <b>78</b>
	<b>Trabalhos futuros</b> . . . . . <b>79</b>
	<b>Apêndices</b> . . . . . <b>80</b>
	<b>Bibliografia</b> . . . . . <b>86</b>

# Abreviações e notações

fdp	:	função de densidade de probabilidade.
fda	:	função de distribuição acumulada.
BS	:	Birnbaum-Saunders.
SN	:	senh-normal.
D	:	Conj. das unidades experimentais associadas ao tempo de falha.
C	:	Conj. das unidades experimentais associadas ao tempo censurado.
$T \sim \text{BS}(\alpha, \beta)$	:	$T$ segue uma distribuição Birnbaum-Saunders com parâmetros $\alpha$ e $\beta$ .
$Y \sim \text{SN}(\alpha, \gamma, \sigma)$	:	$Y$ segue uma distribuição senh-normal com parâmetros $\alpha$ , $\gamma$ e $\sigma$ .
$\phi$	:	função de densidade de probabilidade da distribuição normal padrão.
$\Phi$	:	função de distribuição acumulada da distribuição normal padrão.
$\Omega$	:	espaço de perturbação.
$\omega$	:	vetor de perturbação.
$\ell(\theta)$	:	função de log-verossimilhança de $\theta$ para o modelo adotado.
$\ell(\theta \omega)$	:	função de log-verossimilhança de $\theta$ para o modelo perturbado.
$\hat{\theta}$	:	estimador de máxima verossimilhança de $\theta$ no modelo adotado.
$\hat{\theta}_\omega$	:	estimador de máxima verossimilhança de $\theta$ no modelo perturbado.
$\nabla_f$	:	vetor gradiente da função $f$ .
$\mathbf{I}$	:	matriz associada a primeira forma fundamental.
$\mathbf{\Pi}$	:	matriz associada a segunda forma fundamental.
$C_{\mathbf{l}}$	:	curvatura normal na direção do vetor não-nulo $\mathbf{l}$ .
$B_{\mathbf{l}}$	:	curvatura normal conforme na direção do vetor não-nulo $\mathbf{l}$ .
$\bar{\lambda}_i$	:	autovalor normalizado.
$q$ -influyente	:	ordem de magnitude de influência igual a $q$ .
$m(q)_j$	:	contribuição agregada do $j$ -ésimo vetor básico de perturbação $\mathbf{E}_j$ .
$m_j$	:	contribuição agregada total.

# Introdução

Em uma análise estatística, faz-se necessário avaliar a estabilidade dos resultados inferenciais, no que diz respeito à perturbação nos dados ou no modelo estatístico adotado. Para tanto, podemos fazer uso da técnica de análise de diagnóstico de influência cujo objetivo é verificar possíveis violações das hipóteses necessárias para a utilização do modelo adotado. Paralelamente, busca-se detectar a existência de observações mal ajustadas (*outliers*) e a presença de observações influentes, ou seja, aquelas que interferem de forma desproporcional nos resultados obtidos, quando a mesma for excluída ou sujeita à pequenas perturbações.

As técnicas de diagnóstico baseavam-se, comumente, na análise de resíduos. Entretanto, Cook (1986) propôs uma metodologia inovadora, denominada influência local, que consiste em avaliar por meio de uma medida apropriada de influência, baseada na curvatura normal, a estabilidade das estimativas fornecidas pelo modelo mediante pequenas perturbações nos dados ou no próprio modelo. Além disso, essa metodologia possibilita avaliar a influência conjunta de todos os pontos.

Porém, Poon & Poon (1999) chamam atenção para o fato de que a curvatura normal pode assumir qualquer valor real e não é invariante sob uma mudança uniforme de escala, ocasionando perda de objetividade no julgamento da grandeza da curvatura. Com o objetivo de solucionar esse problema e, conseqüentemente, aperfeiçoar o método de influência local, eles propõem que façamos uso da curvatura normal conforme, que está relacionada com a curvatura normal, mas assume valores em um intervalo limitado da reta e é invariante sob uma classe de reparametrizações denominadas reparametriza-

ções conformes.

Neste trabalho, expomos a técnica de influência local mediante a curvatura normal conforme proposta por Poon & Poon (1999) e aplicamos essa ferramenta nos modelos de regressão Birnbaum-Saunders e senh-normal. A distribuição Birnbaum-Saunders apresentada por Birnbaum-Saunders (1969a) relaciona o tempo até a ocorrência de falha com um dano acumulativo que se supõem seguir uma distribuição normal padrão. A distribuição senh-normal desenvolvida por Rieck & Nedelmam (1991) é simétrica e possui curtose maior ou menor do que a distribuição normal padrão e tem como caso particular a distribuição Birnbaum-Saunders sob uma transformação logarítmica.

Esta dissertação se encontra dividida em cinco capítulos. No primeiro capítulo fazemos uma breve explanação sobre alguns tópicos da análise de sobrevivência. O segundo capítulo consiste no estudo das distribuições Birnbaum-Saunders e senh-normal, bem como na modelagem de regressão do modelo log-Birnbaum-Saunders. No terceiro capítulo apresentamos a formulação da influência local através da curvatura normal conforme. No quarto capítulo, aplicamos a técnica de influência local por meio da curvatura normal conforme no modelo de regressão log-linear para a distribuição Birnbaum-Saunders e ilustramos a teoria a um conjunto de dados reais. No quinto capítulo consideramos o modelo de regressão senh-normal, ou seja, estendemos os resultados de diagnóstico de influência desenvolvidos por Leiva et al. (2007) sob modelos de regressão linear log-BS, considerando agora que o parâmetro  $\sigma$  é desconhecido. Para este modelo derivamos as curvaturas normais da influência local conforme sob os esquemas de perturbação: ponderação de casos e perturbação da variável resposta e ilustramos a teoria a um conjunto de dados reais.

# Capítulo 1

## Preliminares

Na análise de sobrevivência, a variável resposta é, geralmente, o tempo até a ocorrência de um evento de interesse. Por exemplo, Colosimo & Giolo (2001) citam um estudo de natureza ambiental sobre mangueiras sujeitas a determinada praga capaz de matar a planta, no qual o interesse recai sobre o tempo de vida, em anos, até a morte da árvore; Lawless (1982), reportando-se a Nelson (1970b), cita um experimento no qual novos modelos de uma ferramenta foram testados e em que a variável de interesse era o número de ciclos completos até a falha.

Em muitos experimentos o acompanhamento das unidades estatísticas é, comumente, interrompido, seja porque o estudo terminou para análise dos dados ou, a falha ocorreu devido a causa diferente da estudada. Neste caso, a informação correspondente à resposta se resume ao conhecimento de que o tempo de falha é superior ao observado, implicando em uma resposta parcial que, na análise de sobrevivência, chamamos de censura. A censura é uma das principais características dos dados de sobrevivência. Na ausência da mesma, utilizamos as técnicas clássicas da Estatística para análise de dados. Porém, a presença de censura inibe o uso dessas técnicas, exigindo o emprego dos métodos da análise de sobrevivência, pois estes possibilitam incorporar na análise a informação contida nos dados censurados.

## 1.1 Caracterização de dados de sobrevivência

Os dados de sobrevivência são caracterizados pelos tempos de falha e, comumente, pelas censuras. Esses dois componentes constituem a resposta. Além disso, o uso de variáveis explicativas é comum em algumas situações.

O tempo de falha é constituído por três componentes: o tempo inicial, a escala de medida e o evento de interesse (a falha). O tempo inicial é o marco temporal que determina o início do estudo e onde as unidades são comparadas pela primeira vez. Já a escala de medida é, geralmente, o tempo real; mas há outras escalas, como por exemplo, o número de ciclos de um motor e o número de fissuras em uma coluna de concreto. O terceiro e último elemento, a falha, deve ser claramente definida antes do início do estudo, na maioria das situações as falhas são eventos indesejáveis, tais como queima de um componente eletrônico, quebra de um motor, inapropriação de produtos ao consumo, rompimento de uma coluna.

Em conformidade com Lawless (1982), dizemos que uma observação é censurada à direita em  $L$  se o valor correspondente à observação é desconhecido, mas com a garantia de que é no mínimo  $L$ . De forma análoga, uma observação é censurada à esquerda em  $L$  se o valor associado à mesma é desconhecido e no máximo  $L$ . Segundo o mesmo autor, em estudos de sobrevivência, a censura à direita é comum, porém a censura à esquerda é, geralmente, rara. Em nosso trabalho, consideraremos censura à direita.

Em algumas situações, fixamos um período limitado de tempo para realização do experimento. Conseqüentemente, podemos nos deparar com unidades experimentais que não venham a falhar nesse período pré-determinado, acarretando observações censuradas. Neste caso, dizemos que os dados estão sob censura tipo I. Em outras situações, o experimento prossegue até que um número pré-fixado de unidades falhem. Este número, digamos  $r$ , não é necessariamente igual ao número  $n$  de unidades amostradas, mas é tal que  $r \leq n$ . Uma vez fixado  $r$ , é evidente o fato de que, no exato momento em que a  $r$ -ésima unidade falhar, podem existir espécimes que não tenham falhado, implicando em observações censuradas, neste caso dizemos que os dados estão sob censura tipo II. Nos dois tipos de censura citados anteriormente o pesquisador influencia,

de forma direta, nos dados censurados. Um terceiro mecanismo de censura é a censura aleatória, que ocorre quando algumas unidades experimentais são retiradas do estudo sem ter ocorrido a falha, ou também, por exemplo, se a falha deve-se a razão diferente da estudada.

Na presença de covariáveis, os dados de sobrevivência correspondentes à  $i$ -ésima unidade experimental sob estudo são, em geral, representados por  $(t_i, \delta_i, \mathbf{x}_i)$ , sendo  $t_i$  o tempo de falha ou de censura,  $\delta_i$  a variável indicadora de falha ou de censura tal que  $\delta_i = 1$  se  $t_i$  é tempo de falha, ou  $\delta_i = 0$  se  $t_i$  é tempo censurado, e  $\mathbf{x}_i$  é o vetor de variáveis explicativas.

## 1.2 Função de sobrevivência, taxa de risco e função de risco acumulativa

### 1.2.1 Modelos contínuos

Seja  $T$  uma variável aleatória contínua não-negativa, representando o tempo até a ocorrência de falha de unidades estatísticas que compõem uma mesma população. Associada à variável  $T$ , definimos uma função  $S: \mathbb{R} \rightarrow \mathbb{R}$ , conforme Lawless (1982) tal que

$$S(t) = P(T \geq t). \quad (1.1)$$

A função definida pela equação (1.1) é denominada função de sobrevivência.

**Exemplo 1** *Seja  $T$  uma variável aleatória cuja função de densidade de probabilidade (fdp) é*

$$f(t) = \lambda \exp(-\lambda t) \ , \ t \geq 0.$$

*Ou seja,  $T$  possui uma distribuição exponencial com parâmetro  $\lambda > 0$ . A função de sobrevivência de  $T$  é dada por*

$$S(t) = \exp(-\lambda t) \ , \ t \geq 0.$$

A Figura 1.1 apresenta o gráfico da função de sobrevivência da distribuição exponencial para diferentes valores do parâmetro  $\lambda$ . Como podemos observar, a função de

sobrevivência assume o máximo em zero, e à medida que o tempo cresce, a função de sobrevivência vai tendendo a zero.

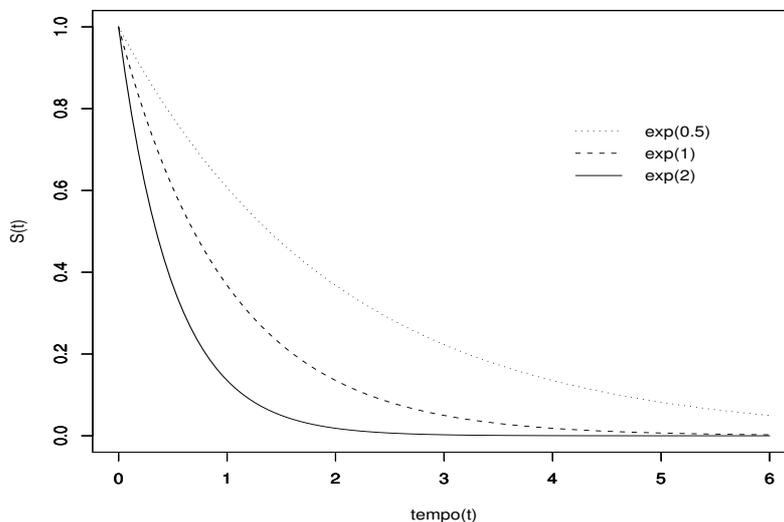


Figura 1.1: gráfico da função de sobrevivência da distribuição exponencial para  $\lambda = 0.5$ ,  $\lambda = 1$  e  $\lambda = 2$ .

**Propriedade 1.1** *Se  $T$  é uma variável aleatória absolutamente contínua e não-negativa, então sua função de sobrevivência  $S$  possui as seguintes propriedades:*

- (i)  $S$  é monótona não-crescente;
- (ii)  $S(0) = 1$  e  $\lim_{t \rightarrow +\infty} S(t) = 0$ ;
- (iii)  $S$  é contínua.

**Prova:** (i) sejam  $t_1, t_2 \in \mathbb{R}$  tais que  $t_1 < t_2$ . Logo, o evento  $A = \{t \in \mathbb{R}; t \geq t_1\}$  contém o evento  $B = \{t \in \mathbb{R}; t \geq t_2\}$ . Conseqüentemente,

$$S(t_1) = P(T \geq t_1) \geq P(T \geq t_2) = S(t_2);$$

assegurando-nos que  $S$  é não-crescente.

Agora provaremos (ii), por definição  $S(0) = P(T \geq 0) = \int_0^{\infty} f(t)dt$ , sendo  $f$  a fdp da variável  $T$ . Dado que  $T$  é não-negativa, segue que  $\int_0^{\infty} f(t)dt = 1$ , portanto,

$S(0) = 1$ . A fim de provarmos que  $\lim_{t \rightarrow +\infty} S(t) = 0$ , seja  $\{t_n\}_{n \geq 1}$  uma seqüência não-decrescente de números reais tal que  $t_n \rightarrow +\infty$ . Conseqüentemente, a seqüência de eventos  $\{B_n\}_{n \geq 1}$  com  $B_n = \{t \in \mathbb{R}; t \geq t_n\}$  é tal que

$$B_{n+1} \subset B_n, \forall n \geq 1$$

e

$$\lim_{n \rightarrow +\infty} B_n = \bigcap_{n \geq 1} B_n = \emptyset.$$

Ou seja,  $\{B_n\}_{n \geq 1}$  é uma seqüência de eventos que decresce monotonicamente para  $\emptyset$ . Logo,

$$\lim_{t_n \rightarrow +\infty} S(t_n) = \lim_{t_n \rightarrow +\infty} P(T \geq t_n) = \lim_{t_n \rightarrow +\infty} P(B_n) = P\left(\lim_{n \rightarrow +\infty} B_n\right) = 0.$$

Por fim, provaremos (iii), dado que  $T$  é uma variável aleatória contínua, então  $T$  possui uma fdp  $f$ . Logo,

$$S(t) = P(T \geq t) = 1 - \int_{-\infty}^t f(x)dx,$$

que é uma função contínua. ■

**Propriedade 1.2** *Se  $T$  é uma variável aleatória absolutamente contínua e não-negativa, então*

$$f(t) = -\frac{d}{dt}S(t),$$

*nos pontos em que  $S$  é derivável.*

**Prova:** dado que  $T$  é uma variável contínua, então

$$S(t) = 1 - \int_{-\infty}^t f(x)dx.$$

Conseqüentemente,

$$\frac{d}{dt}S(t) = -f(t),$$

nos pontos em que  $S$  é derivável. ■

Sejam  $t_1, t_2 \in \mathbb{R}$  tais que  $t_1 < t_2$ . Colosimo & Giolo (2006) definem a taxa de falha no intervalo  $[t_1, t_2)$  como sendo a razão entre a probabilidade de que a falha ocorra

nesse intervalo, dado que não ocorreu antes de  $t_1$ ; e o comprimento do intervalo  $[t_1, t_2)$ . Ou seja, a taxa de falha é dada por

$$\frac{P(t_1 \leq T < t_2 | T \geq t_1)}{t_2 - t_1}. \quad (1.2)$$

Usando o fato que

$$\begin{aligned} P(t_1 \leq T < t_2 | T \geq t_1) &= \frac{P(t_1 \leq T < t_2)}{P(T \geq t_1)} \\ &= \frac{P(T \geq t_1) - P(T \geq t_2)}{P(T \geq t_1)} \\ &= \frac{S(t_1) - S(t_2)}{S(t_1)}, \end{aligned}$$

a razão (1.2) pode ser reescrita como

$$\frac{S(t_1) - S(t_2)}{[t_2 - t_1]S(t_1)}. \quad (1.3)$$

Neste ponto, destacamos o fato de que as taxas de falha não assumem valores negativos, pois a função de sobrevivência é não-crescente.

Escrevendo  $t_1 = t$  e  $t_2 = t + \Delta t$ , a expressão (1.3), torna-se:

$$\frac{S(t) - S(t + \Delta t)}{\Delta t S(t)}.$$

Definimos a função taxa de falha da variável  $T$ , ou taxa de risco, e indicamos por  $h$ , como sendo o limite:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}.$$

Ou, equivalentemente,

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} = -\frac{1}{S(t)} \frac{d}{dt} S(t), \quad (1.4)$$

A função  $h(t)$  representa a taxa instantânea de falha no tempo  $t$ , condicional à sobrevivência até o tempo  $t$ .

Na Figura 1.2 apresentamos os gráficos das taxas de risco da distribuição gama, para diferentes valores do parâmetro de forma  $k$  e parâmetro de escala  $\lambda = 1$ .

Observemos que, para  $k = 2$  e  $\lambda = 1$ , a função de risco da distribuição gama é crescente para todo  $t > 0$ . Porém, para  $k = 1$  e  $\lambda = 1$  tal função é constante.

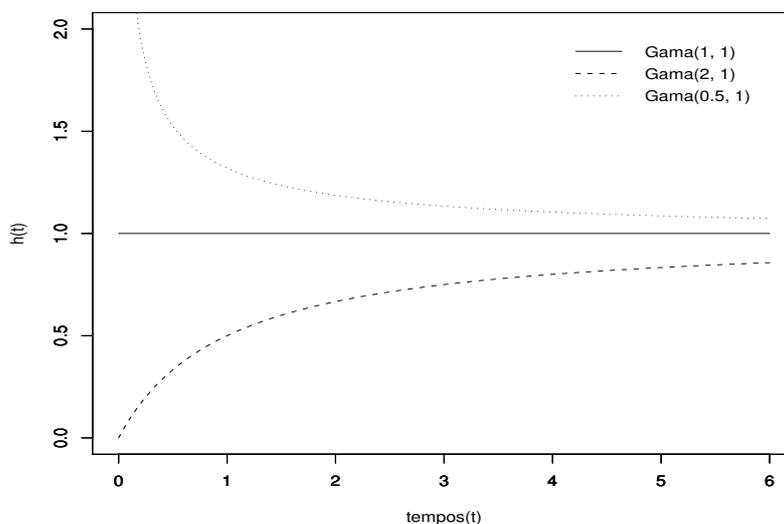


Figura 1.2: gráfico das taxas de risco da distribuição gama para  $k = 0, 5, 1$  e  $2$  e  $\lambda = 1$ .

Finalmente, para  $k = 0.5$  e  $\lambda = 1$  a função é decrescente. Na verdade, para  $k > 1$  a função de risco da distribuição gama é crescente, com  $h(0) = 0$  e  $\lim_{t \rightarrow +\infty} h(t) = \lambda$ , para  $0 < k < 1$  a mesma é decrescente com  $\lim_{t \rightarrow 0^+} h(t) = +\infty$ . Além disso, para  $k = 1$  tal função é constante igual a  $\lambda$  (ver Lawless, 1982).

Destacamos o fato de que as taxas de risco não são necessariamente funções monótonas. Por exemplo, a taxa de risco da distribuição log-normal com parâmetros  $\mu = 0$  e  $\sigma = 0.5$  não é uma função monótona. Observemos o gráfico dessa função na Figura 1.3, onde perceberemos a não monotonicidade dessa função, pois  $h(0) = 0$ , depois cresce para um máximo e por fim decresce quando  $t$  tende ao infinito (para mais detalhes veja Goldthwaite (1961)).

A seguir, apresentaremos algumas propriedades das taxas de risco.

**Propriedade 1.3** *Se  $T$  é uma variável aleatória absolutamente contínua e não-negativa, então*

$$h(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t).$$

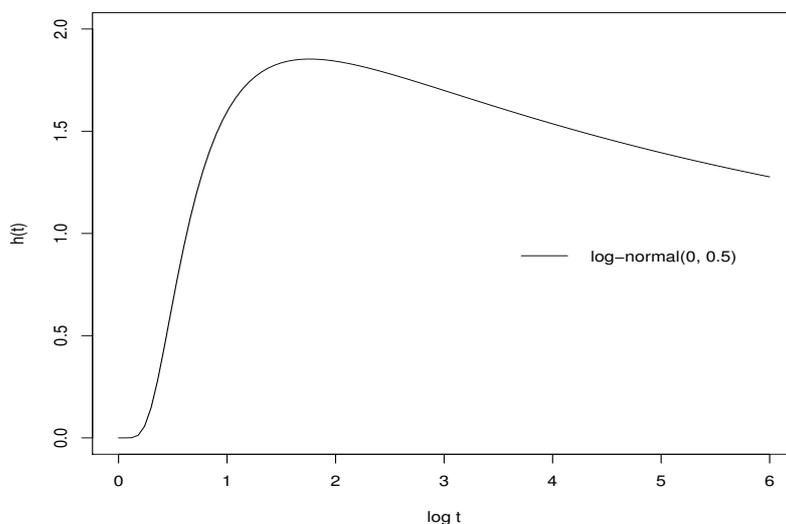


Figura 1.3: gráfico da taxa de risco da distribuição log-normal para  $\mu = 0$  e valores distintos do parâmetro  $\sigma$ .

**Prova:** da equação (1.4), temos

$$h(t) = -\frac{1}{S(t)} \frac{d}{dt} S(t).$$

Usando a Propriedade 1.2, obtemos

$$h(t) = \frac{f(t)}{S(t)}.$$

Por outro, fazendo uso de propriedades da função logarítmica, concluímos que

$$h(t) = -\frac{1}{S(t)} \frac{d}{dt} S(t) = -\frac{d}{dt} \log S(t).$$

■

**Propriedade 1.4** *A taxa de risco  $h$  de uma dada distribuição de tempo de vida satisfaz as condições:*

(i)  $h(t) \geq 0$  para todo  $t \in \mathbb{R}$ ; e

(ii)  $\int_0^{+\infty} h(t) dt = +\infty$ .

**Prova:** a condição (i) é uma consequência imediata do fato de que a taxa risco é dada pela razão entre duas funções não-negativas. Mostremos, agora, a condição (ii). Por definição,

$$\int_0^{+\infty} h(t) dt = \lim_{x \rightarrow +\infty} \int_0^x h(t) dt.$$

Pela Propriedade 1.3, temos

$$\int_0^x h(t) dt = \int_0^x -\frac{d}{dt} \log S(t) dt = -\log S(x).$$

Portanto,

$$\int_0^{+\infty} h(t) dt = \lim_{x \rightarrow +\infty} \int_0^x h(t) dt = \lim_{x \rightarrow +\infty} [-\log S(x)] = +\infty.$$

■

**Propriedade 1.5** *Se  $T$  é uma variável aleatória absolutamente contínua e não-negativa, então*

$$S(t) = \exp \left( - \int_0^t h(u) du \right).$$

**Prova:** pela Propriedade 1.3,

$$h(t) = -\frac{d}{dt} \log S(t) \tag{1.5}$$

Integrando ambos os membros da equação (1.5) de 0 até  $t$ , temos

$$\int_0^t h(u) du = \int_0^t -\frac{d}{du} \log S(u) du = -\log S(t).$$

Portanto,

$$S(t) = \exp \left( - \int_0^t h(u) du \right).$$

■

Outra função útil na caracterização do tempo de sobrevivência é a função de risco acumulativa, indicada por  $H$  e definida por

$$H(t) = \int_0^t h(u) du.$$

**Propriedade 1.6** *A função de risco acumulativa de uma dada distribuição  $T$  de tempo de falha satisfaz as seguintes propriedades:*

- (i)  $H(t) \geq 0 \forall t \in \mathbb{R}$ ;
- (ii)  $\lim_{t \rightarrow +\infty} H(t) = +\infty$ ;
- (iii)  $S(t) = \exp(-H(t))$ .

**Prova:** das três propriedades citadas acima, as duas primeiras decorrem imediatamente da Propriedade 1.4 e a última é consequência da Propriedade 1.5, combinada com a definição da função de risco acumulativa. ■

## 1.2.2 Modelos discretos

Podemos nos deparar com situações em que a variável aleatória que indica o tempo de sobrevivência é discreta. Por exemplo, o número de ciclos efetuados por um motor até a ocorrência de falha. Seja  $T$  uma variável aleatória discreta, assumindo os valores  $t_1, t_2, \dots$  com  $0 \leq t_1 \leq t_2 \leq \dots \leq t_j \leq t_{j+1} \leq \dots$  e cuja função de probabilidade é

$$p(t_j) = P(T = t_j), \quad j = 1, 2, \dots$$

A função de sobrevivência de  $T$  é definida, segundo Lawless (1982), por

$$S(t) = P(T \geq t) = \sum_{j: t_j \geq t} p(t_j), \quad \forall t \in \mathbb{R}.$$

**Propriedade 1.7** *Se  $T$  é uma variável discreta, então*

$$p(t_j) = S(t_j) - S(t_{j+1}), \quad j = 1, 2, \dots$$

**Prova:** uma vez que,  $0 \leq t_1 \leq t_2 \leq \dots \leq t_j \leq t_{j+1} \leq \dots$ , temos

$$P(T \geq t_j) = P(T = t_j) + P(T \geq t_{j+1}).$$

Ou seja,

$$P(T = t_j) = P(T \geq t_j) - P(T \geq t_{j+1}).$$

Portanto,

$$p(t_j) = S(t_j) - S(t_{j+1}), \quad j = 1, 2, \dots$$



Para o caso discreto, Lawless (1982) define a função de risco por

$$h(t_j) = P(T = t_j | T \geq t_j) ; j = 1, 2, \dots \quad (1.6)$$

Uma vez que

$$P(T = t_j | T \geq t_j) = \frac{P(T = t_j, T \geq t_j)}{P(T \geq t_j)} = \frac{P(T = t_j)}{P(T \geq t_j)} = \frac{p(t_j)}{S(t_j)} , j = 1, 2, \dots ,$$

podemos reescrever a expressão (1.6) como

$$h(t_j) = \frac{p(t_j)}{S(t_j)} , j = 1, 2, \dots$$

**Propriedade 1.8** *Se  $T$  é uma variável aleatória discreta, então*

$$h(t_j) = 1 - \frac{S(t_{j+1})}{S(t_j)} , j = 1, 2, \dots$$

**Prova:** seja  $T$  uma variável aleatória discreta que assume os valores  $t_1, t_2, \dots, t_j, \dots$  tais que  $0 \leq t_1 \leq t_2 \leq \dots \leq t_j \leq t_{j+1} \leq \dots$ . Por definição,

$$h(t_j) = \frac{p(t_j)}{S(t_j)} , j = 1, 2, \dots$$

Pela Propriedade 1.7

$$p(t_j) = S(t_j) - S(t_{j+1}) , j = 1, 2, \dots$$

Logo,

$$h(t_j) = \frac{S(t_j) - S(t_{j+1})}{S(t_j)} = 1 - \frac{S(t_{j+1})}{S(t_j)} , j = 1, 2, \dots$$



## 1.3 Função de verossimilhança, função escore e matriz de informação de Fisher

### 1.3.1 Função de verossimilhança

Seja  $T$  uma variável aleatória cuja função densidade de probabilidade pertence a determinada família  $\mathcal{F}_\theta$  de distribuições, na qual o parâmetro  $\theta$  é, em geral, desconhecido. Denotamos por  $f_T(\theta)$  a fdp de  $T$  com  $\theta \in \Theta \subset \mathbb{R}^q$  e por  $S_T(\theta)$  a função de sobrevivência de  $T$ . Sejam  $t_1, t_2, \dots, t_n$  observações casuais e independentes da variável aleatória  $T$ , com  $r$  falhas e  $n - r$  censuras, observadas ao término do experimento. A função de verossimilhança de  $\theta$ , baseada nas  $n$  observações  $t_1, t_2, \dots, t_n$ , em que as  $r$  primeiras ordenadas são não-censuradas e as demais são censuradas, é definida por

$$L(\theta) \propto \prod_{i=1}^r f_T(t_i, \theta) \prod_{i=r+1}^n S_T(t_i, \theta). \quad (1.7)$$

Na definição dada pela equação (1.7), a contribuição de cada observação não-censurada é dada por meio da função de densidade. Por outro lado, a contribuição de cada observação censurada é dada pela função de sobrevivência. A constante positiva de proporcionalidade pode depender das observações  $t_1, t_2, \dots, t_n$ , mas não do parâmetro  $\theta$ .

A função de verossimilhança deve ser interpretada como uma função do vetor de parâmetros  $\theta$  para um conjunto de dados fixo, e serve para medir o quanto aqueles dados suportam uma hipótese sobre  $\theta$ . Ela expressa a plausividade do vetor  $\theta$  em explicar os dados  $t'_i$ s; neste sentido, a mesma informa a ordem natural de preferência entre diversas possibilidades de  $\theta$ , (ver Cordeiro, 1992).

A seguir, apresentaremos a função de verossimilhança, considerando tipos distintos de censura.

- Censura do tipo I.

Neste caso,

$$S_T(t_i, \theta) = S_T(L, \theta),$$

para  $i = r + 1, r + 2, \dots, n$ . Conseqüentemente,

$$L(\boldsymbol{\theta}) \propto \left[ \prod_{i=1}^r f_T(t_i, \boldsymbol{\theta}) \right] [S_T(L, \boldsymbol{\theta})]^{n-r}.$$

- Censura do tipo II.

Nesta situação fixa-se  $r$  e somente os  $r$  menores tempos são observados. Baseando-se em resultados relacionados com estatísticas de ordem (Veja Apêndice B) tem-se

$$L(\boldsymbol{\theta}) \propto \frac{n!}{(n-r)!} \left[ \prod_{i=1}^r f_T(t_i, \boldsymbol{\theta}) \right] [S_T(t_{(r)}, \boldsymbol{\theta})]^{n-r},$$

em que  $t_{(r)}$  denota o maior tempo observado. Notamos que o fator  $n!/(n-r)!$  independe do parâmetro  $\boldsymbol{\theta}$ , logo pode ser desprezado da expressão acima para fins de estimação.

- Censura do tipo aleatória.

Neste caso,  $T$  é o tempo até a ocorrência de falha e  $C$  é o tempo de censura. Para  $i = 1, 2, \dots, n$ , os dados observados consistem em pares  $(t_i, \delta_i)$  em que  $t_i = \min\{T_i, C_i\}$  e  $\delta_i = 1$  se  $T_i \leq C_i$  ou  $\delta_i = 0$  se  $T_i > C_i$ . Considerando os tempos de falha e censura independentes e denotando por  $g$  e  $G$  as funções de densidade e de sobrevivência de  $C$ , respectivamente, então

$$L(\boldsymbol{\theta}) \propto \prod_{i=1}^r f_T(t_i, \boldsymbol{\theta}) G(t_i) \prod_{i=r+1}^n g(t_i) S_T(t_i, \boldsymbol{\theta}).$$

Sob a hipótese de que o mecanismo de censura é não-informativo, isto é, o mesmo não carrega informações sobre os parâmetros, os fatores  $G(t)$  e  $g(t)$  independem de  $\boldsymbol{\theta}$ , portanto podem ser desprezados para fins de estimação.

Segue da exposição acima que a função de verossimilhança para os mecanismos de censura apresentados pode ser tomada da forma

$$L(\boldsymbol{\theta}) \propto \prod_{i \in D} f_T(t_i, \boldsymbol{\theta}) \prod_{i \in C} S_T(t_i, \boldsymbol{\theta}), \quad (1.8)$$

em que  $D$  e  $C$  denotam os conjuntos das unidades sob análise para os quais  $t_i$  é o tempo de sobrevivência ou tempo censurado, respectivamente.

A função de log-verossimilhança é definida por

$$\ell(\boldsymbol{\theta}) = \log(L(\boldsymbol{\theta})).$$

Assim, a partir da expressão (1.8), temos

$$\ell(\boldsymbol{\theta}) \propto \sum_{i \in D} \log(f_T(t_i, \boldsymbol{\theta})) + \sum_{i \in C} \log(S_T(t_i, \boldsymbol{\theta})).$$

### 1.3.2 Função escore e matriz de informação de Fisher

Definimos a função escore por

$$U(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

O estimador de máxima verossimilhança é o valor de  $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^q$  que maximiza a função de verossimilhança ou, equivalentemente, a função de log-verossimilhança. Portanto, o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ , denotado por  $\hat{\boldsymbol{\theta}}$ , deve satisfazer ao seguinte sistema de equações

$$\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \mathbf{0}.$$

A matriz de informação de Fisher esperada de  $\boldsymbol{\theta}$  contida nos dados  $t'_i$ s é a matriz  $q \times q$  definida por

$$K(\boldsymbol{\theta}) = E[U(\boldsymbol{\theta})U^\top(\boldsymbol{\theta})].$$

A esperança e a covariância da função escore são tais que

$$E[U(\boldsymbol{\theta})] = \mathbf{0}$$

e

$$\text{Cov}[U(\boldsymbol{\theta})] = E\left(-\frac{\partial U^\top(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = E\left(-\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\right) = K(\boldsymbol{\theta}).$$

A matriz de primeiras derivadas da função escore com sinal negativo, ou seja,

$$-\frac{\partial U^\top(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}$$

é denominada matriz de informação observada e a denotaremos por  $-\ddot{L}$ .

## Capítulo 2

# O modelo de regressão log-Birnbaum-Saunders

Em algumas situações experimentais, o comportamento da variável de interesse, denominada de resposta, pode ser explicado por meio de uma estrutura com covariáveis e parâmetros desconhecidos. Este é o contexto de um modelo de regressão. Por exemplo, Lawless (1982) cita um estudo considerado em Krall, Uthoff & Harley (1975) no qual a variável de interesse, o tempo de vida de pacientes com mieloma múltiplo, está relacionada com as seguintes covariáveis: sexo do paciente, idade do indivíduo no diagnóstico, medida de hemoglobina, dentre outras.

Dado que as distribuições de tempo de sobrevivência possuem, em geral, assimetria à direita, alguns modelos, por exemplo, o modelo de regressão normal linear, revela-se inadequado nestas situações. Com o objetivo de solucionar essa dificuldade é comum utilizarmos uma componente determinística não-linear combinada com uma distribuição assimétrica para a componente estocástica, ou transformarmos a resposta a fim de tornar possível a utilização de um modelo conhecido.

Neste capítulo, consideraremos um modelo de regressão linear cujos erros seguem uma distribuição log-Birnbaum-Saunders. A distribuição Birnbaum-Saunders é uma importante distribuição de vida (assimétrica) originada de um problema de fadiga

de materiais que foi desenvolvida por Birnbaum & Saunders (1969a), ver Johnson, Kotz & Balakrishnan (1995, p. 651) e Saunders (2007). A distribuição Birnbaum-Saunders (BS) relaciona o tempo até a ocorrência de alguma falha no material com algum dano cumulativo assumido gaussiano. A distribuição BS tem sido largamente aplicada na área de engenharia, com inúmeros trabalhos publicados. Contudo, devido aos argumentos teóricos utilizados na construção dessa distribuição, é natural encontrar aplicações em outras áreas, tais como áreas médica, de ciências atuariais e de meio ambiente, em que algum tipo de estresse cumulativo pode levar à ocorrência de um evento, ver Podlasky (2008), Leiva, Sanhueza & Angulo (2008) e Leiva, Sanhueza & Saunders (2008).

A distribuição BS é definida em termos da distribuição normal padrão. Para maiores detalhes desta distribuição, ver Marshal & Olkin (2007). Uma derivação mais geral desta distribuição foi feita por Desmond (1985), que mostra que esta distribuição pode ser obtida de diferentes modelos além da normal. Díaz-García & Leiva (2005) geraram a distribuição BS a partir de distribuições elípticas. Este novo modelo é chamado distribuição Birnbaum-Saunders generalizada (BSG). O modelo BSG admite diferentes graus de curtose e assimetria e possui unimodalidade e bimodalidade. Adicionalmente, as distribuições BSG englobam modelos em que o método de máxima verossimilhança produz estimativas para os parâmetros que geralmente não são sensíveis a *outliers*, como foi mostrado por Leiva, Riquelme, Balakrishnan & Sanhueza (2008) e Barros, Paula & Leiva (2008); para outras propostas de estimação nos modelos BSG ver Leiva, Barros, Paula & Sanhueza (2008) e Leiva, Sanhueza, Sen & Paula (2008). Além disso, há argumentos teóricos que justificam o uso das distribuições BSG nos problemas que envolvem dano cumulativo que ocorrem comumente em problemas de engenharia, em problemas ambientais e médicos. Sanhueza, Leiva & Balakrishnan (2008) apresentaram um estudo completo dos resultados e das aplicações relativos às distribuições BSG, mostrando que estes modelos são úteis em diversos campos devido a seus argumentos teóricos.

Diversas propostas de estimação têm sido postuladas por Birnbaum & Saunders (1969b), Engelhardt, Bain & Wright (1981), Ahmad (1988), Achcar (1993), Chang & Tang (1994), Dupuis & Mills (1998), Ng, Kundub & Balakrishnan (2003) e From &

Li (2006). Recentemente, métodos de inferência aprimorados para a distribuição BS foram propostos por Lemonte, Cribari-Neto & Vasconcellos (2007), Lemonte, Simas & Cribari-Neto (2008) e Cysneiros, Cribari-Neto & Araujo (2008). Modelos de regressão BS podem ser usados numa grande variedade de aplicações. Por exemplo, comparar os parâmetros de locação de várias populações ou para avaliar efeitos ambientais e de controle da variável em situações de teste de vida. Modelagem estatística sob distribuição BS vem recebido muita atenção ultimamente. Alguns trabalhos têm sido desenvolvidos, os quais podem ser vistos em Rieck & Nedelman (1991), Owen & Padgett (1999, 2000) e Tsionas (2001). Alguns aspectos relacionados à análise de diagnóstico para modelos de regressão log-BS podem ser visto em Rieck & Nedelman (1991), Tsionas (2001), Galea, Leiva & Paula (2004), Leiva, Barros, Paula & Galea (2007) e Xie & Wei (2007).

## 2.1 A distribuição Birnbaum-Saunders

As distribuições paramétricas comumente empregadas para modelar o tempo até a ocorrência de falha em processos de fadiga são a normal inversa, gama e a lognormal. Birnbaum-Saunders (1969a) nos chamam atenção para o fato de que embora essas distribuições se ajustem satisfatoriamente na região central da distribuição do tempo de vida, nas caudas o ajuste é insatisfatório. Diante disso, eles propuseram uma nova família de distribuições para modelar o tempo de vida de materiais sob processos de fadiga. A seguir, apresentaremos essa família de distribuições, conhecida na literatura estatística como a distribuição Birnbaum-Saunders, e mostraremos algumas de suas propriedades.

Definimos a distribuição BS a partir da distribuição normal padrão por meio de uma variável aleatória  $T$  definida por:

$$T = \beta \left[ \frac{\alpha Z}{2} + \left[ \left[ \frac{\alpha Z}{2} \right]^2 + 1 \right]^{\frac{1}{2}} \right]^2,$$

em que variável aleatória  $Z = \frac{1}{\alpha} \left[ \left[ \frac{t}{\beta} \right]^{\frac{1}{2}} - \left[ \frac{\beta}{t} \right]^{\frac{1}{2}} \right]$  segue uma distribuição normal padrão, e  $\alpha > 0$  e  $\beta > 0$ . Neste caso, dizemos que a variável  $T$  possui distribuição BS com

parâmetros  $\alpha$  e  $\beta$ , que são parâmetros de forma e escala, respectivamente, e indicamos  $T \sim \text{BS}(\alpha, \beta)$ .

A função de distribuição acumulativa (fda) da variável  $T \sim \text{BS}(\alpha, \beta)$  é dada por

$$F_T(t) = \Phi \left( \frac{1}{\alpha} \left[ \left[ \frac{t}{\beta} \right]^{\frac{1}{2}} - \left[ \frac{\beta}{t} \right]^{\frac{1}{2}} \right] \right), \quad t > 0, \quad (2.1)$$

em que  $\Phi(\cdot)$  denota a fda da distribuição normal padrão.

**Proposição 2.1** *Se  $T$  é uma variável aleatória que possui distribuição BS com parâmetros  $\alpha$  e  $\beta$ , então a função de densidade de probabilidade (fdp) de  $T$  é dada por*

$$f_T(t) = \frac{1}{2\alpha\beta\sqrt{2\pi}} \left[ \left[ \frac{\beta}{t} \right]^{\frac{1}{2}} + \left[ \frac{\beta}{t} \right]^{\frac{3}{2}} \right] \exp \left( -\frac{1}{2\alpha^2} \left[ \frac{t}{\beta} + \frac{\beta}{t} - 2 \right] \right), \quad t > 0. \quad (2.2)$$

**Prova:** temos que a fdp de  $T$  é obtida a partir da relação

$$f_T(t) = \frac{d}{dt} F_T(t),$$

em que  $F_T(t)$  é dado pela expressão (2.1). Logo,

$$\begin{aligned} f_T(t) &= \frac{d}{dt} \Phi \left( \frac{1}{\alpha} \left[ \left[ \frac{t}{\beta} \right]^{\frac{1}{2}} - \left[ \frac{\beta}{t} \right]^{\frac{1}{2}} \right] \right) \\ &= \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{1}{2} \left[ \frac{1}{\alpha} \left[ \left[ \frac{t}{\beta} \right]^{\frac{1}{2}} - \left[ \frac{\beta}{t} \right]^{\frac{1}{2}} \right] \right]^2 \right) \frac{1}{2\alpha\beta} \left[ \left[ \frac{\beta}{t} \right]^{\frac{1}{2}} + \left[ \frac{\beta}{t} \right]^{\frac{3}{2}} \right] \\ &= \frac{1}{2\alpha\beta\sqrt{2\pi}} \left[ \left[ \frac{\beta}{t} \right]^{\frac{1}{2}} + \left[ \frac{\beta}{t} \right]^{\frac{3}{2}} \right] \exp \left( -\frac{1}{2\alpha^2} \left[ \frac{t}{\beta} + \frac{\beta}{t} - 2 \right] \right), \quad t > 0. \end{aligned}$$

■

A Figura 2.1 ilustra o comportamento do gráfico da f.d.p. da distribuição BS para diferentes valores dos parâmetros  $\alpha$  e  $\beta = 1$ . Como podemos observar, o parâmetro  $\alpha$  controla a forma da distribuição. A distribuição BS torna-se assimétrica à medida que  $\alpha$  cresce, e simétrica, em torno de  $\beta$ , à medida que  $\alpha$  se aproxima de zero.

Agora, mostraremos algumas propriedades de uma variável que segue uma distribuição Birnbaum-Saunders.

**Propriedade 2.1** *Se  $T \sim \text{BS}(\alpha, \beta)$ , então  $cT \sim \text{BS}(\alpha, c\beta)$  para todo número real  $c > 0$ .*

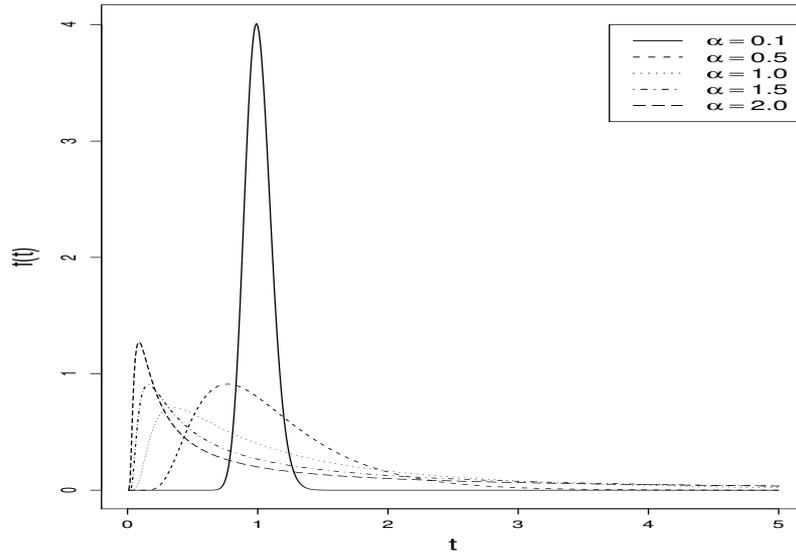


Figura 2.1: gráfico da fdp da distribuição BS para os valores indicados de  $\alpha$  e  $\beta = 1$ .

**Prova:** de fato, suponhamos que a variável aleatória  $Y = cT$ , com  $c > 0$ . Conseqüentemente,

$$F_Y(t) = P(Y \leq t) = P\left(T \leq \frac{t}{c}\right) = F_T\left(\frac{t}{c}\right) = \Phi\left(\frac{1}{\alpha} \left[ \left[\frac{t}{c\beta}\right]^{\frac{1}{2}} - \left[\frac{c\beta}{t}\right]^{\frac{1}{2}} \right]\right).$$

Portanto,  $Y = cT \sim BS(\alpha, c\beta)$ . ■

**Propriedade 2.2** Se  $T \sim BS(\alpha, \beta)$ , então  $1/T \sim BS(\alpha, 1/\beta)$ .

**Prova:** com efeito, definamos a variável  $Y = 1/T$ , temos

$$F_Y(t) = P(Y \leq t) = P(1/T \leq t) = P\left(T \geq \frac{1}{t}\right) = 1 - F_T\left(\frac{1}{t}\right).$$

Conseqüentemente,

$$\begin{aligned} f_Y(t) &= \frac{d}{dt} F_Y(t) = \frac{1}{t^2} f_T\left(\frac{1}{t}\right) \\ &= \frac{1}{t^2} \frac{1}{2\alpha\beta\sqrt{2\pi}} \left[ \left[\frac{\beta}{t^{-1}}\right]^{\frac{1}{2}} + \left[\frac{\beta}{t^{-1}}\right]^{\frac{3}{2}} \right] \exp\left(-\frac{1}{2\alpha^2} \left[\frac{t^{-1}}{\beta} + \frac{\beta}{t^{-1}} - 2\right]\right) \\ &= \frac{1}{2\alpha\beta^{-1}\sqrt{2\pi}} \left[ \left[\frac{\beta^{-1}}{t}\right]^{\frac{1}{2}} + \left[\frac{\beta^{-1}}{t}\right]^{\frac{3}{2}} \right] \exp\left(-\frac{1}{2\alpha^2} \left[\frac{t}{\beta^{-1}} + \frac{\beta^{-1}}{t} - 2\right]\right), \end{aligned}$$

$t > 0$ , o que implica em  $Y = 1/T \sim BS(\alpha, 1/\beta)$ . ■

**Propriedade 2.3** Se  $T \sim BS(\alpha, \beta)$ , então  $\text{Mediana}(T) = \beta$ .

**Prova:** se  $T \sim BS(\alpha, \beta)$ , então a fda de  $T$  é dada por

$$F_T(t) = \Phi \left( \frac{1}{\alpha} \left[ \left( \frac{t}{\beta} \right)^{\frac{1}{2}} - \left( \frac{\beta}{t} \right)^{\frac{1}{2}} \right] \right), \quad t > 0.$$

Conseqüentemente,

$$F_T(\beta) = \Phi \left( \frac{1}{\alpha} \left[ \left( \frac{\beta}{\beta} \right)^{\frac{1}{2}} - \left( \frac{\beta}{\beta} \right)^{\frac{1}{2}} \right] \right) = \Phi(0) = \frac{1}{2}.$$

Portanto,  $\beta = \text{Mediana}(T)$ . ■

A função de sobrevivência da variável  $T \sim BS(\alpha, \beta)$  é dada por

$$S(t) = 1 - \Phi \left( \frac{1}{\alpha} \left[ \left( \frac{t}{\beta} \right)^{\frac{1}{2}} - \left( \frac{\beta}{t} \right)^{\frac{1}{2}} \right] \right) = \Phi \left( -\frac{1}{\alpha} \left[ \left( \frac{t}{\beta} \right)^{\frac{1}{2}} - \left( \frac{\beta}{t} \right)^{\frac{1}{2}} \right] \right), \quad t > 0. \quad (2.3)$$

A Figura 2.2 apresenta o gráfico da função de sobrevivência da distribuição BS para valores específicos dos parâmetros  $\alpha$  e  $\beta = 1$ .

A partir da relação  $h(t) = f(t)/S(t)$ , podemos analisar o comportamento da taxa de risco da distribuição Birnbaum-Saunders, pois  $f(t)$  e  $S(t)$  são dadas pelas expressões (2.2) e (2.3), respectivamente. Observemos através da Figura 2.3 que a taxa de risco  $h(t)$  assume valor zero em  $t = 0$ , cresce até um valor máximo e depois decresce até uma constante positiva.

## 2.2 A distribuição senh-normal

Nesta seção, apresentaremos a distribuição senh-normal (SN) desenvolvida por Rieck & Nedelman (1991). Em seguida, exibiremos algumas propriedades de uma variável aleatória que segue uma distribuição senh-normal e, por fim, mostraremos uma relação existente entre essa distribuição e a distribuição BS sob uma transformação logarítmica.

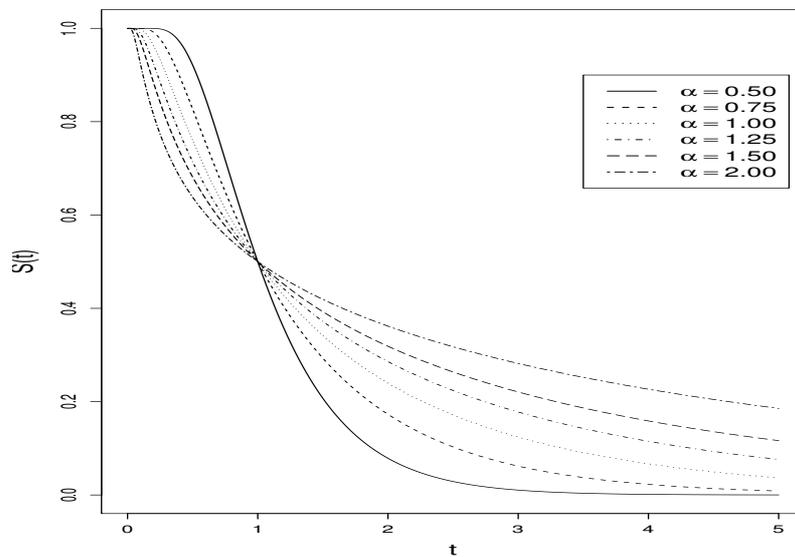


Figura 2.2: gráfico da função de sobrevivência da distribuição BS para os valores indicados de  $\alpha$  e  $\beta = 1$ .

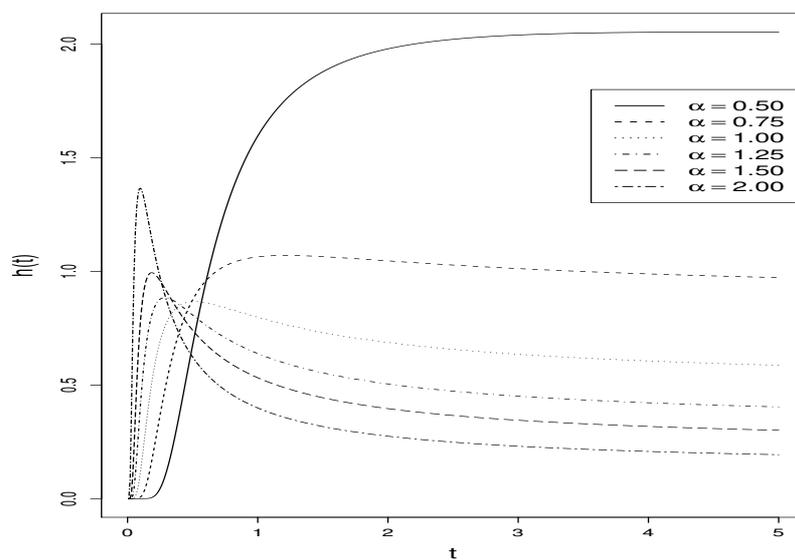


Figura 2.3: gráfico da taxa de risco da distribuição BS para os valores indicados do parâmetro  $\alpha$  e  $\beta = 1$ .

Seja  $Y$  uma variável aleatória tal que a variável  $Z = [2\sinh(\frac{Y-\gamma}{\sigma})]/\alpha$  segue uma distribuição normal padrão. Neste caso, dizemos que a variável

$$Y = \operatorname{arcsenh}\left(\frac{\alpha Z}{2} \sigma\right) + \gamma$$

tem distribuição *senh-normal* com parâmetros de forma, locação e escala, dados respectivamente por,  $\alpha > 0$ ,  $\gamma \in \mathbb{R}$  e  $\sigma > 0$  e indicamos  $Y \sim \operatorname{SN}(\alpha, \gamma, \sigma)$ .

A fda de uma variável  $Y \sim \operatorname{SN}(\alpha, \gamma, \sigma)$  é dada por

$$F_Y(y) = \Phi\left(\frac{2}{\alpha} \operatorname{senh}\left(\frac{y-\gamma}{\sigma}\right)\right), y \in \mathbb{R}.$$

**Propriedade 2.4** *Se  $Y \sim \operatorname{SN}(\alpha, \gamma, \sigma)$ , então a fdp de  $Y$  tem a forma*

$$f_Y(y) = \frac{2}{\alpha\sigma\sqrt{2\pi}} \cosh\left(\frac{y-\gamma}{\sigma}\right) \exp\left(-\frac{2}{\alpha^2} \left[\operatorname{senh}\left(\frac{y-\gamma}{\sigma}\right)\right]^2\right), y \in \mathbb{R}.$$

**Prova:** temos que

$$\begin{aligned} f_Y(y) &= \frac{d}{dy} F_Y(y) = \frac{d}{dy} \Phi\left(\frac{2}{\alpha} \operatorname{senh}\left(\frac{y-\gamma}{\sigma}\right)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{2}{\alpha} \operatorname{senh}\left(\frac{y-\gamma}{\sigma}\right)\right]^2\right) \frac{2}{\alpha\sigma} \cosh\left(\frac{y-\gamma}{\sigma}\right) \\ &= \frac{2}{\alpha\sigma\sqrt{2\pi}} \cosh\left(\frac{y-\gamma}{\sigma}\right) \exp\left(-\frac{2}{\alpha^2} \left[\operatorname{senh}\left(\frac{y-\gamma}{\sigma}\right)\right]^2\right), y \in \mathbb{R}. \end{aligned}$$

■

Na Figura 2.4, observemos o comportamento das funções densidade de probabilidade das distribuições *senh-normal* e normal. Como podemos observar, quando  $\alpha$  cresce, a curtose da distribuição SN também cresce. Em particular, para  $\alpha = 2$ , temos unimodalidade e curtose menor do que o caso normal. Entretanto, para  $\alpha > 2$ , à medida que  $\alpha$  cresce, a distribuição SN começa a apresentar bimodalidade, possui modas que são mais separadas e curtose maior do que o caso normal. O parâmetro  $\gamma$  modifica a locação e o parâmetro  $\sigma$  a escala da distribuição.

A seguir, apresentaremos algumas propriedades de uma variável aleatória  $Y \sim \operatorname{SN}(\alpha, \gamma, \sigma)$ .

**Propriedade 2.5** *Se  $Y \sim \operatorname{SN}(\alpha, \gamma, \sigma)$  e  $a, b \in \mathbb{R}$ , com  $b > 0$ , então a variável aleatória  $W = a + bY \sim \operatorname{SN}(\alpha, a + b\gamma, |b|\sigma)$ .*

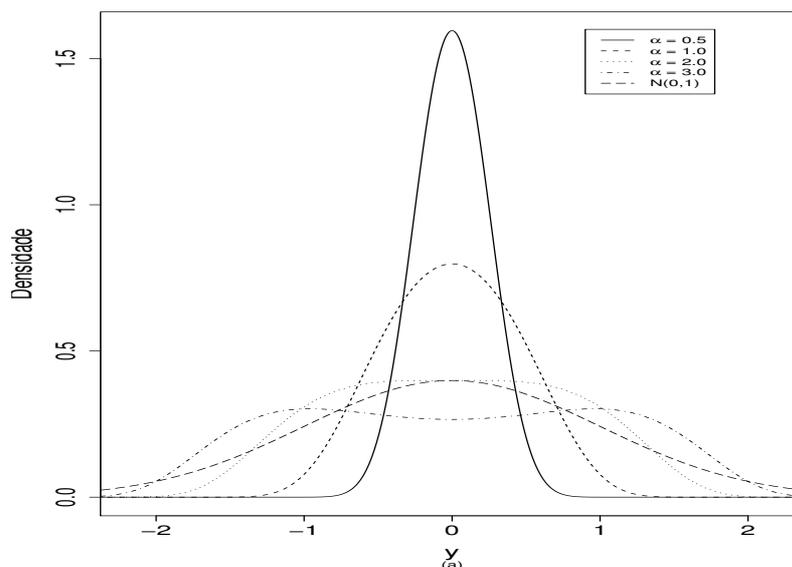


Figura 2.4: gráfico da fdp da distribuição normal e da distribuição senh-normal para os valores indicados do parâmetro  $\alpha$ ,  $\gamma = 0$  e  $\sigma = 2$ .

**Propriedade 2.6** Se  $Y \sim SN(\alpha, \gamma, \sigma)$ , então  $Y$  é simétrica em torno de  $\gamma$ .

**Propriedade 2.7** A variável aleatória  $Y \sim SN(\alpha, \gamma, \sigma)$  é unimodal para  $\alpha \leq 2$  e bimodal para  $\alpha > 2$ .

**Propriedade 2.8** A variável aleatória  $S_\alpha = 2 \left[ \frac{Y - \gamma}{\alpha \sigma} \right]$  converge em distribuição para a normal padrão quando  $\alpha$  aproxima-se de zero.

**Propriedade 2.9** A função geradora de momentos da variável aleatória  $Y \sim SN(\alpha, \gamma, \sigma)$  é dada por

$$m(s) = \exp(\mu s) \left[ \frac{K_a(\delta^{-2}) + K_b(\delta^{-2})}{2K_{\frac{1}{2}}(\delta^{-2})} \right],$$

em que  $a = [\sigma s + 1]/2$ ,  $b = [\sigma s - 1]/2$  e  $K(\cdot)$  é a função de Bessel modificada do terceiro tipo, definida por  $K_\lambda(w) = \frac{1}{2} \left( \frac{w}{2} \right)^\lambda \int_0^{+\infty} y^{-\lambda-1} e^{-y - \frac{w^2}{4y}} dy$ . (Para mais detalhes sobre a função de Bessel veja Gradshteyn & Ryzhik, 2000).

**Propriedade 2.10** Se  $Y \sim SN(\alpha, \gamma, \sigma)$ , então  $E(Y) = \gamma$  e  $Var(Y) = \sigma^2 g(\alpha)$ , em que  $g(\alpha)$  é a variância quando  $\sigma = 1$ .

A maioria das propriedades citadas encontra-se demonstradas em Rieck (1989).

Rieck & Nedelman (1991) mostraram que há uma relação entre a distribuição SN e a distribuição BS sob uma transformação logarítmica. A proposição a seguir explicita tal relação.

**Proposição 2.2** *Se  $T$  é uma variável aleatória cuja distribuição é BS com parâmetros de forma e escala iguais a  $\alpha$  e  $\beta$ , respectivamente, então a variável  $Y = \log(T)$  possui distribuição SN com parâmetros de forma, locação e escala dados por  $\alpha$ ,  $\gamma = \log(\beta)$  e  $\sigma = 2$ , respectivamente.*

**Prova:** seja  $T$  uma variável aleatória tal que  $T \sim \text{BS}(\alpha, \beta)$ , conseqüentemente a função de distribuição acumulada de  $T$  é dada por:

$$F_T(t) = \Phi \left( \frac{1}{\alpha} \left[ \left( \frac{t}{\beta} \right)^{\frac{1}{2}} - \left( \frac{\beta}{t} \right)^{\frac{1}{2}} \right] \right), \quad t > 0.$$

Definamos a variável  $Y = \log(T)$ , segue que

$$\begin{aligned} F_Y(y) = F_T(\exp(y)) &= \Phi \left( \frac{1}{\alpha} \left[ \left( \frac{\exp(y)}{\beta} \right)^{\frac{1}{2}} - \left( \frac{\beta}{\exp(y)} \right)^{\frac{1}{2}} \right] \right) \\ &= \Phi \left( \frac{1}{\alpha} \left[ \left( \frac{\exp(y)}{\exp(\log(\beta))} \right)^{\frac{1}{2}} - \left( \frac{\exp(y)}{\exp(\log(\beta))} \right)^{-\frac{1}{2}} \right] \right) \\ &= \Phi \left( \frac{1}{\alpha} \left[ (\exp(y - \log(\beta)))^{\frac{1}{2}} - (\exp(y - \log(\beta)))^{-\frac{1}{2}} \right] \right) \\ &= \Phi \left( \frac{1}{\alpha} \left[ \exp \left( \frac{y - \log(\beta)}{2} \right) - \exp \left( -\frac{y - \log(\beta)}{2} \right) \right] \right) \\ &= \Phi \left( \frac{2}{\alpha} \cdot \frac{1}{2} \left[ \exp \left( \frac{y - \log(\beta)}{2} \right) - \exp \left( -\frac{y - \log(\beta)}{2} \right) \right] \right) \\ &= \Phi \left( \frac{2}{\alpha} \sinh \left[ \frac{y - \log(\beta)}{2} \right] \right), \quad y \in \mathbb{R}. \end{aligned}$$

Portanto,  $Y = \log(T) \sim \text{SN}(\alpha, \log(\beta), 2)$ . ■

## 2.3 O modelo de regressão log-Birnbaum-Saunders

### 2.3.1 Caracterização

Sejam  $T_1, T_2, \dots, T_n$  variáveis aleatórias independentes tais que  $T_i \sim \text{BS}(\alpha_i, \eta_i)$  para  $i = 1, 2, \dots, n$ . Admitamos que a distribuição de cada variável  $T_i$  depende de um vetor

de  $p$  variáveis explicativas  $\mathbf{x}_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$ , em outras palavras, assumiremos que

- (i)  $\eta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$  para  $i = 1, 2, \dots, n$ , em que  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$  é um vetor de parâmetros desconhecidos;
- (ii) O parâmetro de forma  $\alpha_i$  independe do vetor de covariáveis  $\mathbf{x}_i^\top$ , isto é,  $\alpha_i = \alpha$  para  $i = 1, 2, \dots, n$ .

As suposições acima combinadas com a Propriedade 2.1 implicam  $T_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta}) \zeta_i$  em que  $\zeta_i \sim \text{BS}(\alpha, 1)$ , para  $i = 1, 2, \dots, n$ . Considerando  $Y_i = \log(T_i)$  obtemos

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \log(\zeta_i),$$

para  $i = 1, 2, \dots, n$ . Ou seja, temos um modelo de regressão log-linear para distribuição BS cujo componente aleatório  $\varepsilon_i$  é igual a  $\log(\zeta_i)$ .

Agora, usando a Proposição 2.2, que relaciona as distribuições BS e SN, concluímos

$$\varepsilon_i = \log(\zeta_i) \sim \text{SN}(\alpha, 0, 2), i = 1, 2, \dots, n.$$

As hipóteses e conclusões obtidas acima leva-nos ao seguinte modelo de regressão:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \tag{2.4}$$

em que  $Y_i$  consiste no logaritmo neperiano do tempo de sobrevivência ou tempo de censura associado às unidades experimentais;  $\varepsilon_i$ 's são variáveis aleatórias independentes e identicamente distribuídas tais que  $\varepsilon_i \sim \text{SN}(\alpha, 0, 2)$ ,  $\mathbf{x}_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip})$  são valores observados de  $p$  covariáveis e  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$  é o vetor de parâmetros desconhecidos a ser estimado, com  $i = 1, 2, \dots, n$ . Esse modelo é denominado modelo de regressão log-Birnbaum-Saunders (log-BS) e sobre o mesmo interessa-nos estimar o vetor  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$ .

### 2.3.2 Estimação

Sejam  $y_1, y_2, \dots, y_n$  observações casuais e independentes da variável aleatória  $Y_i$  definida pela expressão (2.4). Dado que  $\varepsilon_i \sim \text{SN}(\alpha, 0, 2)$ , para  $i = 1, 2, \dots, n$ , pela

Propriedade 2.5 temos  $Y_i \sim \text{SN}(\alpha, \mathbf{x}_i^\top \boldsymbol{\beta}, 2)$ , para  $i = 1, 2, \dots, n$ . Denotamos por  $D$  o conjunto das unidades experimentais para o qual  $y_i$  representa o logaritmo neperiano do tempo de vida e por  $C$  o conjunto das unidades em que  $y_i$  é o logaritmo neperiano do tempo censurado. Neste trabalho, supomos censura não-informativa e independência entre os tempos de vida e de censura.

A função de log-verossimilhança total de  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$  baseada nas observações  $y_1, y_2, \dots, y_n$  é dada por

$$\ell(\boldsymbol{\theta}) \propto \sum_{i \in D} \log(f_{Y_i}(y_i, \boldsymbol{\theta})) + \sum_{i \in C} \log(S_{Y_i}(y_i, \boldsymbol{\theta})).$$

Uma vez que  $Y_i \sim \text{SN}(\alpha, \mathbf{x}_i^\top \boldsymbol{\beta}, 2)$  para  $i = 1, 2, \dots, n$ , temos

$$\log(f_{Y_i}(y_i, \boldsymbol{\theta})) = \log\left(\frac{2}{\alpha} \cosh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)\right) - \frac{1}{2} \left[ \log(8\pi) + \left[ \frac{2}{\alpha} \sinh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right) \right]^2 \right]$$

e

$$\log(S_{Y_i}(y_i, \boldsymbol{\theta})) = \log(1 - F_{Y_i}(y_i, \boldsymbol{\theta})) = \log\left(1 - \Phi\left(\frac{2}{\alpha} \sinh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)\right)\right).$$

Neste ponto, façamos

$$\xi_{i1} = \frac{2}{\alpha} \cosh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right) \quad (2.5)$$

e

$$\xi_{i2} = \frac{2}{\alpha} \sinh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right), \quad (2.6)$$

para  $i = 1, 2, \dots, n$ . Conseqüentemente, podemos reescrever a função de log-verossimilhança total de  $\boldsymbol{\theta}$  na forma

$$\ell(\boldsymbol{\theta}) \propto \sum_{i \in D} \left\{ \log(\xi_{i1}) - \frac{1}{2} [\log(8\pi) + \xi_{i2}^2] \right\} + \sum_{i \in C} \log(\Phi(-\xi_{i2})). \quad (2.7)$$

A seguir, mostraremos que as derivadas parciais  $\partial \ell(\boldsymbol{\theta}) / \partial \alpha$  e  $\partial \ell(\boldsymbol{\theta}) / \partial \beta_j$  para  $j = 1, 2, \dots, p$  são dadas por

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \frac{1}{\alpha} \left[ \sum_{i \in D} \{\xi_{i2}^2 - 1\} + \sum_{i \in C} \xi_{i2} h(\xi_{i2}) \right] \quad (2.8)$$

e

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i \in D} \left\{ \frac{x_{ij}}{\alpha^2} \sinh(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) - \frac{x_{ij}}{2} \tanh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right) \right\} + \sum_{i \in C} \left\{ \frac{x_{ij}}{2} \xi_{i1} h(\xi_{i2}) \right\} \quad (2.9)$$

para  $j = 1, 2, \dots, p$ , em que  $h(\xi_{i2}) = \phi(\xi_{i2})/\Phi(-\xi_{i2})$ , sendo  $\phi(\cdot)$  e  $\Phi(\cdot)$  representando as fdp e fda da distribuição normal padrão, respectivamente.

De fato, a derivada parcial de  $\ell(\boldsymbol{\theta})$  em relação ao parâmetro  $\alpha$  é dada por

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \sum_{i \in D} \left\{ \frac{1}{\xi_{i1}} \frac{\partial \xi_{i1}}{\partial \alpha} - \xi_{i2} \frac{\partial \xi_{i2}}{\partial \alpha} \right\} - \sum_{i \in C} \left\{ \frac{1}{[\Phi(-\xi_{i2})]} \frac{\partial \Phi(\xi_{i2})}{\partial \alpha} \frac{\partial \xi_{i2}}{\partial \alpha} \right\}. \quad (2.10)$$

Como

$$\frac{\partial \xi_{i1}}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left[ \frac{2}{\alpha} \cosh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2} \right) \right] = -\frac{1}{\alpha} \xi_{i1} \quad (2.11)$$

e

$$\frac{\partial \xi_{i2}}{\partial \alpha} = \frac{\partial}{\partial \alpha} \left[ \frac{2}{\alpha} \sinh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2} \right) \right] = -\frac{1}{\alpha} \xi_{i2}, \quad (2.12)$$

substituindo as derivadas parciais dadas pelas equações (2.11) e (2.12) na equação (2.10), obtemos

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \frac{1}{\alpha} \left[ \sum_{i \in D} \{\xi_{i2}^2 - 1\} + \sum_{i \in C} \xi_{i2} h(\xi_{i2}) \right]. \quad (2.13)$$

Agora, a derivada parcial de  $\ell(\boldsymbol{\theta})$  em relação a  $\beta_j$  para  $j = 1, 2, \dots, p$  é dada por

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i \in D} \left\{ \frac{1}{\xi_{i1}} \frac{\partial \xi_{i1}}{\partial \beta_j} - \xi_{i2} \frac{\partial \xi_{i2}}{\partial \beta_j} \right\} - \sum_{i \in C} \left\{ \frac{1}{[1 - \Phi(\xi_{i2})]} \phi(\xi_{i2}) \frac{\partial \xi_{i2}}{\partial \beta_j} \right\}. \quad (2.14)$$

Mas,

$$\frac{\partial \xi_{i1}}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \left[ \frac{2}{\alpha} \cosh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2} \right) \right] = -\frac{x_{ij}}{2} \xi_{i2} \quad (2.15)$$

e

$$\frac{\partial \xi_{i2}}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \left[ \frac{2}{\alpha} \sinh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2} \right) \right] = -\frac{x_{ij}}{2} \xi_{i1}, \quad (2.16)$$

para  $j = 1, 2, \dots, p$ . Substituindo as derivadas parciais dadas pelas expressões (2.15) e (2.16) na equação (2.14), obtemos

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i \in D} \left\{ -\frac{x_{ij}}{2} \left[ \frac{\xi_{i2}}{\xi_{i1}} - \xi_{i1} \xi_{i2} \right] \right\} + \sum_{i \in C} \left\{ \frac{x_{ij}}{2} \xi_{i1} h(\xi_{i2}) \right\} \quad (2.17)$$

para  $j = 1, 2, \dots, p$ . Neste ponto, observamos que

$$\frac{\xi_{i2}}{\xi_{i1}} = \tanh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2} \right) \quad (2.18)$$

e

$$\xi_{i1} \xi_{i2} = \frac{2}{\alpha^2} \sinh(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}). \quad (2.19)$$

Finalmente, ao substituirmos os resultados dados pelas expressões (2.18) e (2.19) na equação (2.17), obtemos

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \sum_{i \in D} \left\{ \frac{x_{ij}}{\alpha^2} \sinh(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) - \frac{x_{ij}}{2} \tanh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right) \right\} + \sum_{i \in C} \left\{ \frac{x_{ij}}{2} \xi_{i1} h(\xi_{i2}) \right\},$$

para  $j = 1, 2, \dots, p$ .

Os estimadores de máxima verossimilhança dos coeficientes  $\beta'_j$ s e do parâmetro  $\alpha$  no modelo de regressão dado pela expressão (2.4), denotados por  $\hat{\beta}_j$  para  $j = 1, 2, \dots, p$  e  $\hat{\alpha}$ , respectivamente, devem satisfazer ao sistema de equações abaixo:

$$\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} \right|_{\beta_j = \hat{\beta}_j, \alpha = \hat{\alpha}} = 0, \quad j = 1, 2, \dots, p \quad (2.20)$$

e

$$\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} \right|_{\beta = \hat{\beta}, \alpha = \hat{\alpha}} = 0. \quad (2.21)$$

No caso sob estudo, o sistema de equações formado pelas equações (2.20) e (2.21) é não-linear e não possui soluções analíticas, portanto faz necessário o uso de métodos numéricos para obtenção de soluções. Neste trabalho, usaremos o método BFGS (veja detalhes em Press et al, 1992) com derivadas parciais analíticas; esse método vem sendo utilizados por pesquisadores quando se objetiva a maximização da função de log-verossimilhança.

O método de máxima verossimilhança permite a construção de intervalos de confiança para os coeficientes  $\beta'_j$ s e para o parâmetro de forma  $\alpha$  associados ao modelo dado pela equação (2.4); isso devido a duas propriedades importantes dos estimadores de máxima verossimilhança. Uma delas nos assegura que, para grandes amostras, o estimador de máxima verossimilhança de  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha)^\top$ , denotado por  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\alpha})^\top$ , possui distribuição aproximadamente normal  $(p + 1)$ -variada. A outra nos garante que o estimador  $\hat{\boldsymbol{\theta}}$  é assintoticamente não-viesado para  $\boldsymbol{\theta}$  e, à medida que  $n$  cresce, sua matriz de variâncias e covariâncias  $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$  pode ser aproximada pela matriz  $-\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}$ , em que  $\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}$  representa a matriz de informação de Fisher observada.

Para o modelo de regressão considerado neste capítulo, a matriz  $\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ , obtida por Leiva, Barros, Paula & Galea (2007), é dada por

$$\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \ddot{\mathbf{L}}_{\beta\alpha} \\ \ddot{\mathbf{L}}_{\alpha\beta} & \ddot{\mathbf{L}}_{\alpha\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{V} \mathbf{X} & \mathbf{X}^\top \mathbf{K} \\ \mathbf{K}^\top \mathbf{X} & \text{tr}(\mathbf{G}) \end{bmatrix},$$

em que  $\mathbf{G} = \text{diag}\{g_1(\boldsymbol{\theta}), g_2(\boldsymbol{\theta}), \dots, g_n(\boldsymbol{\theta})\}$  com

$$g_i(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\alpha^2} - \frac{3(\xi_{i2})^2}{\alpha^2}, & \text{se } i \in D \\ -\frac{2}{\alpha^2} \xi_{i2} h(\xi_{i2}) - \frac{1}{\alpha^2} (\xi_{i2})^2 h'(\xi_{i2}), & \text{se } i \in C \end{cases};$$

$\mathbf{K} = (k_1(\boldsymbol{\theta}), k_2(\boldsymbol{\theta}), \dots, k_n(\boldsymbol{\theta}))^\top$ , com

$$k_i(\boldsymbol{\theta}) = \begin{cases} -\frac{2}{\alpha^3} \sinh(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}), & \text{se } i \in D \\ -\frac{1}{2\alpha} \xi_{i1} h(\xi_{i2}) - \frac{1}{\alpha^3} \sinh(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}) h'(\xi_{i2}), & \text{se } i \in C \end{cases}$$

e

$\mathbf{V} = \text{diag}\{\nu_1(\boldsymbol{\theta}), \nu_2(\boldsymbol{\theta}), \dots, \nu_n(\boldsymbol{\theta})\}$  com

$$\nu_i(\boldsymbol{\theta}) = \begin{cases} \frac{1}{4} \text{sech}^2\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right) - \frac{1}{\alpha^2} \cosh(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}), & \text{se } i \in D \\ -\frac{1}{4} \xi_{i2} h(\xi_{i2}) - \frac{1}{4} (\xi_{i2})^2 h'(\xi_{i2}), & \text{se } i \in C \end{cases},$$

sendo  $h(\xi_{i2}) = \phi(\xi_{i2})/\Phi(-\xi_{i2})$ ,  $\xi_{i1}$  e  $\xi_{i2}$  são dados pelas equações (2.5) e (2.6), respectivamente, e  $h'$  denota a derivada da função  $h$ .

## Capítulo 3

# Influência local mediante a curvatura normal conforme

O método de influência local, proposto por Cook (1986), é uma técnica de diagnóstico que consiste em avaliar, por meio de uma medida apropriada de influência, a sensibilidade das estimativas fornecidas por um modelo estatístico adotado mediante pequenas perturbações nos dados ou no próprio modelo. Ela baseia-se em conceitos da geometria diferencial para estudar o comportamento local do gráfico de uma função denominada de afastamento pela verossimilhança, mais precisamente, a curvatura normal  $C_{\mathbf{l}}$  em um ponto  $\omega_o$  numa dada direção  $\mathbf{l}$  é obtida. Cook (1986) sugere que examinemos as direções em que a curvatura normal é máxima, pois a partir dessas podemos identificar observações influentes sob o esquema de perturbação no qual sujeitamos o modelo postulado.

Poon & Poon (1999) chamam atenção para o fato de que a curvatura normal pode assumir qualquer valor real e não é invariante sob uma mudança uniforme de escala, ocasionando perda de objetividade no julgamento da grandeza da curvatura. Com o objetivo de solucionar esse problema e, conseqüentemente, aperfeiçoar o método de influência local, Poon & Poon (1999) propõem que façamos uso da curvatura normal conforme, que está relacionada com a curvatura normal, mas assume valores em um

intervalo limitado da reta e é invariante sob uma classe de reparametrizações denominadas reparametrizações conformes. A curvatura normal conforme e suas propriedades nos dão suporte para construirmos valores de referência que permitem julgarmos a grandeza dessa curvatura.

### 3.1 Influência local, gráfico de influência e curvatura normal

Seja  $\ell(\boldsymbol{\theta})$  a função de log-verossimilhança para um modelo estatístico postulado, em que  $\boldsymbol{\theta} \in \mathbb{R}^q$  é um vetor  $q \times 1$  de parâmetros desconhecidos. Denotemos por  $\ell(\boldsymbol{\theta} | \boldsymbol{\omega})$  a função de log-verossimilhança para o modelo perturbado em que  $\boldsymbol{\omega}^\top = (\omega_1, \omega_2, \dots, \omega_n)$  é um vetor de perturbação em um aberto  $\Omega \subset \mathbb{R}^n$ . Assumimos que existe um  $\boldsymbol{\omega}_o \in \Omega$ , denominado vetor de não-perturbação, tal que  $\ell(\boldsymbol{\theta} | \boldsymbol{\omega}_o) = \ell(\boldsymbol{\theta})$ , para todo  $\boldsymbol{\theta} \in \mathbb{R}^q$ , isto é, estamos admitindo que o modelo postulado está encaixado no modelo perturbado. Prosseguindo, sejam  $\hat{\boldsymbol{\theta}}$  e  $\hat{\boldsymbol{\theta}}_\omega$  os estimadores de máxima verossimilhança para  $\boldsymbol{\theta}$  nos modelos postulado, e perturbado, respectivamente. Finalmente, admitimos que  $\ell(\boldsymbol{\theta} | \boldsymbol{\omega})$  é de classe  $C^2$  (possui derivadas de segunda ordem contínua) em  $(\boldsymbol{\theta}^\top, \boldsymbol{\omega}^\top)^\top$ .

Uma comparação entre os estimadores  $\hat{\boldsymbol{\theta}}$  e  $\hat{\boldsymbol{\theta}}_\omega$  quando  $\boldsymbol{\omega}$  varia em  $\Omega$  irá nos auxiliar na verificação da “robustez” das estimativas obtidas, no que diz respeito ao esquema de perturbação utilizado. Mais especificamente, se a distância entre  $\hat{\boldsymbol{\theta}}$  e  $\hat{\boldsymbol{\theta}}_\omega$  é pequena quando  $\boldsymbol{\omega}$  varia em  $\Omega$ , então há indícios de estabilidade no modelo ajustado no que se refere ao esquema de perturbação empregado. Cook (1986) sugere que comparemos esses estimadores por meio de uma função apropriada, denominada afastamento de verossimilhança e definida por

$$f(\boldsymbol{\omega}) = 2[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}} | \boldsymbol{\omega})], \boldsymbol{\omega} \in \Omega. \quad (3.1)$$

Dado que  $\hat{\boldsymbol{\theta}}$  é o estimador de máxima verossimilhança no modelo postulado, temos  $\ell(\hat{\boldsymbol{\theta}}) \geq \ell(\hat{\boldsymbol{\theta}} | \boldsymbol{\omega})$ , para todo  $\boldsymbol{\omega} \in \Omega$ . Conseqüentemente,  $f(\boldsymbol{\omega}) \geq 0$ , para todo  $\boldsymbol{\omega} \in \Omega$ . Além disso, notamos que  $\boldsymbol{\omega}_o$  é um ponto de mínimo local da função afastamento de verossimilhança.

O estudo de influência local fundamenta-se em analisar o comportamento da

função  $f$  em uma vizinhança de  $\boldsymbol{\omega}_o$ . Com esse objetivo, consideramos a superfície geométrica  $S \subset \mathbb{R}^{n+1}$ , formada pelo pontos

$$\boldsymbol{\alpha}(\boldsymbol{\omega}) = (\boldsymbol{\omega}^\top, f(\boldsymbol{\omega}))^\top,$$

quando  $\boldsymbol{\omega}$  varia em  $\Omega$ . A superfície  $S$  definida acima é denominada gráfico de influência. Analisar como essa superfície desvia-se de seu plano tangente em  $\boldsymbol{\omega}_o$  é uma das maneiras de estudarmos o compartimento local da função  $f$  em  $\boldsymbol{\omega}_o$ . Essa análise pode ser feita por meio das curvaturas das secções normais da superfície  $S$  em  $\boldsymbol{\omega}_o$ , ou seja, mediante a curvatura normal de  $S$  em  $\boldsymbol{\omega}_o$ . A curvatura normal é dada em termos da primeira e segunda formas fundamentais do gráfico da função afastamento de verossimilhança cujas matrizes associadas denotadas por  $\mathbf{I}$  e  $\mathbf{\Pi}$ , respectivamente, são simétricas e dadas por

$$\mathbf{I}_{ij} = \delta_{ij} + \frac{\partial f}{\partial \omega_i} \frac{\partial f}{\partial \omega_j}$$

e

$$\mathbf{\Pi}_{ij} = \frac{1}{[1 + \|\nabla_f\|^2]^{\frac{1}{2}}} \frac{\partial^2 f}{\partial \omega_i \partial \omega_j},$$

em que

$$\delta_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

e  $\|\nabla_f\|$  denota a norma euclidiana do vetor gradiente de  $f$ .

Em notação matricial,

$$\begin{aligned} \mathbf{I} &= I_n + \nabla_f \nabla_f^\top, \\ \mathbf{\Pi} &= \frac{1}{(1 + \|\nabla_f\|^2)^{\frac{1}{2}}} \mathbf{H}_f, \end{aligned} \tag{3.2}$$

em que  $I_n$  representa a matriz identidade de ordem  $n$  e  $\mathbf{H}_f$  denota a matriz hessiana de  $f$ .

Quaisquer que sejam os vetores  $\boldsymbol{\nu}$  e  $\boldsymbol{v}$  em  $\mathbb{R}^n$ , definimos

$$\mathbf{I}(\boldsymbol{\nu}, \boldsymbol{v}) = \boldsymbol{\nu}^\top \mathbf{I} \boldsymbol{v}$$

e

$$\mathbf{\Pi}(\boldsymbol{\nu}, \boldsymbol{v}) = \boldsymbol{\nu}^\top \mathbf{\Pi} \boldsymbol{v}.$$

Dada uma direção arbitrária em  $\mathbb{R}^n$  representada por um vetor  $\boldsymbol{l}$  não-nulo, consideremos a linha reta em  $\Omega$  que passa por  $\boldsymbol{\omega}_o$  e tem direção determinada pelo vetor  $\boldsymbol{l}$ ; essa

reta é dada por  $\boldsymbol{\omega}(a) = \boldsymbol{\omega}_o + a\mathbf{l}$ , em que  $a \in \mathbb{R}$ ,  $\mathbf{l} \in \mathbb{R}^n$  e  $\boldsymbol{\omega}_o \in \Omega$ . A curvatura normal de  $S$  ao longo de  $\mathbf{l}$ , em um ponto  $\boldsymbol{\omega}_o$ , é definida por

$$C_{\mathbf{l}} = C(\mathbf{l}, \mathbf{l}) = \frac{\boldsymbol{\Pi}(\mathbf{l}, \mathbf{l})}{\mathbf{I}(\mathbf{l}, \mathbf{l})} = \frac{\mathbf{l}^\top \mathbf{H}_f \mathbf{l}}{\mathbf{l}^\top (I_n + \nabla_f \nabla_f^\top) \mathbf{l} (1 + \|\nabla_f\|^2)^{\frac{1}{2}}} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_o}. \quad (3.3)$$

Interessa-nos a curvatura normal em pontos críticos do gráfico de  $f$ . Se  $\boldsymbol{\omega}_o$  é um ponto crítico do gráfico de  $f$  e escolhermos  $\mathbf{l}$  tal que  $\mathbf{l}^\top \mathbf{l} = 1$ , então a expressão (3.3) se reduz a

$$C_{\mathbf{l}} = \mathbf{l}^\top \mathbf{H}_f \mathbf{l} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_o}. \quad (3.4)$$

A partir das equações (3.4) e (3.1), obtemos

$$C_{\mathbf{l}} = -2(\mathbf{l}^\top \ddot{\mathbf{F}} \mathbf{l}) \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_o}, \quad (3.5)$$

em que  $\ddot{\mathbf{F}}$  é uma matriz de ordem  $n$  cujos elementos são dados por  $\partial^2 \ell(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}) / \partial \omega_i \partial \omega_j$ , para  $i, j = 1, 2, \dots, n$ .

Cook (1986) mostrou que a equação (3.5) pode ser reescrita na forma

$$C_{\mathbf{l}} = -2[\mathbf{l}^\top \boldsymbol{\Delta}^\top \ddot{\mathbf{L}}^{-1} \boldsymbol{\Delta} \mathbf{l}] \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_o}, \quad (3.6)$$

em que  $\boldsymbol{\Delta}$  é uma matriz cujos elementos são dados por

$$\Delta_{ij} = \frac{\partial^2 \ell(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \theta_i \partial \omega_j}, \quad (3.7)$$

para  $i = 1, 2, \dots, q$  e  $j = 1, 2, \dots, n$ , e  $\ddot{\mathbf{L}}$  é uma matriz de ordem  $q$  tal que seus elementos são da forma

$$\ddot{\mathbf{L}}_{ij} = \frac{\partial^2 \ell(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \theta_i \partial \theta_j} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_o}, \quad (3.8)$$

para  $i, j = 1, 2, \dots, q$ .

A matriz  $-\ddot{\mathbf{F}} = -\boldsymbol{\Delta}^\top \ddot{\mathbf{L}}^{-1} \boldsymbol{\Delta}$  é semidefinida positiva em  $\boldsymbol{\omega}_o$ , pois a função afastamento de verossimilhança atinge seu mínimo nesse ponto. Sejam  $\lambda_{max}$  o maior autovalor de  $-\ddot{\mathbf{F}}$  e  $\mathbf{e}_{max}$  o autovetor correspondente. O  $\lambda_{max}$  está associado ao máximo da curvatura normal,  $C_{max}$ , e  $\mathbf{e}_{max}$  indica a direção na qual a curvatura normal é máxima. Cook (1986) menciona que valores grandes de  $C_{max} = \lambda_{max}$  é uma indicação de problemas locais.

### 3.2 A curvatura normal conforme

Poon & Poon (1999) definem a curvatura normal conforme ao longo de uma dada direção  $\mathbf{l}$ , em um ponto  $\boldsymbol{\omega}_0$  pertencente ao gráfico de  $\boldsymbol{\alpha}$ , por

$$B_{\mathbf{l}} = \frac{\boldsymbol{\Pi}(\mathbf{l}, \mathbf{l})}{\mathbf{I}(\mathbf{l}, \mathbf{l})[\text{tr}(\boldsymbol{\Pi}^2)]^{\frac{1}{2}}} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \quad (3.9)$$

Notemos que, exceto quando a matriz associada a segunda forma fundamental é a matriz nula, a curvatura normal conforme está bem definida. Consideremos

$$\|\mathbf{H}_f\| = \sqrt{\text{tr}(\mathbf{H}_f^2)}.$$

Conseqüentemente, a partir da equação (3.2), obtemos

$$\|\mathbf{H}_f\| = [1 + \|\nabla_f\|^2]^{\frac{1}{2}} \|\boldsymbol{\Pi}\| = [1 + \|\nabla_f\|^2]^{\frac{1}{2}} [\text{tr}(\boldsymbol{\Pi}^2)]^{\frac{1}{2}}.$$

Daí, pela definição de curvatura normal conforme dada na equação (3.9), teremos

$$\begin{aligned} B_{\mathbf{l}} &= \frac{\mathbf{l}^\top \boldsymbol{\Pi} \mathbf{l}}{\mathbf{l}^\top \mathbf{I} \mathbf{l} [\text{tr}(\boldsymbol{\Pi}^2)]^{\frac{1}{2}}} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ &= \frac{\mathbf{l}^\top \mathbf{H}_f \mathbf{l}}{\mathbf{l}^\top [I_n + \nabla_f \nabla_f^\top] \mathbf{l} [1 + \|\nabla_f\|^2]^{\frac{1}{2}} [\text{tr}(\boldsymbol{\Pi}^2)]^{\frac{1}{2}}} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\ &= \frac{\mathbf{l}^\top \mathbf{H}_f \mathbf{l}}{\mathbf{l}^\top [I_n + \nabla_f \nabla_f^\top] \mathbf{l} \|\mathbf{H}_f\|} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \end{aligned}$$

Agora, mostraremos que se  $\boldsymbol{\omega}_0$  é um ponto crítico de  $f$  e o vetor  $\mathbf{l}$  é unitário, então

$$B_{\mathbf{l}} = \frac{\mathbf{l}^\top \mathbf{H}_f \mathbf{l}}{\|\mathbf{H}_f\|} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}.$$

Basta mostrarmos que sob tais hipóteses,  $\mathbf{l}^\top [I_n + \nabla_f \nabla_f^\top] \mathbf{l} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = 1$ . Com efeito, se  $\boldsymbol{\omega}_0$  é um ponto crítico de  $f$ , então  $\nabla_f(\boldsymbol{\omega}_0) = [0, 0, \dots, 0]^\top$ . Logo,

$$\mathbf{l}^\top [I_n + \nabla_f \nabla_f^\top] \mathbf{l} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} = \mathbf{l}^\top I_n \mathbf{l} = \|\mathbf{l}\|^2 = 1,$$

implicando na relação

$$B_{\mathbf{l}} = \frac{C_{\mathbf{l}}}{\|\mathbf{H}_f\|} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0}. \quad (3.10)$$

Dado que  $\mathbf{H}_f = -2\ddot{\mathbf{F}} = -2[\Delta^\top \ddot{\mathbf{L}}^{-1} \Delta]$ , temos

$$\begin{aligned}
B_{\mathbf{l}} &= \frac{\mathbf{l}^\top \mathbf{H}_f \mathbf{l}}{\|\mathbf{H}_f\|} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\
&= \frac{\mathbf{l}^\top [-2\ddot{\mathbf{F}}] \mathbf{l}}{\| -2\ddot{\mathbf{F}} \|} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\
&= -\frac{\mathbf{l}^\top \ddot{\mathbf{F}} \mathbf{l}}{[\text{tr}(\ddot{\mathbf{F}}^2)]^{\frac{1}{2}}} \Big|_{\boldsymbol{\omega}=\boldsymbol{\omega}_0} \\
&= -\frac{\mathbf{l}^\top \Delta^\top \ddot{\mathbf{L}}^{-1} \Delta \mathbf{l}}{[\text{tr}([\Delta^\top \ddot{\mathbf{L}}^{-1} \Delta]^2)]^{\frac{1}{2}}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}.
\end{aligned}$$

Ou seja,

$$B_{\mathbf{l}} = -\frac{\mathbf{l}^\top \Delta^\top \ddot{\mathbf{L}}^{-1} \Delta \mathbf{l}}{[\text{tr}([\Delta^\top \ddot{\mathbf{L}}^{-1} \Delta]^2)]^{\frac{1}{2}}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}, \boldsymbol{\omega}=\boldsymbol{\omega}_0}. \quad (3.11)$$

Analisando a equação (3.11) observamos que, em relação ao esforço computacional, o cálculo de  $B_{\mathbf{l}}$  equipara-se ao de  $C_{\mathbf{l}}$ . Entretanto, a curvatura normal conforme  $B_{\mathbf{l}}$  é detentora de propriedades relevantes quando comparada com a curvatura normal  $C_{\mathbf{l}}$ . Essas propriedades serão apresentadas a seguir.

### 3.3 Propriedades da curvatura normal conforme

A curvatura normal conforme é detentora de propriedades que nos dão suporte para construir valores de referência que permitem julgarmos, de forma objetiva, a magnitude dessa curvatura. Nesta seção mostraremos estas propriedades.

**Definição 3.1** *Seja  $\mathbf{M}$  uma matriz quadrada de ordem  $n$ . Dizemos que  $\mathbf{M}$  é uma matriz conforme se existe um número real positivo  $\tau$  tal que  $\mathbf{M}\mathbf{M}^\top = \tau I_n$ .*

As matrizes ortogonais são exemplos de matrizes conformes. Em particular, as matrizes identidades.

A seguir apresentamos algumas propriedades das matrizes conformes.

**Propriedade 3.1** *Se  $\mathbf{M}$  é uma matriz conforme, então  $\mathbf{M}$  é não-singular.*

**Prova:** seja  $\mathbf{M}$  uma matriz conforme, isso equivale a dizermos que existe  $\tau > 0$  tal

que  $\mathbf{M}\mathbf{M}^\top = \tau I_n$ . Conseqüentemente,  $[\det(\mathbf{M})]^2 = \tau \det(I_n) = \tau > 0$ , que implica em  $\det(\mathbf{M}) \neq 0$ . Portanto,  $\mathbf{M}$  é não-singular. ■

**Propriedade 3.2** *Uma condição necessária e suficiente para que uma matriz  $\mathbf{M}_{n \times n}$  seja uma matriz conforme é  $\mathbf{M}^\top = \tau \mathbf{M}^{-1}$  para algum número real positivo  $\tau$ .*

**Prova:** primeiramente, supomos que  $\mathbf{M}$  é uma matriz conforme, ou seja, existe  $\tau > 0$  tal que  $\mathbf{M}\mathbf{M}^\top = \tau I_n$ . Pela Propriedade 3.1 temos que  $\mathbf{M}$  é não-singular, isto é, existe  $\mathbf{M}^{-1}$  tal que  $\mathbf{M}\mathbf{M}^{-1} = \mathbf{M}^{-1}\mathbf{M} = I_n$ . Logo,

$$\mathbf{M}^\top = I_n \mathbf{M}^\top = \mathbf{M}^{-1} \mathbf{M} \mathbf{M}^\top = \mathbf{M}^{-1} \tau I_n = \tau \mathbf{M}^{-1}.$$

Por outro lado, admitimos que  $\mathbf{M}^\top = \tau \mathbf{M}^{-1}$  com  $\tau > 0$ . Conseqüentemente,

$$\mathbf{M}\mathbf{M}^\top = \mathbf{M}\tau \mathbf{M}^{-1} = \tau I_n,$$

levando-nos a concluir que  $\mathbf{M}$  é uma matriz conforme. ■

**Propriedade 3.3** *A inversa de uma matriz conforme é uma matriz conforme.*

**Prova:** seja  $\mathbf{M}$  uma matriz conforme. Pelas Propriedades 3.1 e 3.2, temos que  $\mathbf{M}$  é inversível e  $\mathbf{M}^\top = \tau \mathbf{M}^{-1}$  com  $\tau > 0$ . Com isso,

$$\mathbf{M}^{-1}(\mathbf{M}^{-1})^\top = \frac{1}{\tau} \mathbf{M}^\top (\mathbf{M}^{-1})^\top = \frac{1}{\tau} \mathbf{M}^\top (\mathbf{M}^\top)^{-1} = \frac{1}{\tau} I_n,$$

sendo  $\frac{1}{\tau} > 0$ ; assegurando-nos que  $\mathbf{M}^{-1}$  é conforme. ■

**Propriedade 3.4** *Se  $\mathbf{A}_{n \times n}$  e  $\mathbf{B}_{n \times n}$  são matrizes conformes, então  $\mathbf{AB}$  é uma matriz conforme.*

**Prova:** Por hipótese, temos que  $\mathbf{M}_1$  e  $\mathbf{M}_2$  são matrizes conformes, isso significa que existem  $\tau, \delta > 0$  tais que  $\mathbf{M}_1 \mathbf{M}_1^\top = \tau I_n$  e  $\mathbf{M}_2 \mathbf{M}_2^\top = \delta I_n$ . Logo,

$$[\mathbf{M}_1 \mathbf{M}_2][\mathbf{M}_1 \mathbf{M}_2]^\top = \mathbf{M}_1 \mathbf{M}_2 \mathbf{M}_2^\top \mathbf{M}_1^\top = \mathbf{M}_1 \delta I_n \mathbf{M}_1^\top = \delta \mathbf{M}_1 \mathbf{M}_1^\top = \tau \delta I_n.$$

Portanto,  $\mathbf{M}_1 \mathbf{M}_2$  é conforme. ■

**Definição 3.2** *Uma aplicação suave  $\varphi: \Omega \rightarrow \Theta$  é dita reparametrização conforme em  $\omega_o \in \Omega$  se a matriz jacobiana de  $\varphi$  em  $\omega_o$  é uma matriz conforme.*

Um exemplo de reparametrização conforme é  $\varphi(\boldsymbol{\omega}) = \mathbf{M}\boldsymbol{\omega} + c$ , sendo  $\mathbf{M}$  uma matriz conforme.

Reportando-se a Cook (1986), Poon & Poon (1999) consideram reparametrizações como sendo modificações no esquema de perturbação utilizado.

**Teorema 3.3** *Se  $\varphi: \Omega \rightarrow \Theta$  é uma reparametrização conforme em um ponto crítico  $\boldsymbol{\omega}_o \in \Omega$  do gráfico de  $f$ , então a curvatura normal conforme em  $\boldsymbol{\omega}_o$ , em qualquer direção  $\mathbf{l}$ , é invariante sob tal reparametrização.*

**Prova:** sejam  $\varphi(\boldsymbol{\omega})$  uma reparametrização de  $\Omega$  em  $\Theta$  e  $\psi(\mathbf{x})$  sua função inversa. Primeiramente, mostremos que se  $\boldsymbol{\omega}_0$  é um ponto crítico de  $f$ , então  $\varphi(\boldsymbol{\omega}_0)$  é um ponto crítico de  $g(\mathbf{x}) = f \circ \psi(\mathbf{x})$ . De fato, a diferencial de  $g$  em  $\varphi(\boldsymbol{\omega}_0)$ , denotada por  $\mathbf{d}g|_{\varphi(\boldsymbol{\omega}_0)}$ , é tal que

$$\mathbf{d}g|_{\varphi(\boldsymbol{\omega}_0)} = \mathbf{d}(f \circ \psi)|_{\varphi(\boldsymbol{\omega}_0)} = \mathbf{d}f|_{\psi(\varphi(\boldsymbol{\omega}_0))} \mathbf{d}\psi|_{\varphi(\boldsymbol{\omega}_0)} = \mathbf{d}f|_{\boldsymbol{\omega}_0} \mathbf{d}\psi|_{\varphi(\boldsymbol{\omega}_0)}.$$

Ora, por hipótese  $\boldsymbol{\omega}_0$  é um ponto crítico de  $f$ , ou seja,  $\mathbf{d}f|_{\boldsymbol{\omega}_0} = 0$ . Conseqüentemente,  $\mathbf{d}g|_{\varphi(\boldsymbol{\omega}_0)} = 0$ , assegurando-nos que  $\varphi(\boldsymbol{\omega}_0)$  é um ponto crítico de  $g = f \circ \psi$ .

Agora, obtenhamos a matriz hessiana de  $g = f \circ \psi$  em termos da hessiana de  $f$

$$\mathbf{H}_g = \frac{\partial^2 f \circ \psi(\mathbf{x})}{\partial \omega_i \partial \omega_j} = \frac{\partial^2 f(\psi(\mathbf{x}))}{\partial \omega_i \partial \omega_j} \frac{\partial \psi(\mathbf{x})}{\partial \omega_i} \frac{\partial \psi(\mathbf{x})}{\partial \omega_j} + \frac{\partial f(\psi(\mathbf{x}))}{\partial \omega_j} \frac{\partial^2 \psi(\mathbf{x})}{\partial \omega_i \partial \omega_j}.$$

Logo, em  $\mathbf{x} = \varphi(\boldsymbol{\omega}_0)$  temos

$$\mathbf{H}_g = \frac{\partial^2 f(\boldsymbol{\omega}_0)}{\partial \omega_i \partial \omega_j} \frac{\partial \psi(\varphi(\boldsymbol{\omega}_0))}{\partial \omega_i} \frac{\partial \psi(\varphi(\boldsymbol{\omega}_0))}{\partial \omega_j} + \frac{\partial f(\boldsymbol{\omega}_0)}{\partial \omega_j} \frac{\partial^2 \psi(\varphi(\boldsymbol{\omega}_0))}{\partial \omega_i \partial \omega_j}.$$

Uma vez que  $\boldsymbol{\omega}_0$  é um ponto crítico de  $f$ , isto é,  $\frac{\partial f(\boldsymbol{\omega}_0)}{\partial \omega_j} = 0$  para  $j = 1, 2, \dots, n$ , obtemos

$$\mathbf{H}_g = \frac{\partial \psi(\varphi(\boldsymbol{\omega}_0))}{\partial \omega_j} \frac{\partial^2 f(\boldsymbol{\omega}_0)}{\partial \omega_i \partial \omega_j} \frac{\partial \psi(\varphi(\boldsymbol{\omega}_0))}{\partial \omega_i} = \boldsymbol{\Psi}^\top \mathbf{H}_f \boldsymbol{\Psi}, \quad (3.12)$$

em que  $\boldsymbol{\Psi}$  representa a matriz jacobiana de  $\psi$  em  $\varphi(\boldsymbol{\omega}_0)$ .

Prosseguindo, a hipótese de que  $\varphi$  é conforme em  $\boldsymbol{\omega}_0$  significa dizer que a matriz jacobiana de  $\varphi$ , denotada por  $\boldsymbol{\Upsilon}$ , é conforme em  $\boldsymbol{\omega}_0$ , ou seja,  $\tau \boldsymbol{\Upsilon}^\top = \boldsymbol{\Upsilon}^{-1}$  para algum número real positivo  $\tau$ . Conseqüentemente, tal hipótese combinada com o fato de que  $\boldsymbol{\Psi} = \boldsymbol{\Upsilon}^{-1}$  implicam

$$\boldsymbol{\Psi}^\top \boldsymbol{\Psi} = [\boldsymbol{\Upsilon}^{-1}]^\top \boldsymbol{\Upsilon}^{-1} = [\tau \boldsymbol{\Upsilon}^\top]^\top \boldsymbol{\Upsilon}^{-1} = \tau \boldsymbol{\Upsilon} \boldsymbol{\Upsilon}^{-1} = \tau \mathbf{I}_n. \quad (3.13)$$

Ou seja, a matriz jacobiana de  $\psi$  em  $\varphi(\boldsymbol{\omega}_0)$  é uma matriz conforme. Em outras palavras,  $\psi$  é uma aplicação conforme em  $\varphi(\boldsymbol{\omega}_0)$ . A partir das equações (3.12) e (3.13), obtemos

$$\|\mathbf{H}_g\| = [\text{tr}(\boldsymbol{\Psi}^\top \mathbf{H}_f \boldsymbol{\Psi})^2]^{\frac{1}{2}} = [\text{tr}(\boldsymbol{\Psi}^\top \boldsymbol{\Psi} \mathbf{H}_f \mathbf{H}_f \boldsymbol{\Psi}^\top \boldsymbol{\Psi})]^{\frac{1}{2}} = [\text{tr}(\tau I_n \mathbf{H}_f^2 \tau I_n)]^{\frac{1}{2}} = \tau \|\mathbf{H}_f\|.$$

A reparametrização considerada leva a direção  $\mathbf{l}$  em  $\boldsymbol{\omega}_0$  para  $\boldsymbol{\Upsilon}(\mathbf{l})$  em  $\varphi(\boldsymbol{\omega}_0)$ . Em  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$  temos  $\boldsymbol{\Upsilon} = \frac{1}{\tau} \boldsymbol{\Psi}^\top$ , daí pela definição da curvatura normal dada na equação (3.3), temos

$$\begin{aligned} \hat{C}_{\boldsymbol{\Upsilon}(\mathbf{l})} &= \frac{[\boldsymbol{\Upsilon} \mathbf{l}]^\top \mathbf{H}_g \boldsymbol{\Upsilon} \mathbf{l}}{[\boldsymbol{\Upsilon} \mathbf{l}]^\top \boldsymbol{\Upsilon} \mathbf{l}} = \frac{[\frac{1}{\tau} \boldsymbol{\Psi}^\top \mathbf{l}]^\top \boldsymbol{\Psi}^\top \mathbf{H}_f \boldsymbol{\Psi} \frac{1}{\tau} \boldsymbol{\Psi}^\top \mathbf{l}}{[\frac{1}{\tau} \boldsymbol{\Psi}^\top \mathbf{l}]^\top \frac{1}{\tau} \boldsymbol{\Psi}^\top \mathbf{l}} = \frac{\mathbf{l}^\top \boldsymbol{\Psi} \frac{1}{\tau} \boldsymbol{\Psi}^\top \mathbf{H}_f \boldsymbol{\Psi} \frac{1}{\tau} \mathbf{l}}{\mathbf{l}^\top \boldsymbol{\Psi} \frac{1}{\tau} \frac{1}{\tau} \boldsymbol{\Psi}^\top \mathbf{l}} \\ &= \frac{\mathbf{l}^\top \frac{1}{\tau} \tau I_n \mathbf{H}_f \tau I_n \frac{1}{\tau} \mathbf{l}}{\mathbf{l}^\top \frac{1}{\tau} \tau I_n \frac{1}{\tau} \mathbf{l}} = \tau \left[ \frac{\mathbf{l}^\top \mathbf{H}_f \mathbf{l}}{\mathbf{l}^\top \mathbf{l}} \right] = \tau C_{\mathbf{l}}. \end{aligned}$$

Pela relação (3.10), concluímos

$$\hat{B}_{\boldsymbol{\Upsilon}(\mathbf{l})} = \frac{\hat{C}_{\boldsymbol{\Upsilon}(\mathbf{l})}}{\|\mathbf{H}_g\|} = \frac{\tau C_{\mathbf{l}}}{\tau \|\mathbf{H}_f\|} = B_{\mathbf{l}},$$

conforme havíamos enunciado no teorema. ■

No modelo de regressão normal linear, quando  $\boldsymbol{\omega}^\top = (\omega_1, \omega_2, \dots, \omega_n)$  é o vetor de perturbação na ponderação de casos para uma amostra aleatória de tamanho  $N = n$ , a curvatura normal conforme é invariante com respeito à reparametrização  $\varphi(\boldsymbol{\omega}) = (1 + \omega_1/2, 1 + \omega_2/2, \dots, 1 + \omega_n/2)$ , estudada por Loynes (1986), pois tal reparametrização é conforme em  $\boldsymbol{\omega}_0 = (1, 1, \dots, 1)^\top$ .

**Teorema 3.4** *A curvatura normal conforme  $B_{\mathbf{l}}$  é tal que  $|B_{\mathbf{l}}| \leq 1$ , para qualquer direção  $\mathbf{l}$ .*

**Prova:** seja  $\mathcal{B} = \{\mathbf{e}_i ; 1 \leq i \leq n\}$  uma base ortonormal formada por autovetores de  $\boldsymbol{\Pi}$ , cujos autovalores correspondentes são  $\lambda_1, \lambda_2, \dots, \lambda_n$ . Conseqüentemente,

- (i)  $\mathbf{e}_i$  é ortogonal a  $\mathbf{e}_j$ , para  $i \neq j$  e  $1 \leq i, j \leq n$ . Ou seja,  $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0$  para  $i \neq j$  com  $i, j = 1, 2, \dots, n$ .
- (ii)  $\|\mathbf{e}_i\| = 1$ , para  $i = 1, 2, \dots, n$ .
- (iii)  $\boldsymbol{\Pi} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$  é a matriz de representação, na base  $\mathcal{B}$ , da segunda forma fundamental.

Para qualquer direção  $\mathbf{l} \in \mathbb{R}^n$ , com  $\mathbf{l} = \sum_{i=1}^n a_i \mathbf{e}_i = (a_1, a_2, \dots, a_n)$  e  $a_i \in \mathbb{R}$ , para todo  $i = 1, 2, \dots, n$ , temos

$$\mathbf{I}(\mathbf{l}; \mathbf{l}) = \langle \mathbf{l}; \mathbf{l} \rangle = \sum_{i=1}^n a_i^2;$$

$$\mathbf{\Pi}(\mathbf{l}; \mathbf{l}) = \mathbf{l}^\top \mathbf{\Pi} \mathbf{l} = \sum_{i=1}^n \lambda_i a_i^2$$

e

$$\text{tr}(\mathbf{\Pi}^2) = \sum_{i=1}^n \lambda_i^2.$$

Logo,

$$|B_{\mathbf{l}}| = \left| \frac{\mathbf{\Pi}(\mathbf{l}; \mathbf{l})}{\mathbf{I}(\mathbf{l}; \mathbf{l}) \{\text{tr}(\mathbf{\Pi}^2)\}^{\frac{1}{2}}} \right| = \frac{|\sum_{i=1}^n \lambda_i a_i^2|}{|\sum_{i=1}^n a_i^2| |\sum_{i=1}^n \lambda_i^2|^{\frac{1}{2}}} = \frac{|\sum_{i=1}^n \lambda_i a_i^2|}{\sum_{i=1}^n a_i^2 [\sum_{i=1}^n \lambda_i^2]^{\frac{1}{2}}}. \quad (3.14)$$

Neste ponto, observemos que

$$\sum_{i=1}^n \lambda_i a_i^2 = \langle (\lambda_1, \lambda_2, \dots, \lambda_n); (a_1^2, a_2^2, \dots, a_n^2) \rangle.$$

Pela desigualdade de Cauchy-Schwarz, temos

$$\left| \sum_{i=1}^n \lambda_i a_i^2 \right| \leq \|(\lambda_1, \lambda_2, \dots, \lambda_n)\| \cdot \|a_1^2, a_2^2, \dots, a_n^2\| = \left[ \sum_{i=1}^n \lambda_i^2 \right]^{\frac{1}{2}} \left[ \sum_{i=1}^n a_i^4 \right]^{\frac{1}{2}}.$$

Voltando a expressão (3.14), temos

$$|B_{\mathbf{l}}| \leq \frac{[\sum_{i=1}^n \lambda_i^2]^{\frac{1}{2}} [\sum_{i=1}^n a_i^4]^{\frac{1}{2}}}{\sum_{i=1}^n a_i^2 [\sum_{i=1}^n \lambda_i^2]^{\frac{1}{2}}} = \frac{[\sum_{i=1}^n a_i^4]^{\frac{1}{2}}}{\sum_{i=1}^n a_i^2}.$$

Uma vez que  $\sum_{i=1}^n a_i^4 \leq [\sum_{i=1}^n a_i^2]^2$ , concluímos

$$|B_{\mathbf{l}}| \leq \frac{[\sum_{i=1}^n a_i^4]^{\frac{1}{2}}}{\sum_{i=1}^n a_i^2} \leq \frac{[\sum_{i=1}^n a_i^2]^2]^{\frac{1}{2}}}{\sum_{i=1}^n a_i^2} = 1. \quad \blacksquare$$

O Teorema 3.4 nos assegura que a curvatura normal conforme assume valores em um intervalo limitado da reta real. Isso facilita a interpretação de sua magnitude.

**Teorema 3.5** *Se  $\mathcal{B} = \{\mathbf{e}_i; 1 \leq i \leq n\}$  é uma coleção de autovetores ortonormais de  $\mathbf{\Pi}$ , então a curvatura normal conforme  $B_{\mathbf{e}_i}$  na direção do vetor  $\mathbf{e}_i$  é igual ao autovalor*

normalizado,  $\hat{\lambda}_i$ , associado ao autovetor  $\mathbf{e}_i$ . Além disso,  $\sum_{i=1}^n B_{\mathbf{e}_i}^2 = 1$ .

**Prova:** se  $\mathcal{B} = \{\mathbf{e}_i; 1 \leq i \leq n\}$  é uma coleção de autovetores ortonormais de  $\mathbf{\Pi}$ , então para todo  $\mathbf{e}_i \in \mathcal{B}$  e  $i = 1, 2, \dots, n$ , temos

$$\mathbf{I}(\mathbf{e}_i; \mathbf{e}_i) = \mathbf{e}_i^\top \mathbf{e}_i = \|\mathbf{e}_i\|^2 = 1$$

e

$$\mathbf{\Pi}(\mathbf{e}_i; \mathbf{e}_i) = \mathbf{e}_i^\top \mathbf{\Pi} \mathbf{e}_i = \mathbf{e}_i^\top \lambda_i \mathbf{e}_i = \lambda_i \mathbf{e}_i^\top \mathbf{e}_i = \lambda_i.$$

Conseqüentemente,

$$B_{\mathbf{e}_i} = \frac{\mathbf{\Pi}(\mathbf{e}_i, \mathbf{e}_i)}{\mathbf{I}(\mathbf{e}_i; \mathbf{e}_i) [\text{tr}(\mathbf{\Pi}^2)]^{\frac{1}{2}}} = \frac{\lambda_i}{[\sum_{i=1}^n \lambda_i^2]^{\frac{1}{2}}} = \hat{\lambda}_i.$$

Além disso,

$$\sum_{i=1}^n B_{\mathbf{e}_i}^2 = \sum_{i=1}^n \frac{\lambda_i^2}{\sum_{k=1}^n \lambda_k^2} = \frac{1}{\sum_{k=1}^n \lambda_k^2} \sum_{i=1}^n \lambda_i^2 = 1.$$

■

Uma vez que  $\sum_{i=1}^n B_{\mathbf{e}_i}^2 = 1$ , se as curvaturas são iguais para todos os autovetores  $\mathbf{e}_i$  e  $i = 1, \dots, n$ , então cada uma delas é igual a  $1/\sqrt{n}$ . De fato, se  $B_{\mathbf{e}_i} = B$  para todo  $i = 1, 2, \dots, n$ , então

$$\sum_{i=1}^n B_{\mathbf{e}_i}^2 = \sum_{i=1}^n B^2 = nB^2.$$

Ora, pelo Teorema 3.5

$$\sum_{i=1}^n B_{\mathbf{e}_i}^2 = 1.$$

Portanto,

$$B = \frac{1}{\sqrt{n}}.$$

O número real  $1/\sqrt{n}$  pode ser usado na construção de um valor de referência para que possamos julgar a magnitude da curvatura e, conseqüentemente, avaliar a influência local de forma objetiva.

### 3.4 Influência local por meio da curvatura normal conforme

Notemos que a curvatura normal conforme e a curvatura normal diferem apenas por um fator positivo. Logo o autovetor  $\mathbf{e}_{max}$  que maximiza uma delas também

maximizará a outra. Entretanto, a propriedade de invariância sob reparametrizações conformes e o fato de ser limitada facilitam a interpretação da curvatura normal conforme. A discussão na Seção 3.3 sugere que um valor de referência para julgar o efeito de  $B_{\mathbf{e}_i}$  e  $B_{\mathbf{l}}$  em vários níveis pode ser determinado. Com essa finalidade, Poon & Poon (1999) definem:

**Definição 3.6** *Um autovetor  $\mathbf{e}$  é dito  $q$ -influyente se  $|B_{\mathbf{e}}| \geq q/\sqrt{n}$ .*

### 3.4.1 Influência de um único autovetor

Seja  $\{\mathbf{E}_i ; 1 \leq i \leq n\}$  a base canônica do  $\mathbb{R}^n$ . Chamaremos  $\mathbf{E}_i$  o  $i$ -ésimo vetor básico de perturbação. A fim de analisar a contribuição de vetores básicos de perturbação a influência de um único vetor  $\mathbf{e}$   $q$ -influyente, encontramos os vetores básicos de perturbação que estão próximos de  $\mathbf{e}$ , no sentido de que a distância entre cada um deles e o vetor  $\mathbf{e}$  é pequena. Seja  $\mathcal{C} = \{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  uma coleção de autovetores ortonormais cujos autovalores normalizados correspondentes são  $\bar{\lambda}_1, \bar{\lambda}_2, \dots, \bar{\lambda}_n$ . Segue que se  $\mathbf{e}_i = \sum_{j=1}^n a_{ij} \mathbf{E}_j$ , para  $i = 1, 2, \dots, n$ , então  $\sum_{j=1}^n a_{ij}^2 = 1$ . De fato, para todo  $i = 1, 2, \dots, n$ , temos

$$\sum_{j=1}^n a_{ij}^2 = \langle \mathbf{e}_i, \mathbf{e}_i \rangle = \|\mathbf{e}_i\|^2 = 1,$$

sendo  $\langle \cdot, \cdot \rangle$  o produto interno usual em  $\mathbb{R}^n$  e  $\|\cdot\|$  a norma euclidiana. De forma que para cada  $i$  fixado,  $i = 1, 2, \dots, n$ , se a contribuição  $a_{ij}$  é a mesma para todos os  $\mathbf{E}_j$ 's, então  $|a_{ij}| = 1/\sqrt{n}$ . Com efeito, dado que a contribuição  $a_{ij}$  é a mesma para todo  $j = 1, 2, \dots, n$ , temos para cada  $i$  fixado

$$\sum_{j=1}^n a_{ij}^2 = n a_{ij}^2,$$

para qualquer que seja  $j = 1, 2, \dots, n$ . Como  $\sum_{j=1}^n a_{ij}^2 = 1$ , então  $n a_{ij}^2 = 1$ , ou seja,  $|a_{ij}| = 1/\sqrt{n}$ . Como foi sugerido na seção anterior, esse número pode ser usado na construção de um valor de referência. Essa metodologia pode ser aplicada para estudar o  $\mathbf{e}_{max}$  ou a influência individual de qualquer autovetor.

### 3.4.2 Contribuição agregada de vetores básicos de perturbação

A idéia apresentada na Seção 3.4.1, pode ser estendida para analisar a influência de vetores de perturbação básicos para todos os autovetores  $q$ -influentes. Com esse objetivo, definamos  $\mu_{(i)} = |\bar{\lambda}_i|$  para  $i = 1, 2, \dots, n$ . Em seguida, ordenemos esses valores da seguinte forma:

$$\mu_{max} = \mu_{(1)} \geq \mu_{(2)} \geq \dots \geq \mu_{(k)} \geq \frac{q}{\sqrt{n}} > \mu_{(k+1)} \geq \dots \geq \mu_{(n)} \geq 0.$$

Por fim, usemos  $a_{ij}$  para denotar a  $j$ -ésima componente do autovetor normalizado associado a  $\mu_{(i)} = |\bar{\lambda}_i|$ . Esses elementos levaram Poon & Poon (1999) a formular a seguinte definição.

**Definição 3.7** *A contribuição agregada do  $j$ -ésimo vetor básico de perturbação  $\mathbf{E}_j$  para todos os autovetores  $q$ -influentes, denotada por  $m(q)_j$ , é o número real dado por*

$$m(q)_j = \left[ \sum_{i=1}^k \mu_{(i)} a_{ij}^2 \right]^{\frac{1}{2}}.$$

Neste ponto, observemos que se a contribuição de todos os vetores de perturbação é a mesma. Então cada uma delas, denotadas uniformemente por  $\bar{m}(q)$ , é dada por

$$\bar{m}(q) = \left[ \frac{1}{n} \sum_{i=1}^k \mu_{(i)} \right]^{\frac{1}{2}}.$$

De fato, primeiramente vejamos que

$$\sum_{j=1}^n m(q)_j^2 = \sum_{j=1}^n \left\{ \sum_{i=1}^k \mu_{(i)} a_{ij}^2 \right\} = \sum_{i=1}^k \mu_{(i)} \left\{ \sum_{j=1}^n a_{ij}^2 \right\} = \left[ \sum_{i=1}^k \mu_{(i)} \right] 1 = \sum_{i=1}^k \mu_{(i)}. \quad (3.15)$$

Por outro lado, se a contribuição  $m(q)_j = \bar{m}(q)$ , para todo  $j = 1, 2, \dots, n$ , temos

$$\sum_{j=1}^n m(q)_j^2 = n[\bar{m}(q)]^2. \quad (3.16)$$

A partir das equações (3.15) e (3.16), concluímos

$$\bar{m}(q) = \left[ \frac{1}{n} \sum_{i=1}^k \mu_{(i)} \right]^{\frac{1}{2}}. \quad (3.17)$$

Devemos tomar  $\bar{m}(q)$  como referência quando tivermos por objetivo determinar a significância da contribuição de cada um dos vetores básicos de perturbação sobre os autovetores  $q$ -influentes.

Existem dois casos extremos que devem ser levados em consideração ao fazermos uso desse método. O primeiro é permitir  $q$  suficientemente grande de tal forma que consideremos a contribuição de cada um dos vetores básicos de perturbação para o  $\mathbf{e}_{max}$  apenas. Sendo assim,

$$m(q)_j = [\mu_{(1)} a_{1j}^2]^{\frac{1}{2}} = [\mu_{max} a_{1j}^2]^{\frac{1}{2}} = [\mu_{max}]^{\frac{1}{2}} |a_{1j}|.$$

Neste caso, tal método equivale a comparar  $a_{ij}$  com  $1/\sqrt{n}$  como sugerido na Seção 3.4.1. O outro extremo é permitir  $q = 0$ , dessa forma todos autovetores são incluídos em nossa análise. Levando-nos a contribuição total, denotada por  $m_j$  e definida por

$$m_j = m(0)_j = \left[ \sum_{i=1}^n \mu_{(i)} a_{ij}^2 \right]^{\frac{1}{2}}. \quad (3.18)$$

Assim, se a contribuição de todos os vetores básicos de perturbação é a mesma, então cada uma delas, denotadas uniformemente por  $\bar{m}$ , é tal que

$$\bar{m} = \left[ \frac{1}{n} \sum_{i=1}^n \mu_{(i)} \right]^{\frac{1}{2}} = \left[ \frac{1}{n} \sum_{i=1}^n |\bar{\lambda}_i| \right]^{\frac{1}{2}} = \left[ \frac{1}{n} \sum_{i=1}^n \left| \frac{\lambda_i}{\sqrt{\sum_{k=1}^n \lambda_k^2}} \right| \right]^{\frac{1}{2}} = \left[ \frac{\sum_{i=1}^n |\lambda_i|}{n \sqrt{\sum_{i=1}^n \lambda_i^2}} \right]^{\frac{1}{2}}. \quad (3.19)$$

A contribuição total  $m_j$  e a curvatura normal conforme  $B_{E_j}$  estão fortemente relacionadas. A seguir explicitaremos essa relação. Sejam  $\mathbf{E}_j = \sum_{i=1}^n a_{ij} \mathbf{e}_i$  e  $\mathbf{B}$  a matriz associada a forma bilinear  $B_{\mathbf{l}}$ . Dado que  $B_{\mathbf{v}} = \mathbf{v}^{\top} \mathbf{B} \mathbf{v}$  e  $\mathbf{B} \mathbf{e}_i = \hat{\lambda}_i \mathbf{e}_i$ , temos

$$\begin{aligned} B_{\mathbf{E}_j} &= \mathbf{E}_j^{\top} \mathbf{B} \mathbf{E}_j = \left[ \sum_{i=1}^n a_{ij} \mathbf{e}_i \right]^{\top} \mathbf{B} \left( \sum_{k=1}^n a_{kj} \mathbf{e}_k \right) \\ &= \sum_{i=1}^n a_{ij} \mathbf{e}_i^{\top} \sum_{k=1}^n \mathbf{B}(a_{kj} \mathbf{e}_k) \\ &= \sum_{i=1}^n a_{ij} \mathbf{e}_i^{\top} \sum_{k=1}^n a_{kj} \mathbf{B} \mathbf{e}_k \\ &= \sum_{i=1}^n a_{ij} \mathbf{e}_i^{\top} \sum_{k=1}^n a_{kj} \hat{\lambda}_k \mathbf{e}_k \\ &= \sum_{i=1}^n a_{ij} \sum_{k=1}^n a_{kj} \hat{\lambda}_k \mathbf{e}_i^{\top} \mathbf{e}_k. \end{aligned}$$

Neste ponto, usemos o fato de que  $\mathcal{C} = \{\mathbf{e}_i; 1 \leq i \leq n\}$  é uma coleção de vetores

ortonormais. Logo,

$$B_{\mathbf{E}_j} = \sum_{i=1}^n \bar{\lambda}_i a_{ij}^2.$$

Além disso, se todos os autovalores são não-negativos, então  $B_{\mathbf{E}_j}$  é igual ao quadrado da contribuição total do  $j$ -ésimo vetor básico de perturbação. De fato, se  $\bar{\lambda}_i \geq 0$  para todo  $i = 1, 2, \dots, n$ , então

$$\sum_{i=1}^n \mu_{(i)} a_{ij}^2 = \sum_{i=1}^n \bar{\lambda}_i a_{ij}^2 \geq 0.$$

Conseqüentemente,

$$m_j^2 = \left[ \left[ \sum_{i=1}^n \mu_{(i)} a_{ij}^2 \right]^{\frac{1}{2}} \right]^2 = \left[ \left[ \sum_{i=1}^n \bar{\lambda}_i a_{ij}^2 \right]^{\frac{1}{2}} \right]^2 = B_{\mathbf{E}_j}.$$

As idéias expostas acima podem ser resumidas no seguinte teorema.

**Teorema 3.8** *Se todos os autovalores da matriz hessiana  $\mathbf{H}_f$  são não-negativos, então  $m_j^2 = B_{\mathbf{E}_j}$ , para todo  $j = 1, 2, \dots, n$ .*

**Corolário 3.9** *Se a contribuição  $m_j$  é a mesma para todo vetor básico de perturbação  $\mathbf{E}_j$ , então*

$$B_{\mathbf{E}_j} = \bar{m}^2 = \frac{\text{tr}(\mathbf{\Pi})}{n [\text{tr}(\mathbf{\Pi}^2)]^{\frac{1}{2}}},$$

para todo  $j = 1, 2, \dots, n$ .

**Prova:** pelo Teorema 3.8,

$$B_{\mathbf{E}_j} = m_j^2,$$

para todo  $j = 1, 2, \dots, n$ . Sob a hipótese de que a contribuição de todos os vetores básicos de perturbação é mesma, temos

$$B_{\mathbf{E}_j} = m_j^2 = \bar{m}^2 = \left[ \left[ \frac{\sum_{i=1}^n \lambda_i}{n \sqrt{\sum_{i=1}^n \lambda_i^2}} \right]^{\frac{1}{2}} \right]^2 = \frac{\sum_{i=1}^n \lambda_i}{n [\sum_{i=1}^n \lambda_i^2]^{\frac{1}{2}}} = \frac{\text{tr}(\mathbf{\Pi})}{n [\text{tr}(\mathbf{\Pi}^2)]^{\frac{1}{2}}}$$

■

## Capítulo 4

# Influência local no modelo de regressão log-Birnbaum-Saunders

Neste capítulo, aplicaremos a técnica de influência local mediante a curvatura normal conforme proposta por Poon & Poon (1999) no modelo de regressão log-BS. Apresentaremos as matrizes apropriadas para obter a influência local nos parâmetros estimados do modelo sob os esquemas de perturbação: ponderação de casos, perturbação da variável resposta e perturbação da variável explicativa, obtidos por Leiva, Barros, Paula & Galea (2007). A título de ilustração, analisaremos o conjunto de dados considerados em Krall et al. (1975) e apresentados por Lawless (1982, pp.336-337). Esses dados foram analisados por Jin et al. (2003) e Ghosh & Ghosal (2006), que usaram um modelo semiparamétrico de tempo de falha acelerado. Mais recentemente, Leiva, Barros, Paula & Galea (2007) realizaram uma análise de diagnóstico desses dados baseada em resíduos e na metodologia de influência local por meio da curvatura normal proposta por Cook (1986).

## 4.1 Esquemas de perturbações

Sejam  $\boldsymbol{\omega}^\top = (\omega_1, \omega_2, \dots, \omega_n) \in \Omega \subset \mathbb{R}^n$  um vetor de perturbação e  $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$  a função de log-verossimilhança perturbada correspondente ao modelo adotado e a forma de perturbação imposta. Conforme Escobar & Meeker (1992), admitiremos que

- (i) Existe  $\boldsymbol{\omega}_o \in \Omega$ , denominado vetor de não-perturbação, tal que  $\ell(\boldsymbol{\theta}|\boldsymbol{\omega}_o) = \ell(\boldsymbol{\theta})$  para todo  $\boldsymbol{\theta} \in \Theta$ ;
- (ii)  $\ell(\boldsymbol{\theta}|\boldsymbol{\omega})$  é duas vezes diferenciável em  $(\boldsymbol{\theta}^\top, \boldsymbol{\omega}^\top)^\top$ ;
- (iii) Para cada  $\boldsymbol{\omega} \in \Omega$  fixado,  $\ell(\boldsymbol{\theta})$  possui um único máximo em  $\hat{\boldsymbol{\theta}}_\boldsymbol{\omega}$ ;
- (iv)  $\hat{\boldsymbol{\theta}}_\boldsymbol{\omega}$  pertence ao interior de  $\Theta$ .

### 4.1.1 Ponderação de casos

Neste esquema de perturbação, a função de log-verossimilhança perturbada tem a forma

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) \propto \sum_{i \in D} \omega_i \log(f(y_i, \boldsymbol{\omega})) + \sum_{i \in C} \omega_i \log(S(y_i, \boldsymbol{\omega})),$$

em que  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^\top \in \Omega \subset \mathbb{R}^n$  com  $0 \leq \omega_i \leq 1$ , para  $i = 1, 2, \dots, n$ ,  $\boldsymbol{\omega}_0 = (1, 1, \dots, 1)^\top$  é o vetor de não-perturbação e os conjuntos  $D$  e  $C$  foram definidos na equação (1.8).

Em particular, para o modelo de regressão log-Birnbaum-Saunders definido na equação(2.4), temos

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) \propto \sum_{i \in D} \omega_i \left[ \log(\xi_{i1}) - \frac{1}{2} [\log(8\pi) + \xi_{i2}^2] \right] + \sum_{i \in C} \omega_i \log(\Phi(-\xi_{i2})),$$

em que  $\xi_{i1} = \frac{2}{\alpha} \cosh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)$  e  $\xi_{i2} = \frac{2}{\alpha} \sinh\left(\frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{2}\right)$ .

No que se refere à matriz

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{ij})_{(p+1) \times n} = \left( \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \theta_i \partial \omega_j} \right)$$

em que  $i = 1, 2, \dots, p+1$  e  $j = 1, 2, \dots, n$ , Leiva, Barros, Paula & Galea (2007), considerando o modelo de regressão sob estudo e o esquema de ponderação de casos, mostraram que tal matriz  $\boldsymbol{\Delta}$  é dada por

$$\Delta = \begin{pmatrix} \Delta_{\beta} \\ \Delta_{\alpha} \end{pmatrix},$$

em que  $\Delta_{\beta} = \mathbf{X}^{\top} \text{diag}\{\hat{b}_1, \hat{b}_2, \dots, \hat{b}_n\}$  com

$$\hat{b}_i = \begin{cases} \frac{1}{2} \left[ \hat{\xi}_{i1} \hat{\xi}_{i2} - \frac{\hat{\xi}_{i2}}{\hat{\xi}_{i1}} \right], & \text{se } i \in D \\ \frac{\hat{\xi}_{i1}}{2} h(\hat{\xi}_{i2}), & \text{se } i \in C \end{cases} \quad (4.1)$$

e  $\Delta_{\alpha} = (\hat{a}_1, \hat{a}_2, \dots, \hat{a}_n)$  com

$$\hat{a}_i = \begin{cases} -\frac{1}{\hat{\alpha}} + \frac{[\hat{\xi}_{i2}]^2}{\hat{\alpha}}, & \text{se } i \in D \\ \frac{\hat{\xi}_{i2}}{2} h(\hat{\xi}_{i2}), & \text{se } i \in C, \end{cases} \quad (4.2)$$

sendo  $\hat{\xi}_{i1}$  e  $\hat{\xi}_{i2}$  as estimativas de máxima verossimilhança de  $\xi_{i1}$  e  $\xi_{i2}$ , respectivamente, e  $h(\cdot)$  dada por  $h(\hat{\xi}_{i2}) = \phi(\hat{\xi}_{i2})/\Phi(-\hat{\xi}_{i2})$ .

Sob a hipótese de que apenas os casos não-censurados sejam perturbados, as equações (4.1) e (4.2) reduzem-se, respectivamente, a

$$\hat{b}_i = \begin{cases} \frac{1}{2} \left[ \hat{\xi}_{i1} \hat{\xi}_{i2} - \frac{\hat{\xi}_{i2}}{\hat{\xi}_{i1}} \right], & \text{se } i \in D \\ 0, & \text{se } i \in C \end{cases}$$

e

$$\hat{a}_i = \begin{cases} -\frac{1}{\hat{\alpha}} + \frac{[\hat{\xi}_{i2}]^2}{\hat{\alpha}}, & \text{se } i \in D \\ 0, & \text{se } i \in C \end{cases}$$

Por outro lado, se apenas os casos censurados sejam perturbados, temos

$$\hat{b}_i = \begin{cases} 0, & \text{se } i \in D \\ \frac{\hat{\xi}_{i1}}{2} h(\hat{\xi}_{i2}), & \text{se } i \in C \end{cases}$$

e

$$\hat{a}_i = \begin{cases} 0, & \text{se } i \in D \\ \frac{\hat{\xi}_{i2}}{2} h(\hat{\xi}_{i2}), & \text{se } i \in C. \end{cases}$$

#### 4.1.2 Perturbação na variável resposta

Neste caso, sujeitamos a resposta  $y_i$  a uma perturbação da forma  $y_{i\omega} = y_i + \omega_i s_y$ , em que  $\omega_i \in \mathbb{R}$  para todo  $i = 1, 2, \dots, n$ ,  $s_y$  é um fator de escala. Considerando o modelo

de regressão sob estudo e o tipo de perturbação descrito anteriormente, a função de log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\theta} | \boldsymbol{\omega}) \propto \sum_{i \in D} \left\{ \log(\xi_{i1\omega_1}) - \frac{1}{2} [\log(8\pi) + \xi_{i2\omega_1}^2] \right\} + \sum_{i \in C} \log(\Phi(-\xi_{i2\omega_1}))$$

em que

$$\xi_{i1\omega_1} = \frac{2}{\alpha} \cosh \left( \frac{y_{i\omega} - \mathbf{x}_i^\top \boldsymbol{\beta}}{2} \right)$$

e

$$\xi_{i2\omega_1} = \frac{2}{\alpha} \sinh \left( \frac{y_{i\omega} - \mathbf{x}_i^\top \boldsymbol{\beta}}{2} \right)$$

Para esse esquema de perturbação, a matriz

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_\beta \\ \boldsymbol{\Delta}_\alpha \end{pmatrix},$$

é tal que

$$\boldsymbol{\Delta}_\beta = \mathbf{X}^\top \text{diag}\{\hat{d}_1, \hat{d}_2, \dots, \hat{d}_n\}$$

com

$$\hat{d}_i = \begin{cases} s_y \left[ \frac{1}{\hat{\alpha}^2} \cosh \left( y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} \right) - \frac{1}{4} \left[ \text{sech} \left( \frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}{2} \right) \right]^2 \right], & \text{se } i \in D \\ \frac{s_y}{4} \left[ \hat{\xi}_{i2} h(\hat{\xi}_{i2}) + (\hat{\xi}_{i2})^2 h'(\hat{\xi}_{i2}) \right], & \text{se } i \in C \end{cases}$$

e  $\boldsymbol{\Delta} = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$  e com

$$\hat{c}_i = \begin{cases} \frac{s_y}{\hat{\alpha}} \hat{\xi}_{i1} \hat{\xi}_{i2}, & \text{se } i \in D \\ \frac{s_y}{2\hat{\alpha}} \left[ \hat{\xi}_{i2} h(\hat{\xi}_{i2}) + \hat{\xi}_{i1} \hat{\xi}_{i2} h'(\hat{\xi}_{i2}) \right], & \text{se } i \in C, \end{cases}$$

em que  $h(\cdot)$  e  $h'(\cdot)$  são dados em (2.3.2), e  $\hat{\xi}_{i1}$  e  $\hat{\xi}_{i2}$  são  $\xi_{i2}$  e  $\xi_{i2}$ , avaliados em  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\alpha})^\top$ .

### 4.1.3 Perturbação em uma variável explicativa

Nesse esquema, impomos uma perturbação aditiva em uma das variáveis explicativas que seja contínua. Formalmente, sujeitamos a variável explicativa  $\mathbf{x}_t$  em que  $t \in \{1, 2, \dots, p\}$  a perturbação

$$x_{it\omega} = x_{it} + \omega_i s_x,$$

em que  $\omega_i \in \mathbb{R}$ ,  $s_x$  é um fator de escala.

Aqui, a função de log-verossimilhança perturbada assume a forma

$$\ell(\boldsymbol{\theta} | \boldsymbol{\omega}) \propto \sum_{i \in D} \left\{ \log(\xi_{i1\omega_2}) - \frac{1}{2} [\log(8\pi) + \xi_{i2\omega_2}^2] \right\} + \sum_{i \in C} \log(\Phi(-\xi_{i2\omega_2})),$$

em que

$$\xi_{i1\omega_2} = \frac{2}{\alpha} \cosh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - \beta_t \omega_i s_x}{2} \right)$$

e

$$\xi_{i2\omega_2} = \frac{2}{\alpha} \sinh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta} - \beta_t \omega_i s_x}{2} \right).$$

Para esse esquema de perturbação, a matriz

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Delta}_{\boldsymbol{\beta}} \\ \boldsymbol{\Delta}_{\alpha} \end{pmatrix},$$

obtida por Leiva, Barros, Paula & Galea (2007), é tal que  $\boldsymbol{\Delta}_{\boldsymbol{\beta}}$  é uma matriz  $p \times n$  cujos elementos  $\boldsymbol{\Delta}_{\beta_{ij}}$  quando  $j \neq t$  são dados por

$$\boldsymbol{\Delta}_{\beta_{ij}} = \begin{cases} s_x \hat{\beta}_t x_{ij} \left[ \frac{1}{4} \operatorname{sech}^2 \left( \frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_j}{2} \right) - \frac{1}{\alpha^2} \cosh(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}) \right], & \text{se } i \in D; \\ -\frac{s_x \hat{\beta}_t x_{ij}}{4} \left[ \hat{\xi}_{i2} h(\hat{\xi}_{i2}) + (\hat{\xi}_{i1})^2 h'(\hat{\xi}_{i2}) \right], & \text{se } i \in C. \end{cases}$$

E quando  $j = t$ ,

$$\boldsymbol{\Delta}_{\beta_{it}} = \begin{cases} s_x \hat{\beta}_t x_{it} \left[ \frac{1}{4} \operatorname{sech}^2 \left( \frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}{2} \right) - \frac{1}{\alpha^2} \cosh(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}) \right] + \\ + s_x \left[ \frac{1}{\alpha^2} \sinh(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}) - \frac{1}{2} \tanh \left( \frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}{2} \right) \right], & \text{se } i \in D; \\ -\frac{s_x \hat{\beta}_t x_{ij}}{4} \left[ \hat{\xi}_{i2} h(\hat{\xi}_{i2}) + (\hat{\xi}_{i1})^2 h'(\hat{\xi}_{i2}) \right] + \frac{s_x}{2} \hat{\xi}_{i1} h(\hat{\xi}_{i2}), & \text{se } i \in C. \end{cases}$$

e  $\boldsymbol{\Delta}_{\alpha} = (\hat{\kappa}_1, \dots, \hat{\kappa}_n)$  com

$$\hat{\kappa}_i = \begin{cases} -\frac{2}{\alpha^3} s_x \hat{\beta}_t \sinh(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}), & \text{se } i \in D; \\ -\frac{s_x \hat{\beta}_t}{2\alpha^2} \left[ \hat{\xi}_{i2} h(\hat{\xi}_{i2}) + \frac{2}{\alpha^2} \sinh(y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}) h'(\hat{\xi}_{i2}) \right], & \text{se } i \in C. \end{cases}$$

Para mais detalhes veja Leiva, Barros, Paula & Galea (2007).

## 4.2 Aplicação

Esta seção tem como objetivo ilustrar a metodologia apresentada no Capítulo 3 nos modelos de regressão log-BS. Consideraremos um conjunto de dados referente a um estudo clínico que tem como objetivo modelar o tempo de sobrevivência ( $T$ ), em meses, de 65 pacientes com mieloma múltiplo. Para nosso estudo consideramos as seguintes variáveis:

- logaritmo da medida de uréia nitrogenada no sangue no diagnóstico ( $x_1$ );
- medida de hemoglobina no diagnóstico ( $x_2$ );
- idade no diagnóstico ( $x_3$ );
- sexo ( $x_4$ )(0: masculino e 1: feminino) e
- medida de cálcio sérico no diagnóstico ( $x_5$ )

Os dados apresentam os tempos de sobrevivência (em meses) para 65 pacientes que foram tratados com agentes *alkylating*, dos quais 48 morreram durante o estudo e apenas 17 sobreviveram. Os dados estão apresentados em Lawless (1982, p. 332-333). Estes já foram analisados por Jin, Lin, Wei & Ying (2003) e Ghosh & Ghosal (2006) que usaram um modelo semiparamétrico de tempo de falha acelerado. Recentemente, Leiva, Barros, Paula & Galea (2007) ajustaram estes dados a um modelo de regressão log-BS, usando o argumento de que em muitos problemas médicos (por exemplo, doenças cardíacas ou diferentes tipos de câncer), um tipo de dano acumulado provocado por vários fatores de risco é detectado. Então, essa degradação leva a um processo de fadiga, cuja propagação do tempo de sobrevivência pode ser adequadamente modelada pela distribuição BS. Eles fizeram uma análise de diagnóstico baseada em dois tipos de resíduos, a saber: resíduo tipo-Martingale e resíduo componente do desvio. Bem como, utilizaram a técnica de influência local, baseada na curvatura normal, para analisar possíveis pontos influentes e selecionar o modelo adequado para estes dados. Nosso objetivo aqui é analisar novamente este conjunto de dados, considerando o modelo log-BS, mas utilizando a técnica de influência local baseada na curvatura normal conforme,

uma vez que, agora, temos um valor de referência para poder julgar de forma mais objetiva se um ponto é influente ou não.

Na análise feita em Leiva, Barros, Paula & Galea (2007) foram construídos gráficos normais de probabilidade que indicaram que não existem indícios de afastamento da suposição de que o modelo de regressão log-BS seja adequado para os dados, uma vez que, em geral, os resíduos permanecem dentro das bandas de confiança dos envelopes simulados.

Assim, considerando o modelo de regressão

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \varepsilon_i, \quad i = 1, 2, \dots, 65$$

em que os erros  $\varepsilon_i \sim \text{SN}(\alpha, 0, 2)$  e a resposta  $y_i$  é o logaritmo natural do tempo de sobrevivência, temos que as estimativas de máxima verossimilhança dos coeficientes  $\beta_i$ s e do parâmetro de forma  $\alpha$ , assim como os respectivos níveis descritivos são dados na Tabela 4.1.

Tabela 4.1: Estimativas dos parâmetros e os respectivos níveis descritivos.

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\alpha}$
Estimativa	4,500	-1,596	0,142	0,010	0,209	-0,141	1,082
Nível descritivo	0,000	0,000	0,004	0,455	0,461	0,042	

Notemos que os coeficientes  $\beta_3$  e  $\beta_4$  não são significativos a um nível de significância de 10%.

Com o objetivo de verificar a existência de possíveis pontos influentes nos dados, aplicamos a técnica de influência local baseada na curvatura normal conforme, sob o esquema de perturbação ponderação de casos, tanto para as observações censuradas quanto para as não-censuradas. Consideramos os valores para  $q = 1, 2, 3$  e  $4$  a fim de determinar quais são os autovetores (direções) associadas às máximas variações da função deslocamento pela verossimilhança. Na Figura 4.1(a) podemos observar o gráfico do módulo dos autovalores normalizados e os valores de  $q$ . Temos que  $n = 65$ , e que para  $q = 4$  temos apenas um autovalor acima do valor de referência  $q/\sqrt{n}$ , para  $q = 3$  temos apenas dois autovalores acima do valor de referência, para  $q = 2$  temos

5 autovalores e para  $q = 1$  temos 7 autovalores, isto é, temos um único autovetor 4-influente, dois autovetores 3-influente, cinco autovetores 2-influente e sete autovetores 1-influente. Os autovalores normalizados associados a estes autovetores são 0,6020, 0,4764, 0,3548, 0,3230, 0,2887, 0,2376 e 0,2147. Desta forma, a máxima curvatura normal conforme  $B_{\mathbf{I}_{\max}}$  é 0,6020, ou seja  $\mathbf{I}_{\max}$  sintetiza as maiores variações da função afastamento de verossimilhança quando o modelo está sendo perturbado.

No gráfico da Figura 4.1 (b) estão apresentados os valores da contribuição agregada de todos os autovetores ( $q = 0$ ). Como podemos notar, as observações 1, 2, 3, 4, 5, 12, 40, 44, 48 e 62 estão acima do valor de referência  $2b$  sugerido por Poon & Poon (1999) e são consideradas potencialmente influentes. Já as observações 9 e 17 aparecem como sendo marginalmente influentes. Na Figura 4.2, temos as contribuições agregadas dos autovetores associados aos maiores autovalores, correspondentes a  $q = 4$  e  $q = 3$ . Neste caso, para  $q = 4$  as observações 3, 4 e 40 são detectadas como potencialmente influentes, pois estão acima do valor de referência  $\bar{m}(q)\sqrt{2}$  e a observação 44 é considerada marginalmente influente por estar muito próxima deste valor. Enquanto que para  $q = 3$  as observações influentes são 3, 4, 5, 40 e 44, e incrementamos a observação 48 como sendo marginalmente influente (ver Figura 4.2 (b)). A observação 40 é a que mais se destaca em todos os gráficos e se refere a um paciente do sexo masculino, com 74 anos de idade e que sobreviveu 51 meses ao tratamento. As taxas de uréia nitrogenada do sangue e cálcio sérico são maiores que a média dos 65 pacientes e a taxa de hemoglobina é menor que a média dos 65 pacientes. Neste ponto, destacamos o fato de que na análise feita em Leiva, Barros, Paula & Galea (2007), não foi detectada a observação 4 como influente.

Na Tabela 4.2 estão apresentados o número de autovetores  $N$  que são  $q$ -influente, os valores críticos para  $B_{\mathbf{E}_j}$  e  $m_j(q)$ , com  $q = 3, 4$  e as contribuições agregadas associadas aos autovetores.

Agora, com o objetivo de verificar o impacto que as observações 3, 4, 5, 40 e 44 exercem sobre as estimativas dos parâmetros, reajustaremos o modelo, removendo individualmente cada uma dessas observações e em seguida excluindo todas. Na Tabela 4.3 constam as mudanças relativas em porcentagem,  $RC_{\theta_j}$ , de cada estimativa, definida por  $RC_{\theta_j} = \left( \frac{\hat{\theta}_j - \hat{\theta}_{j^{(i)}}}{\hat{\theta}_j} \right) \times 100$ , em que  $\hat{\theta}_{j^{(i)}}$  denota a estimativa de máxima verossimilhança

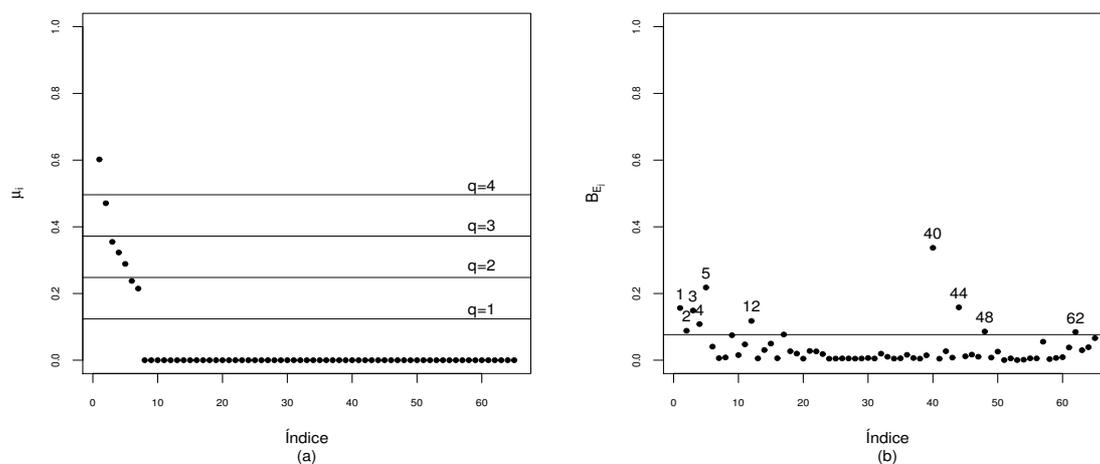


Figura 4.1: autovalores normalizados em módulo e valores de  $q$  (a), e gráfico da contribuição agregada total  $B_{E_j}$  (b), para o esquema de perturbação ponderação de casos.

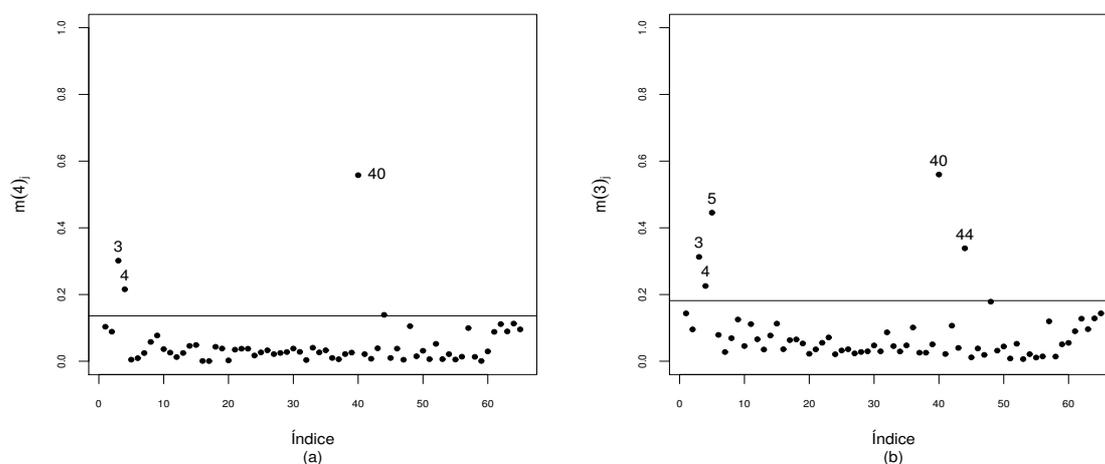


Figura 4.2: contribuição agregada do autovetor 4-influente correspondente ao maior autovalor ( $q = 4$ ) (a), e contribuição agregada dos autovetores 3-influentes correspondentes aos dois maiores autovalores ( $q = 3$ ) (b) para o esquema de perturbação ponderação de casos.

de  $\hat{\theta}_j$  obtida após remoção da  $i$ -ésima observação e os níveis descritivos correspondentes a testes sobre os coeficientes  $\beta'_i$ s. Voltamos a Tabela 4.1, a partir de uma análise sobre a mesma percebemos que, fixado o nível de significância em 10%, as variáveis explica-

Tabela 4.2: Medidas de influência correspondentes as observações influentes.

Contrib. Agregada	N	Valor de referência	obs 3	obs 4	obs 5	obs 40	obs 44
$m_j(4)$	1	0,1361	0,3015	0,2157	0,0470	0,5576	0,1961
$m_j(3)$	2	0,1817	0,3128	0,2225	0,4454	0,5598	0,3385
$B_{E_j}$	65	0,0767	0,1489	0,1083	0,2180	0,3370	0,1584

tivas  $x_3$  e  $x_4$  devem ser retiradas do modelo. Analisando a Tabela 4.3, constatamos que, ao nível de significância de 10%, com a exclusão, individualmente, de cada uma observações 3, 4, 5, 40 e 40 como também a remoção de todas elas os coeficientes associados às variáveis  $x_3$  e  $x_4$  permanecem como sendo não-significativos. Percebemos que a exclusão da observação 3 faz com que o coeficiente associado a variável  $x_5$  deixe de ser significativo. Possivelmente, a não-significância desse coeficiente tenha sido mascarada por essa observação, levando-nos a sugerir a retirada da covariável  $x_5$  do modelo. Confirmando a análise feita em Leiva, Barros, Paula & Galea (2007).

Desta forma, o modelo final fica dado por

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, \quad i = 1, \dots, 65. \quad (4.3)$$

As estimativas de máxima verossimilhança dos parâmetros (erro padrão entre parênteses) são

$$\hat{\beta}_0 = 4,409 (0,772), \quad \hat{\beta}_1 = -1,869 (0,423), \quad \hat{\beta}_2 = 0,109 (0,049) \text{ e } \hat{\alpha} = 1,140 (0,118).$$

A interpretação dos coeficientes estimados do modelo final pode ser feita da mesma forma que em Leiva et al. (2007). Ou seja, o tempo de sobrevivência esperado decresce aproximadamente 85%  $[(1 - e^{-1,869}) \times 100\%]$  quando a taxa de uréia nitrogenada do sangue (em logaritmo) cresce em uma unidade, permanecendo fixa a taxa de hemoglobina. Similarmente, o tempo de sobrevivência esperado cresce aproximadamente 12%  $[e^{0,109} \times 100\%]$  quando a taxa de hemoglobina cresce em uma unidade enquanto a taxa de uréia nitrogenada do sangue (em logaritmo) é mantida fixa.

Tabela 4.3: Mudanças relativas (RC, em %) e os correspondentes níveis descritivos entre parênteses.

Observações removidas	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\alpha}$
Observação 3	15 (0,005)	-3 (0,000)	3 (0,005)	-62 (0,250)	-4 (0,438)	30 (0,188)	1
Observação 4	11 (0,002)	5 (0,000)	-4 (0,002)	-38 (0,299)	5 (0,478)	7 (0,051)	2
Observação 5	-20 (0,000)	-3 (0,000)	25 (0,046)	42 (0,660)	68 (0,814)	-9 (0,023)	4
Observação 40	-14 (0,000)	0 (0,000)	-14 (0,000)	71 (0,826)	-12 (0,377)	-34 (0,011)	6
Observação 44	12 (0,002)	-1 (0,000)	-14 (0,001)	-22 (0,350)	-46 (0,283)	12 (0,063)	3
Observações 3,4,5,40 e 44	12 (0,009)	-1 (0,000)	-8 (0,003)	-34 (0,343)	-5 (0,406)	16 (0,119)	14

## Capítulo 5

# Influência local em modelos de regressão lineares com erro senh-normal

Neste capítulo, consideramos o modelo de regressão linear cujo erro segue uma distribuição SN. Diferentemente da abordagem do Capítulo 4, onde tínhamos a hipótese de que o parâmetro de escala  $\sigma$  era conhecido e igual a 2, agora, tratamos o caso em que  $\sigma$  é desconhecido. O interesse neste modelo é justificado porque ele é pouco explorado na literatura e sempre é visto em seu caso particular, o modelo log-BS. Porém, este tem a vantagem de ser um modelo simétrico com curtose maior ou menor do que o modelo normal e portanto, é uma alternativa interessante ao modelo normal, por ser mais flexível. Estendemos os resultados de diagnóstico de influência desenvolvidos por Leiva, Barros, Paula & Galea (2007) para observações censuradas sob modelos de regressão linear log-BS, considerando agora que o parâmetro  $\sigma$  é desconhecido. Ou seja, derivamos as curvaturas normais da influência local conforme sob os esquemas de perturbação, ponderação de casos e perturbação da variável resposta, para o modelo de regressão SN. Adicionalmente, propomos o resíduo ordinário, para avaliar possíveis afastamentos da suposição de que os erros seguem distribuição SN, bem como para

detectar observações extremas com alguma influência desproporcional nos resultados do ajuste. Consideramos o conjunto de dados *rat* analisado por Poon & Poon (1999) e através da influência local conforme para o modelo SN conseguimos detectar o ponto que estava mascarando a significatividade dos parâmetros do modelo.

## 5.1 Caracterização e estimação

Suponhamos que a resposta seja uma variável aleatória que possa ser expressa como

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (5.1)$$

em que os erros  $\varepsilon_i$ 's são variáveis aleatórias independentes e identicamente distribuídas tais que  $\varepsilon_i \sim \text{SN}(\alpha, 0, \sigma)$ , para  $i = 1, 2, \dots, n$ ,  $\mathbf{x}_i^\top$ 's são valores observados de  $p$  covariáveis,  $\boldsymbol{\beta}$  é um vetor de parâmetros desconhecidos e os parâmetros de forma  $\alpha$  e escala  $\sigma$  são desconhecidos. Sobre esse modelo interessa-nos estimar o vetor  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha, \sigma)^\top$ .

Sejam  $y_1, y_2, \dots, y_n$  observações casuais e independentes da variável resposta definida na equação (5.1). Dado que  $\varepsilon_i \sim \text{SN}(\alpha, 0, \sigma)$ , para  $i = 1, 2, \dots, n$ , temos

$$y_i \sim \text{SN}(\alpha, \mathbf{x}_i^\top \boldsymbol{\beta}, \sigma), \quad i = 1, 2, \dots, n.$$

Logo, a função de log-verossimilhança de  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha, \sigma)^\top$  baseada nas observações  $y_1, y_2, \dots, y_n$  pode ser escrita da seguinte forma

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \left\{ \log(\xi_{i1}^*) - \frac{1}{2} [\log(8\pi) + (\xi_{i2}^*)^2] \right\}, \quad (5.2)$$

em que

$$\xi_{i1}^* = \frac{2}{\alpha} \cosh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \quad (5.3)$$

e

$$\xi_{i2}^* = \frac{2}{\alpha} \sinh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right), \quad (5.4)$$

para  $i = 1, 2, \dots, n$ .

As derivadas parciais da função de log-verossimilhança, dada na equação (5.2), em relação a  $\boldsymbol{\beta}_j$  ( $j = 1, 2, \dots, p$ ),  $\alpha$  e  $\sigma$  podem ser expressas como

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} = \frac{1}{\sigma} \sum_{i=1}^n x_{ij} \left[ \xi_{i1}^* \xi_{i2}^* - \frac{\xi_{i2}^*}{\xi_{i1}^*} \right], \quad j = 1, 2, \dots, p,$$

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} = \frac{1}{\alpha} \left[ \sum_{i=1}^n \{(\xi_{i2}^*)^2 - 1\} \right],$$

e

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} = \frac{1}{\sigma^2} \sum_{i=1}^n [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \left[ \xi_{i1}^* \xi_{i2}^* - \frac{\xi_{i2}^*}{\xi_{i1}^*} \right].$$

Os estimadores de máxima verossimilhança dos coeficientes  $\beta'_j$ s, do parâmetro de forma  $\alpha$  e do parâmetro de escala  $\sigma$ , denotados por  $\hat{\beta}'_j$ s,  $\hat{\alpha}$  e  $\hat{\sigma}$ , respectivamente, devem satisfazer ao seguinte sistema de equações:

$$\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_j} \right|_{\beta_j = \hat{\beta}_j, \alpha = \hat{\alpha}, \sigma = \hat{\sigma}} = 0, \quad j = 1, 2, \dots, p,$$

$$\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} \right|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}, \alpha = \hat{\alpha}, \sigma = \hat{\sigma}} = 0 \quad \text{e}$$

$$\left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} \right|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}, \alpha = \hat{\alpha}, \sigma = \hat{\sigma}} = 0.$$

Mais uma vez, temos um sistema de equações não-lineares que não possui soluções analíticas, sendo necessário o uso de métodos numéricos para obtenção de soluções.

A matriz de variâncias e covariâncias de  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\alpha}, \hat{\sigma})^\top$  pode ser aproximada pela matriz de informação de Fisher observada,  $-\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}$ , avaliada em  $\hat{\boldsymbol{\theta}}$ . Para o modelo de regressão considerado neste capítulo, temos

$$\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{bmatrix} \ddot{\mathbf{L}}_{\beta\beta} & \ddot{\mathbf{L}}_{\beta\alpha} & \ddot{\mathbf{L}}_{\beta\sigma} \\ \ddot{\mathbf{L}}_{\alpha\beta} & \ddot{\mathbf{L}}_{\alpha\alpha} & \ddot{\mathbf{L}}_{\alpha\sigma} \\ \ddot{\mathbf{L}}_{\sigma\beta} & \ddot{\mathbf{L}}_{\sigma\alpha} & \ddot{\mathbf{L}}_{\sigma\sigma} \end{bmatrix},$$

em que as entradas da matriz  $\ddot{\mathbf{L}}_{\boldsymbol{\theta}\boldsymbol{\theta}}$  são obtidas a partir das seguintes expressões:

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \beta_l} = \sum_{i=1}^n \frac{1}{\sigma^2} x_{ij} x_{il} \left[ \operatorname{sech}^2 \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) - \frac{4}{\alpha^2} \cosh \left( \frac{2[y_i - \mathbf{x}_i^\top \boldsymbol{\beta}]}{\sigma} \right) \right],$$

para  $j, l = 1, 2, \dots, p$ .

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \beta_j} = -\frac{4}{\alpha \sigma^3} \sum_{i=1}^n x_{ij} \sinh \left( \frac{2[y_i - \mathbf{x}_i^\top \boldsymbol{\beta}]}{\sigma} \right),$$

para  $j = 1, 2, \dots, p$ .

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \alpha} = \sum_{i=1}^n \left\{ \frac{1}{\alpha^2} - \frac{3\zeta_{i2}^*}{\alpha^2} \right\}.$$

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_j \partial \sigma} &= -\sum_{i=1}^n \frac{2x_{ij}}{\sigma^2 \alpha^2} \left[ \sinh \left( \frac{2[y_i - \mathbf{x}_i^\top \boldsymbol{\beta}]}{\sigma} \right) - \frac{2}{\sigma} \cosh \left( \frac{2[y_i - \mathbf{x}_i^\top \boldsymbol{\beta}]}{\sigma} \right) \right] \\ &+ \sum_{i=1}^n \frac{1}{\sigma^2} x_{ij} \left[ \tanh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) + \frac{1}{\sigma} [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \operatorname{sech}^2 \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \right], \end{aligned}$$

para  $j = 1, 2, \dots, p$ .

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma \partial \sigma} &= -\sum_{i=1}^n [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \left[ \frac{4}{\alpha^2 \sigma^3} \sinh \left( \frac{2[y_i - \mathbf{x}_i^\top \boldsymbol{\beta}]}{\sigma} \right) \right. \\ &+ \left. \frac{4}{\sigma^4 \alpha^2} [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \cosh \left( \frac{2[y_i - \mathbf{x}_i^\top \boldsymbol{\beta}]}{\sigma} \right) \right] \\ &+ \sum_{i=1}^n \frac{2}{\sigma^3} [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \left[ \tanh \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) - \frac{1}{\sigma^4} [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \operatorname{sech}^2 \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \right]. \end{aligned}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \sigma} = -\frac{4}{\alpha^2 \sigma^3} \sum_{i=1}^n [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \sinh \left( \frac{2[y_i - \mathbf{x}_i^\top \boldsymbol{\beta}]}{\sigma} \right).$$

## 5.2 Influência local

Para este modelo iremos considerar apenas os esquemas de perturbação: ponderação de casos e perturbação da variável resposta.

### 5.2.1 Ponderação de casos

Consideremos o modelo de regressão definido na equação (5.1) e o vetor de perturbação  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^\top \in \Omega \subset \mathbb{R}^n$ . A função de log-verossimilhança perturbada para o

esquema de ponderação de casos é

$$\ell(\boldsymbol{\theta}|\boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \left\{ \log(\xi_{i1}^*) - \frac{1}{2} [\log(8\pi) + (\xi_{i2}^*)^2] \right\}, \quad (5.5)$$

em que  $0 \leq \omega_i \leq 1$ , para  $i = 1, 2, \dots, n$  e  $\boldsymbol{\omega}_0 = (1, 1, \dots, 1)^\top$  é o vetor de não-perturbação.

Diferenciando a função de log-verossimilhança perturbada dada pela equação (5.5) em relação a  $\beta_j$  com  $j = 1, 2, \dots$ , obtemos

$$\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \beta_j} = \sum_{i=1}^n \omega_i \left[ \frac{x_{ij}}{\sigma} \left( \xi_{i1}^* \xi_{i2}^* - \frac{\xi_{i2}^*}{\xi_{i1}^*} \right) \right]$$

para  $j = 1, 2, \dots, p$ . Logo,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i \partial \beta_j} = \frac{x_{ij}}{\sigma} \left[ \xi_{i1}^* \xi_{i2}^* - \frac{\xi_{i2}^*}{\xi_{i1}^*} \right],$$

para  $j = 1, 2, \dots, p$  e  $i = 1, 2, \dots, n$ . Por outro lado,

$$\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \alpha} = \frac{1}{\alpha} \left[ \sum_{i=1}^n \omega_i [(\xi_{i2}^*)^2 - 1] \right].$$

Daí,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i \partial \alpha} = \frac{1}{\alpha} [(\xi_{i2}^*)^2 - 1],$$

para  $i = 1, 2, \dots, n$ . Por fim,

$$\frac{\partial \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \sigma} = \frac{1}{\sigma^2} \sum_{i=1}^n \omega_i [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \left[ \xi_{i1}^* \xi_{i2}^* - \frac{\xi_{i2}^*}{\xi_{i1}^*} \right];$$

logo,

$$\frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i \partial \sigma} = \frac{1}{\sigma^2} [y_i - \mathbf{x}_i^\top \boldsymbol{\beta}] \left[ \xi_{i1}^* \xi_{i2}^* - \frac{\xi_{i2}^*}{\xi_{i1}^*} \right],$$

para  $i = 1, 2, \dots, n$ .

As derivadas parciais obtidas anteriormente são úteis na determinação da matriz

$$\boldsymbol{\Delta} = (\Delta)_{ij} = \left( \frac{\partial^2 \ell(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \omega_i \partial \theta_j} \right) \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p+2$$

e, conseqüentemente, no cálculo das curvaturas.

### 5.2.2 Perturbação na variável resposta

Neste caso, como descrito no Capítulo 4, sujeitamos a resposta  $y_i$  a uma perturbação da forma  $y_{i\omega} = y_i + \omega_i s_y$ , em que  $\omega_i \in \mathbb{R}$  para todo  $i = 1, 2, \dots, n$ ,  $s_y$  é um fator de escala. Consideremos o modelo de regressão definido em (5.1) e o vetor de perturbação  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_n)^\top \in \Omega \subset \mathbb{R}^n$ . A função de log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\theta} | \boldsymbol{\omega}) = \sum_{i=1}^n \left\{ \log(\xi_{i1\omega_1}^*) - \frac{1}{2} [\log(8\pi) + (\xi_{i2\omega_1}^*)^2] \right\}$$

em que

$$\xi_{i1\omega_1}^* = \frac{2}{\alpha} \cosh \left( \frac{y_{i\omega} - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right) \quad (5.6)$$

e

$$\xi_{i2\omega_1}^* = \frac{2}{\alpha} \sinh \left( \frac{y_{i\omega} - \mathbf{x}_i^\top \boldsymbol{\beta}}{\sigma} \right). \quad (5.7)$$

Para esse esquema de perturbação, a matriz  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta, \boldsymbol{\Delta}_\alpha, \boldsymbol{\Delta}_\sigma)^\top$  é tal que

$$\boldsymbol{\Delta}_\beta = \mathbf{X}^\top \text{diag}\{\hat{d}_1, \hat{d}_2, \dots, \hat{d}_n\}$$

com

$$\hat{d}_i = \frac{s_y}{\hat{\sigma}^2} \left[ \frac{4}{\hat{\alpha}^2} \cosh \left( \frac{2[y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}]}{\hat{\sigma}} \right) - \text{sech}^2 \left( \frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}{\hat{\sigma}} \right) \right],$$

$\boldsymbol{\Delta}_\alpha = (\hat{c}_1, \hat{c}_2, \dots, \hat{c}_n)$  com

$$\hat{c}_i = \frac{2s_y}{\hat{\alpha}\hat{\sigma}} \xi_{i1}^* \xi_{i2}^*,$$

e  $\boldsymbol{\Delta}_\sigma = (\hat{s}_1, \hat{s}_2, \dots, \hat{s}_n)$ , sendo

$$\hat{s}_i = \frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}{\hat{\sigma}^2} \left[ -\tanh \left( \frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}{\hat{\sigma}} \right) + \frac{2}{\hat{\alpha}^2} \sinh \left( \frac{2[y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}]}{\hat{\sigma}} \right) \right],$$

em que  $\hat{\xi}_{i1}^*$  e  $\hat{\xi}_{i2}^*$  são  $\xi_{i2}^*$  e  $\xi_{i1}^*$ , dadas nas equações (5.6) e (5.7), respectivamente, avaliados em  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^\top, \hat{\alpha}, \hat{\sigma})^\top$ .

## 5.3 Aplicação

Esta aplicação tem como objetivo enfatizar a importância do modelo de regressão linear senh-normal. Para isso, consideramos o conjunto de dados *rat* analisado em Poon

& Poon (1999) apresentado em Weisberg (1985, p.121). Estes dados se referem a uma experiência que foi conduzida para investigar a quantidade de absorção de uma particular droga no fígado de um rato. Dezenove ratos foram selecionados aleatoriamente, pesados, anestesiados e submetidos a uma dose oral da droga. Como fígados grandes deveriam absorver mais do que fígados menores, a dose real que um animal recebeu foi determinada aproximadamente como 40 mg da droga por quilograma da massa corporal. Sabe-se que a massa do fígado está relacionado fortemente com a massa corporal. Após um tempo fixado, cada rato foi sacrificado, tendo seu fígado pesado, e a porcentagem da droga no fígado determinada. A hipótese experimental era a inexistência de relação linear entre a porcentagem da dose no fígado e a massa corporal, a massa do fígado e a dose relativa.

Para nosso estudo, consideramos as seguintes variáveis: porcentagem da dose no fígado ( $Y$ ), a massa corporal ( $x_1$ ), a massa do fígado ( $x_2$ ) e a dose relativa ( $x_3$ ). O interesse recai em modelar  $Y$  que é a porcentagem da dose administrada que foi retida pelo fígado.

Na Figura 5.1 apresentamos gráficos de dispersão para cada variável explicativa contra a variável resposta. Estes revelam que nenhuma das variáveis em estudo parece ter uma relação linear com a porcentagem da dose no fígado. É importante salientar que estes diagramas detectam uma possível relação entre a resposta e as demais variáveis individualmente. Assim não podendo ser devidamente avaliada a relação entre as variáveis dependente e independente conjuntamente.

Na Tabela 5.1, estão dispostas as correlações amostrais entre as variáveis. A análise desta tabela é útil no sentido de detectar multicolinearidade e de verificar relação linear entre a resposta e cada um dos preditores. Como podemos observar, as correlações amostrais entre as variáveis explicativas com a resposta são pequenas, sugerindo que talvez não exista uma relação linear entre elas, ratificando a análise feita através dos gráficos de dispersão. Porém vale ressaltar que a correlação amostral entre as variáveis  $x_1$  e  $x_3$  é altíssima (0,9902).

Por meio de técnicas clássicas da análise de diagnóstico em modelos de regressão normal linear, Weisberg (1985) destacou que a distância de Cook é relativamente grande para a terceira observação. Posteriormente, Cook (1986) detectou como sendo local-

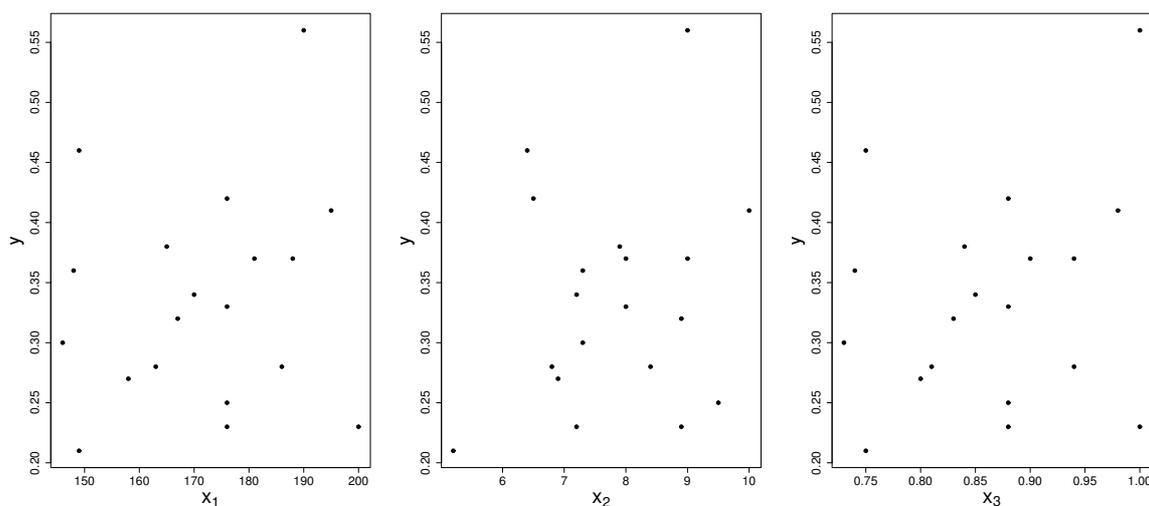


Figura 5.1: diagramas de dispersão entre a variável reposta e as variáveis explicativas.

Tabela 5.1: Matriz de correlação amostral entre as variáveis

	$y$	$x_1$	$x_2$	$x_3$
$y$	1,0000	0,1511	0,2033	0,2275
$x_1$	0,1511	1,0000	0,5000	0,9902
$x_2$	0,2033	0,5000	1,0000	0,4901
$x_3$	0,2275	0,9902	0,4901	1,0000

mente influentes os casos em que os valores da variável  $x_2$  são relativamente pequenos. Mais recentemente, Poon & Poon (1999), através do método de influência local baseado na curvatura normal conforme, sob o esquema de ponderação de casos, constataram as observações 1, 13 e 19 como potencialmente influentes e a observação 5 como marginalmente influente. Porém, os autores se detêm na demonstração da técnica abordada e não enfatizam a questão da significatividade das estimativas dos parâmetros do modelo. Ou seja, se considerarmos o modelo normal linear e fixarmos o nível de significância em 10%, o coeficiente associado a variável  $x_2$  não é significativo. Fizemos uma análise confirmatória com o objetivo de verificar o impacto destas observações influentes nas estimativas dos parâmetros do modelo e observamos que, retirando individualmente e conjuntamente estas observações, o coeficiente associado a  $x_2$  continua não sendo sig-

nificativo para o modelo e as estimativas do coeficiente desta variável sofre as maiores variações, como podemos observar na Tabela 5.2. O que nos levaria a escolher um modelo com apenas as variáveis  $x_1$  e  $x_3$ , ou seja, a massa corporal e a dose relativa explicariam a porcentagem da dose no fígado, contradizendo a hipótese experimental e a análise descritiva.

Tabela 5.2: Mudanças relativas (RC, em %) e os correspondentes níveis descritivos.

Observações removidas	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	-2 (0,1504)	-11 (0,0069)	-78 (0,1540)	-8 (0,0064)
5	38 (0,4487)	15 (0,0482)	66 (0,7991)	10 (0,0294)
13	-54 (0,0660)	-5 (0,0108)	88 (0,9240)	-4 (0,0096)
19	56 (0,5469)	8 (0,0174)	-32 (0,2473)	6 (0,0123)
1, 5, 13 e 19	54 (0,5516)	7 (0,0151)	-4 (0,4612)	4 (0,0081)

Iremos analisar estes dados sob a perspectiva do modelo de regressão senh-normal, utilizando como ferramenta de diagnóstico o método de influência local baseado na curvatura normal conforme abordado por Poon & Poon (1999). Uma vez admitido o modelo de regressão SN definido na equação (5.1) na modelagem, obtemos as estimativas de máxima verossimilhança dos coeficientes  $\beta'_i$ s e dos parâmetros de forma  $\alpha$  e escala  $\sigma$  (ver Tabela 5.3). Como podemos observar, o coeficiente associado a variável  $x_2$  não se mostra significativo.

Tabela 5.3: Estimativas e os respectivos níveis descritivos

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\alpha}$	$\hat{\sigma}$
Estimativas	0,0141	-0,0217	0,0144	4,2681	$3,9903 \cdot 10^{-5}$	3494.756
Nível descritivo	0,1275	0,0025	0,3522	0,0019	0,0000	0,0000

Fizemos uma análise de resíduos com o objetivo de verificar a existência de observações extremas bem como detectar possíveis afastamentos das suposições feitas para

o modelo. Para isto, utilizamos o resíduo ordinário padronizado, dado por

$$r_i = \frac{y_i - E(y_i)}{\sqrt{\text{Var}(y_i)}}.$$

O gráfico normal de probabilidade do resíduo  $r_i$  com os envelopes gerados é apresentado na Figura 5.2. Como podemos observar, no gráfico, não há indícios de afastamento da suposição de que os erros têm distribuição SN.

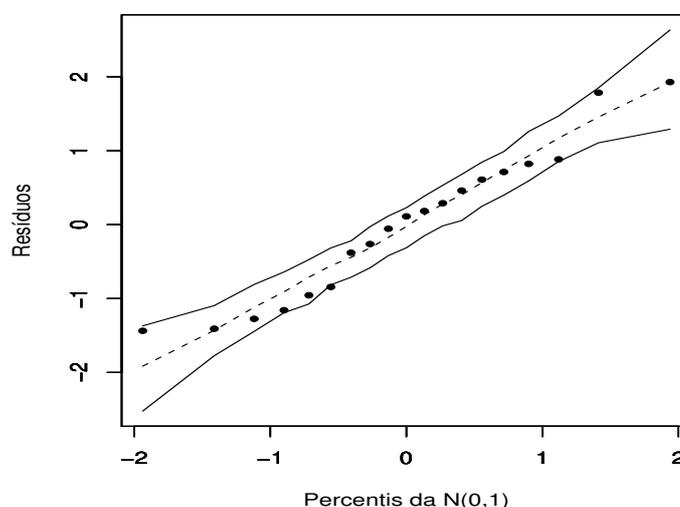


Figura 5.2: gráfico normal de probabilidade para os resíduos padronizados com envelope.

Em seguida, com o objetivo de verificar a existência de possíveis pontos influentes nos dados, aplicamos a técnica de influência local baseada na curvatura normal conforme, sob os seguintes esquemas de perturbação: ponderação de casos e perturbação da variável resposta. Para o esquema ponderação de casos, obtivemos os seguintes autovalores normalizados não-nulos: 0,6026; 0,6322; 0,3519; 0,2607; 0,2084 e 0,0442. Implicando em 2 autovetores 2-influente e 4 autovalores 1-influente (ver Figura 5.3).

Na Figura 5.4 estão apresentados os valores da contribuição agregada de todos os autovetores ( $q = 0$ ). Como podemos notar, as observações 1, 5, 13 e 19 são consideradas potencialmente influentes, porém, as observações 2 e 4 podem ser classificadas como sendo marginalmente influentes, pois o módulo da diferença entre o valor de referência

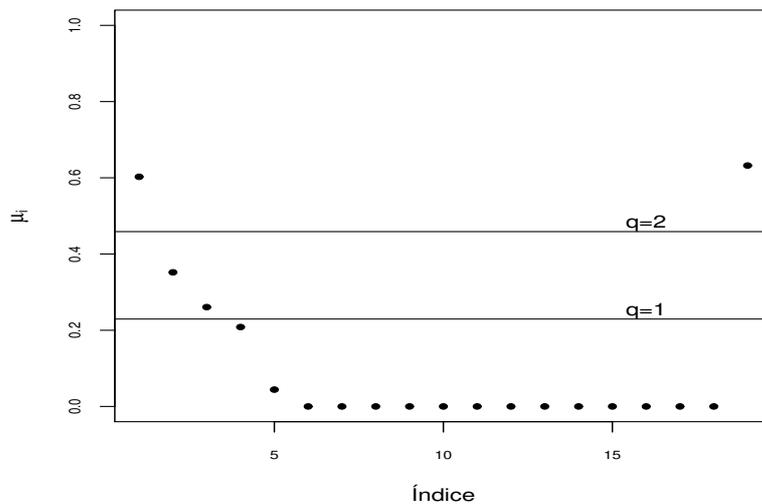


Figura 5.3: autovalores normalizados em módulo e valores de  $q$  para o esquema de perturbação ponderação de casos.

e a contribuição total  $B_{E_j}$  é aproximadamente zero. Na Figura 5.5, constatamos que as observações 1, 13 e 19 são potencialmente influentes quando  $q = 2$  (maior magnitude detectada (a)); quando  $q = 1$ , exatamente as mesmas observações foram detectadas como sendo potencialmente influentes (b).

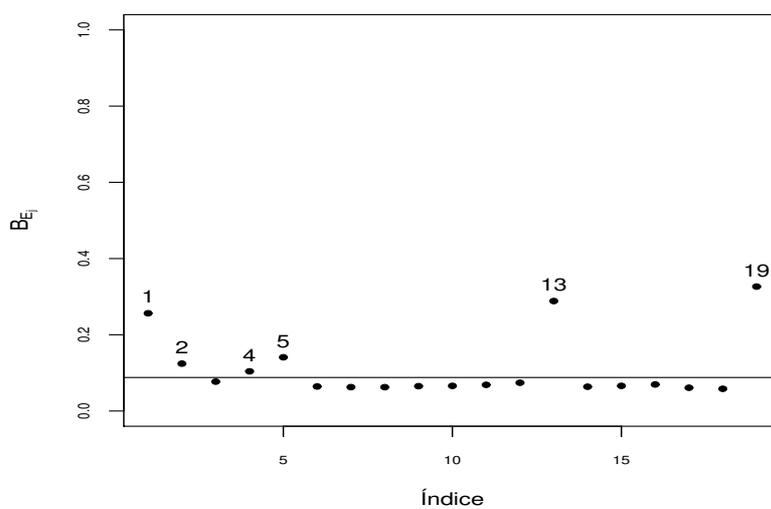


Figura 5.4: gráfico da contribuição agregada total  $B_{E_j}$ .

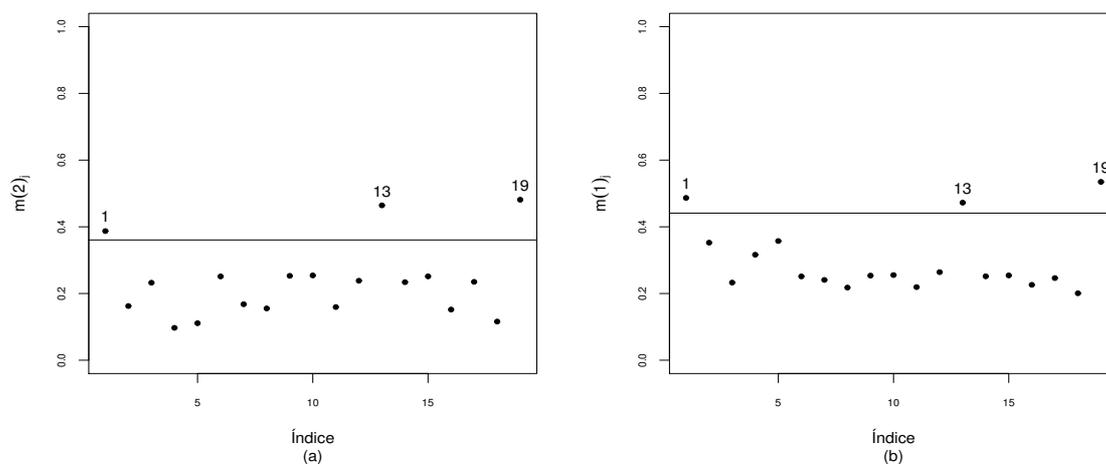


Figura 5.5: contribuição agregada dos autovetores 2-influente correspondente ao maior autovalor ( $q = 2$ ), e (b) contribuição agregada dos autovetores 1-influente correspondentes ( $q = 1$ ) provenientes do esquema de perturbação da ponderação de casos.

A Tabela 5.4, sumariza os resultados obtidos.

Tabela 5.4: medidas de influência utilizando a curvatura normal conforme para o esquema de perturbação ponderação de casos.

Contrib. agregada	N	Valor de referência	Obs 1	Obs 13	Obs 19
$m_j(2)$	2	0,3605	0,3872	0,4627	0,04811
$m_j(1)$	4	0,4409	0,4866	0,4724	0,5347
$B_{E_j}$	19	0.0880	0,2565	0,2885	0,3264

Para o esquema de perturbação da variável resposta, obtivemos cinco autovalores normalizados não-nulos: 0,5000; 0,4995 (de multiplicidade 2); 0,4991 e 0,0454. Implicando em 4 autovetores 1-influente (ver Figura 5.6).

Na Figura 5.7 (a) e (b), estão apresentados os valores da contribuição agregada de todos os autovetores ( $q = 0$ ) e a contribuição agregada dos autovetores 1-influente ( $q = 1$ ), respectivamente, considerando a perturbação na variável resposta. Como podemos notar, apenas a observação 3 se destaca como potencialmente influente em ambos os casos.

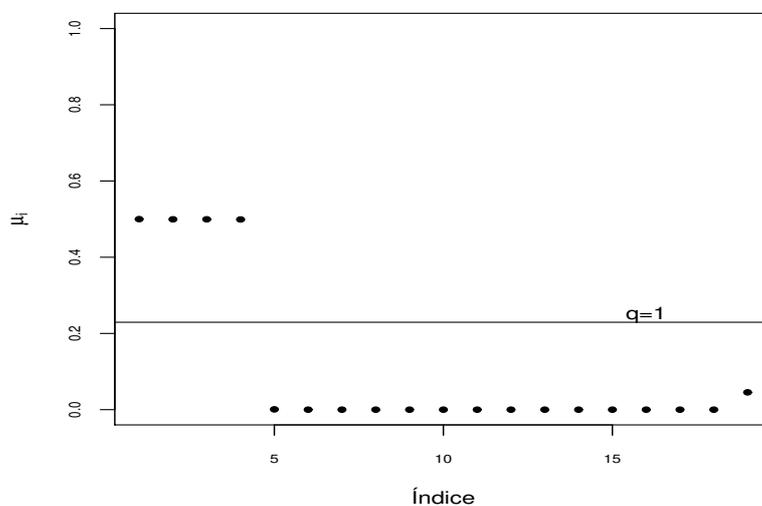


Figura 5.6: autovalores normalizados em módulo e valores de  $q$  para o esquema de perturbação da resposta.

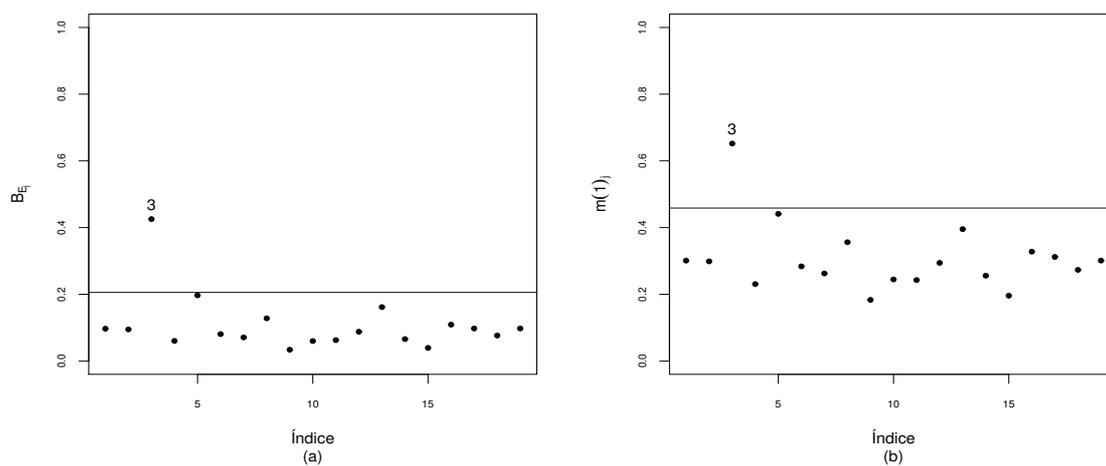


Figura 5.7: contribuição agregada de todos autovetores (a) e a contribuição agregada dos autovetores 1-influente (b) provenientes do esquema de perturbação da variável resposta.

Com a finalidade de verificar o impacto que as observações 1, 3, 13 e 19 exercem sobre as estimativas dos parâmetros, reajustaremos o modelo mediante a exclusão dessas observações. Primeiramente, retiraremos cada uma delas individualmente. Em seguida, excluïremos todas. Na Tabela 5.5 constam as mudanças relativas em porcentagem,  $RC_{\theta_j}$ , de cada estimativa e os níveis descritivos correspondentes a testes sobre os coeficientes  $\beta'_i$ s. Como podemos observar, o maior impacto é causado quando retiramos a observação 3 isoladamente e conjuntamente com as demais. Além disso, há uma mudança inferencial importantíssima, pois na eliminação da observação 3 tanto individualmente como conjuntamente nenhum dos coeficientes é significativo para este modelo, ou seja, a não-significância dos coeficientes estava sendo mascaradas pela observação 3. Esta observação se refere a um rato que tinha massa corporal e massa do fígado maior que a massa média corporal e que a massa média do fígado dos demais ratos, recebeu a dose máxima da droga e teve maior absorção da droga pelo fígado dentre os 19 ratos. Portanto, nenhuma das variáveis, massa corporal, massa do fígado e dose relativa tem relação linear com a porcentagem da dose da droga no fígado, confirmando a hipótese experimental e a análise descritiva. Neste aspecto o diagnóstico de influência foi fundamental para a análise destes dados.

Tabela 5.5: mudanças relativas (RC, em %) e os correspondentes níveis descritivos.

Observações removidas	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
1	-2 (0,0844)	-9 (0,0003)	-76 (0,0873)	-6 (0,0003)
3	-16 (0,0867)	64 (0,6353)	37 (0,5790)	65 (0,6482)
13	-52 (0,0240)	-2 (0,0010)	88 (0,9117)	-1 (0,0007)
19	57 (0,4906)	12 (0,0027)	-30 (0,1725)	9 (0,0014)
1, 3, 13 e 19	15 (0,1608)	44 (0,3191)	-35 (0,1938)	45 (0,3306)

# Conclusão

Neste trabalho apresentamos a formulação da influência local avaliada através da curvatura normal conforme. Fundamentamos a idéia de que a curvatura normal conforme tem propriedades que nos dão suporte para construirmos valores de referência que permitem julgarmos a grandeza dessa curvatura. Aplicamos esta teoria em modelos de regressão log-BS. Consideramos também o modelo de regressão SN onde derivamos as curvaturas normais para dois esquemas de perturbação e apresentamos uma alternativa para se conduzir uma análise residual nesse tipo de modelo. A metodologia utilizada foi ilustrada através de dados reais, mostrando que a análise de diagnóstico desempenha um papel importante na seleção do modelo apropriado.

# Trabalhos futuros

Uma análise análoga a realizada neste trabalho pode ser feita mediante modificações na definição de contribuição agregada. Por exemplo, Poon & Poon (1999) definem a contribuição agregada de segunda ordem do  $j$ -ésimo vetor básico de perturbação  $\mathbf{E}_j$  para todos os autovetores  $q$ -influentes por

$$M(q)_j = \left[ \sum_{i=1}^k \mu_{(i)}^2 a_{ij}^2 \right]^{\frac{1}{2}}.$$

Neste caso,

$$\overline{M}(q) = \left[ \frac{1}{n} \sum_{i=1}^k \mu_{(i)}^2 \right]^{\frac{1}{2}}$$

deverá ser usado para construir valores de referência.

Essa modificação proposta por Poon & Poon (1999) sugere uma extensão na definição de contribuição agregada. Sendo assim, poderíamos definir a contribuição agregada de ordem  $p$  do  $j$ -ésimo vetor básico de perturbação  $\mathbf{E}_j$  para todos os autovetores  $q$ -influentes por

$$m^{(p)}(q)_j = \left[ \sum_{i=1}^k \mu_{(i)}^p a_{ij}^2 \right]^{\frac{1}{2}}.$$

Particularmente, para  $p = 1$  a equação acima se reduz a definição de contribuição agregada dada na Definição 3.7, que foi usada em nossa análise. E, para  $p = 2$  a mesma se reduz a equação sugerida por Poon & Poon (1999).

Se a contribuição agregada de ordem  $p$  é a mesma para todos os autovetores básicos de perturbação, então cada uma delas, denotadas uniformemente por  $\overline{m}^{(p)}$ , é

igual a

$$\bar{m}^{(p)} = \left[ \frac{1}{n} \sum_{i=1}^k \mu_{(i)}^p \right]^{\frac{1}{2}}.$$

Neste caso,  $\bar{m}^{(p)}$  deverá ser usado para construir valores de referência.

# Apêndice A: espaços vetoriais com produto interno

As noções de comprimento, distância e a relação de perpendicularismo podem ser estabelecidas de modo natural a partir do conceito de produto interno. Nesta seção, apresentaremos alguns resultados envolvendo espaços vetoriais munidos de um produto interno. As definições e resultados que serão apresentados podem ser encontrados em livros básicos da literatura que abordam Álgebra Linear, por exemplo, ver Monteiro, 1969.

**Definição .1** *Chama-se produto interno sobre um espaço vetorial real  $E$  a toda aplicação  $\langle \cdot, \cdot \rangle: E \times E \rightarrow \mathbb{R}$  que satisfaz as seguintes condições:*

$$(i) \langle u + v; w \rangle = \langle u; w \rangle + \langle v; w \rangle;$$

$$(ii) \langle \alpha u; v \rangle = \alpha \langle u; v \rangle;$$

$$(iii) \langle u; v \rangle = \langle v; u \rangle;$$

$$(iv) \langle u; u \rangle > 0 \quad \text{se } u \neq 0.$$

**Definição .2** *Chama-se norma de um vetor  $u$ , de um espaço euclidiano  $E$ , ao número real não-negativo*

$$\|u\| = \sqrt{\langle u; u \rangle}.$$

*Particularmente, se  $\|u\| = 1$ , diz-se que  $u$  é um vetor unitário.*

**Proposição .3** *Seja  $\langle \cdot, \cdot \rangle : E \times E \rightarrow \mathbb{R}$  um produto interno sobre o espaço vetorial  $E$ , então vale as relações:*

$$\left\langle \sum_{i=1}^n \alpha_i u_i; v \right\rangle = \sum_{i=1}^n \alpha_i \langle u_i; v \rangle$$

e

$$\left\langle u; \sum_{j=1}^s \beta_j v_j \right\rangle = \sum_{j=1}^s \beta_j \langle u; v_j \rangle.$$

**Proposição .4** *(Desigualdade de Cauchy-Schwarz). Se  $u$  e  $v$  são dois vetores quaisquer de um espaço vetorial  $E$  com produto interno, então vale a seguinte desigualdade:*

$$|\langle u; v \rangle| \leq \|u\| \|v\|.$$

*Particularmente, a igualdade ocorre se, e somente se,  $u$  e  $v$  são linearmente dependentes.*

**Proposição .5** *Se  $u$  e  $v$  são dois vetores quaisquer de um espaço vetorial  $E$  com produto interno e se  $\alpha$  é um número real qualquer, são válidas as propriedades:*

$$(i) \|u\| > 0 \text{ se } u \neq 0;$$

$$(ii) \|\alpha u\| = |\alpha| \|u\|;$$

$$(iii) \|u + v\| \leq \|u\| + \|v\| \text{ (Desigualdade triangular).}$$

**Definição .6** *Sejam  $u$  e  $v$  vetores de um espaço  $E$  com produto interno. Diz-se que  $u$  é ortogonal a  $v$  se, e somente se,  $\langle u; v \rangle = 0$ .  $E$  denotamos tal fato por  $u \perp v$ .*

**Definição .7** *Sejam  $S$  e  $T$  subconjuntos de um espaço vetorial  $E$  com produto interno. Diz-se que  $S$  é ortogonal a  $T$ , notação  $S \perp T$ , se, e somente se, todo vetor de  $S$  é ortogonal a qualquer vetor de  $T$ .*

**Definição .8** *Diz-se que uma família  $\{u_i\}_{1 \leq i \leq p}$  de vetores de  $E$ , é ortogonal se, e somente se,  $u_i \perp u_j$  para  $i \neq j$  ( $1 \leq i, j \leq p$ ).*

**Definição .9** Diz-se que uma família  $\{u_i\}_{1 \leq i \leq p}$ , de vetores de  $E$ , é ortonormal se, e somente se,

- (i) a família  $\{u_i\}_{1 \leq i \leq p}$  é ortogonal;
- (ii)  $\|u_i\| = 1$  para todo  $i = 1, 2, \dots, p$ .

**Definição .10** Seja  $E$  um espaço com produto interno e de dimensão finita  $n$ ; toda base de  $E$  que é uma família ortonormal é denominada base ortonormal de  $E$ .

**Proposição .11** Seja  $E$  um espaço com produto interno e de dimensão finita  $n$ ; se  $\{u_i\}$  é uma base de  $E$ , então para todo índice  $p$ ,  $1 \leq p \leq n$ , existe uma família  $V_p = \{v_j\}_{1 \leq j \leq p}$  de vetores de  $E$ , tal que

- (i)  $V_p$  é ortonormal;
- (ii) Cada vetor  $v_j$  ( $i \leq j \leq p$ ) é uma combinação linear da família  $\{u_i\}_{1 \leq i \leq j}$ .

**Proposição .12** Todo espaço com produto interno e de dimensão finita tem uma base ortonormal.

**Proposição .13** Seja  $E$  um espaço com produto interno e de dimensão finita  $n$  e seja  $\{u_i\}_{1 \leq i \leq n}$  uma base ortonormal de  $E$ .

- (i) Para todo  $u \in E$ , temos  $u = \sum_{i=1}^n \langle u; u_i \rangle u_i$ , isto é, as coordenadas de  $u$  com relação à base ortonormal  $\{u_i\}_{1 \leq i \leq n}$  são  $\langle u; u_1 \rangle, \langle u; u_2 \rangle, \dots, \langle u; u_n \rangle$ ;
- (ii) Se  $u$  e  $v$  são dois vetores quaisquer de  $E$ , então

$$\langle u; v \rangle = \sum_{i=1}^n \langle u; u_i \rangle \langle v; u_i \rangle;$$

- (iii) Para todo vetor  $u$  de  $E$ , temos

$$\|u\|^2 = \sum_{i=1}^n |\langle u; u_i \rangle|^2.$$

## Apêndice B: estatísticas de ordem

Neste apêndice, apresentamos um resultado básico sobre estatísticas de ordem. Caso o leitor deseje um tratamento mais aprofundado acerca do tema, recomendamos David (1970).

**Definição .14** *Seja  $X_1, X_2, \dots, X_n$  uma amostra casual de tamanho  $n$  de uma variável aleatória  $X$  contínua. As variáveis  $X_i$ 's em ordem não-decrescente de magnitude, e denotadas por*

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$$

*são chamadas estatísticas de ordem correspondente à amostra aleatória  $X_1, X_2, \dots, X_n$ .*

**Proposição .15** *A função de densidade de probabilidade conjunta de  $X_{(l_1)}, X_{(l_2)}, \dots, X_{(l_k)}$ , em que  $1 \leq l_1 < l_2 < \dots < l_k \leq n$  e  $1 \leq k \leq n$  pode ser escrita na forma*

$$f(x_{(l_1)}, x_{(l_2)}, \dots, x_{(l_k)}) = \frac{n!}{\prod_{i=1}^{k+1} (l_i - l_{i-1} - 1)!} \prod_{i=1}^{k+1} [F_X(x_{(l_i)}) - F_X(x_{(l_{i-1})})]^{l_i - l_{i-1} - 1} \prod_{i=1}^k f_X(x_{(l_i)}),$$

*em que  $x_{(l_1)} \leq x_{(l_2)} \leq \dots \leq x_{(l_k)}$  e, por conveniência, definimos  $l_0 = 0$ ,  $l_{k+1} = n + 1$ ,  $x_{l_0} = -\infty$  e  $x_{l_{k+1}} = +\infty$ .*

Três casos particulares do resultado acima são os seguintes.

(i) A fdp conjunta de  $Y = (X_{(1)}, X_{(2)}, \dots, X_{(r)})$ , com  $r \leq n$ , é dada por

$$f_Y(x_{(1)}, x_{(2)}, \dots, x_{(r)}) = \frac{n!}{(n-r)!} \left[ \prod_{i=1}^r f_X(x_{(i)}) \right] [1 - F_X(x_{(r)})]^{n-r}.$$

(ii) A fdp de  $X_{(i)}$ , com  $1 \leq i \leq n$ , é dada por

$$f_{X_{(i)}}(x_{(i)}) = \frac{n!}{(i-1)!(n-i)!} f_X(x_{(i)}) [F_X(x_{(i)})]^{i-1} [1 - F_X(x_{(i)})]^{n-i}.$$

(iii) A fdp conjunta de  $(X_{(i)}, X_{(j)})$ , para  $i < j$ , é dada por

$$\begin{aligned} f_{(X_{(i)}, X_{(j)})}(x_{(i)}, x_{(j)}) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} f_X(x_{(i)}) f_X(x_{(j)}) [F_X(x_{(i)})]^{i-1} \times \\ &\times [F_X(x_{(j)}) - F_X(x_{(i)})]^{j-i-1} [1 - F_X(x_{(j)})]^{n-j}. \end{aligned}$$

# Bibliografia

- [1] Achcar, J.A. (1993) Inferences for the Birnbaum-Saunders fatigue life model using Bayesian methods. *Comp. Stat. Data Anal.* 15:67-380.
- [2] Ahmad, L.A. (1988) Jackknife estimation for a family of life distributions. *J. Stat. Comp. Simul.* 29:211-23.
- [3] Barros, M., Paula, G.A., Leiva, V. (2008) A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. *Lifetime Data Anal.* (in press).
- [4] Birnbaum, Z.W., Saunders, S.C. (1969a) A new family of life distributions. *J. Appl. Prob.* 6:319-327.
- [5] Birnbaum, Z.W., Saunders, S.C. (1969b) Estimation for a family of life distributions with applications to fatigue. *J. Appl. Prob.* 6:328-347.
- [6] Chang, D. S., Tang, L. C. (1994) Graphical analysis for Birnbaum-Saunders distribution. *Microel. Rel.* 34:17-22.
- [7] Colosimo, E. A., Giolo, S. R. (2006) *Análise de Sobrevivência Aplicada*. Edgard Blücher, São Paulo.
- [8] Cook R. D. (1986) Assessment of local influence (with discussion) *J. R. Stat. Soc. B* 48:133-169.
- [9] Cordeiro, G. M. (1992) *Introdução à Teoria da Verossimilhança* Associação Brasileira de Estatística. Rio de Janeiro, X SINAPE.
- [10] Cysneiros, A. Cribari-Neto, F., Araujo, C.G.J. (2008) On Birnbaum-Saunders Inference. *Comp. Stat. Data Anal.* 52:4939-4950.

- [11] Desmond, A. (1985) Stochastic models of failure in random environments. *Can. J. Stat.* 13:171-183.
- [12] David, H. A. (1970) *Order Statistics*. New York: Wiley.
- [13] Díaz-García, J.A., Leiva, V. (2005) A new family of life distributions based on elliptically contoured distributions. *J. Stat. Plan. Infer.* 128:445-457 (Erratum: *J. Stat. Plan. Infer.* 137, 1512-1513).
- [14] Dupuis, D.J., Mills J.E. (1998) Robust estimation of the Birnbaum- Saunders distribution. *Trans. Rel.* 47:88-95.
- [15] Engelhardt, M., Bain, L.J., Wright, F.T. (1981) Inferences on the parameters of the Birnbaum-Saunders fatigue life distribution based on maximum likelihood estimation. *Technometrics* 23:251-256.
- [16] Escobar, L. A., Meeker, W. Q. (1992) Assessing influence in regression analysis with censored data, *Biometrics*, 48:507-528.
- [17] From, S.G., Li, L. (2006) Estimation of the parameters of the Birnbaum-Saunders distribution. *Comm. Stat. Theor. Meth.* 35:2157-2169.
- [18] Galea, M., Leiva, V., Paula, G.A. (2004) Influence diagnostics in log-Birnbaum-Saunders regression models. *J. Appl. Stat.* 31, 1049-1064.
- [19] Ghosh, S. K., Ghosal, S. (2006) Semiparametric Accelerated Failure Time Models for Censored Data. In: Upadhyay, S.K., Singh, U., Dey, D.K., (Eds.), *Bayesian Statistics and its Applications*, New Delhi, Anamaya Publishers, 213-229.
- [20] Goldthwaite, L. (1961) Failure rate study for the lognormal lifetime model *Proc. Seventh Nat. Symp. Reliab. Qual. Control.* 208-213.
- [21] Gradshteyn, I.S., Ryzhik, I.M. (2000) *Table of Integrals, Series, and Products*. Academic Press, New York.
- [22] Jin, Z., Lin, D. Y., Wei, L. J., Ying, Z. (2003) Rank-based inference for the accelerated failure time model. *Biometrika* 90: 341-353.
- [23] Johnson, N.L., Kotz, S., Balakrishnan, N. (1995) *Continuous Univariate distributions* Vol-2. Wiley, New York.
- [24] Krall, J. V., Uthoff, V. A., Harley, J. B. (1975) A step-up procedure for selecting variables associated with survival. *Biometrics* 31: 49-57.
- [25] Lawless, J. F. (1982) *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

- [26] Leiva, V., Barros, M., Paula, G.A., Galea, M. (2007) Influence diagnostics in log-Birnbaum-Saunders regression models with censored data. *Comp. Stat. Data Anal.* 51:5694-5707.
- [27] Leiva, V., Riquelme, M., Balakrishnan, N., Sanhueza, A. (2008) Lifetime analysis based on the generalized Birnbaum-Saunders distribution. *Comp. Stat. Data Anal.* 52:2079-2097.
- [28] Leiva, V., Barros, M., Paula, G.A., Sanhueza, A. (2008) Generalized Birnbaum-Saunders distribution applied to air pollutant concentration. *Environmetrics* 19:235-249.
- [29] Leiva, V., Sanhueza, A., Angulo, J.M. (2008) A length-biased version of the Birnbaum-Saunders distribution with application in water quality. *Stoch. Environ. Res. Risk. Assess.* (in press).
- [30] Leiva, V., Sanhueza, A., Sen, P.K., Paula, G.A. (2008) Random number generators for the generalized Birnbaum-Saunders distribution. *J. Stat. Comp. Simul.* (in press).
- [31] Leiva, V., Sanhueza, A., Saunders, S.C. (2008) New developments and applications on life distributions under cumulative damage. To appear in *Reliability Engineering Advances*.
- [32] Lemonte, A., Cribari-Neto, F., Vasconcellos, K.L.P. (2007) Improved statistical inference for the two-parameter Birnbaum-Saunders distribution. *Comp. Stat. Data Anal.* 51:4656-4681.
- [33] Lemonte, A., Simas, A. Cribari-Neto, F. (2008) Bootstrap-based improved estimators for the two-parameter Birnbaum-Saunders distribution. *J. Stat. Comp. Simul.* 78:37-49.
- [34] Marshall, A.W., Olkin, I. 2007. *Life Distributions*. Springer Verlag, New York.
- [35] Monteiro, J. (1970) *Álgebra Linear* Vol. 2. Nobel. São Paulo.
- [36] Nelson, W. B. (1970b) Hazard Plotting methods for analysis of life data with different failure modes. *J. Qual. Technol.*, 2, 126-149.
- [37] Ng, H. K., Kundub, D., Balakrishnan, N. (2003) Modified moment estimation for the two-parameter Birnbaum-Saunders distribution. *Comp. Stat. Data Anal.* 43:283-98.

- [38] Owen, W. J., PAdgett, W. J. (2000) A Birnbaum-Saunders accelerated life model. *IEEE trans. Rel.* 49: 224-229.
- [39] Podlaski, R. (2008) Characterization of diameter distribution data in near-natural forests using the Birnbaum-Saunders distribution. *Can. J. For. Res.* 38:518-527.
- [40] Poon, W. I., Poon, Y. S. (1999) Conformal normal curvature and assessment of local influence. *J. R. Stat. Soc. B* 61: 51-61.
- [41] Press, W. H., Teulosky, S. A., Vetterling, W. T., Flannery, B. P. (1992) *Numerical Recipes in C: The Art of Scientific Computing*. Prentice Hall, London.
- [42] Rieck, J. R. (1989) Statistical analysis for the Birnbaum-Saunders fatigue life distribution, Unpublished PhD Thesis, Clemson University, Departmental of Matematical Sciences.
- [43] Rieck, J.R., Nedelman, J.R. (1991) A log-Linear model for the Birnbaum-Saunders distribution. *Technometrics* 33:51-60.
- [44] Sanhueza, A., Leiva, V., Balakrishnan, N. (2008) The generalized Birnbaum-Saunders distribution and its theory, methodology and application. *Comm. Stat. Theor. Meth.* 37:645-670.
- [45] Tsionas, E.G. (2001) Bayesian inference in Birnbaum-Saunders regression. *Comm. Stat. Theor. Meth.* 30, 179-193.
- [46] Weisberg, S. (1985) *Applied Linear Regression*, 2nd edn. New York: Springer.
- [47] Xie, F.C., Wei, B.C. (2007) Diagnostics analysis for log-Birnbaum-Saunders regression models. *Comp. Stat. Data Anal.* 51:4692-4706.