



ROCÍO PAOLA MAEHARA SÁNCHEZ

MODELO DE REGRESSÃO LINEAR SINH-NORMAL. APLICAÇÕES  
À TEMPO DE VIDAS

CAMPINAS  
2014





**UNIVERSIDADE ESTADUAL DE CAMPINAS**

**INSTITUTO DE MATEMÁTICA, ESTATÍSTICA  
E COMPUTAÇÃO CIENTÍFICA**

**ROCÍO PAOLA MAEHARA SÁNCHEZ**

**MODELO DE REGRESSÃO LINEAR SINH-NORMAL. APLICAÇÕES  
À TEMPO DE VIDAS**

Dissertação apresentada ao Instituto de  
Matemática, Estatística e Computação Científica da Universidade Estadual  
de Campinas como parte dos requisitos exigidos para  
obtenção do título de Mestra em estatística.

**Orientador: Prof. Dr. Filidor Edilson Vilca Labra**

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO  
DEFENDIDA PELA ALUNA ROCÍO PAOLA MAEHARA SÁNCHEZ,  
E ORIENTADA PELO PROF. DR. FILIDOR EDILFONSO VILCA LABRA.

Assinatura do Orientador

A handwritten signature in black ink is written over a horizontal line. The signature is cursive and appears to read "Filidor Edilson Vilca Labra".

**CAMPINAS  
2014**

Ficha catalográfica  
Universidade Estadual de Campinas  
Biblioteca do Instituto de Matemática, Estatística e Computação Científica  
Maria Fabiana Bezerra Muller - CRB 8/6162

M268m Maehara Sánchez, Rocío Paola, 1983-  
Modelo de regressão linear Sinh-Normal. Aplicações à tempo de vidas / Rocío Paola Maehara Sánchez. – Campinas, SP : [s.n.], 2014.

Orientador: Filidor Edilfonso Vilca Labra.  
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Análise de regressão. 2. Modelos lineares (Estatística). 3. Heteroscedasticidade. I. Labra, Filidor Edilfonso Vilca, 1964-. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Informações para Biblioteca Digital

**Título em outro idioma:** Linear Regression model Sinh-Normal. Applications to life times

**Palavras-chave em inglês:**

Regression analysis

Linear models (Statistics)

Heteroscedasticity

**Área de concentração:** Estatística

**Titulação:** Mestra em Estatística

**Banca examinadora:**

Filidor Edilfonso Vilca Labra [Orientador]

Víctor Hugo Lachos Dávila

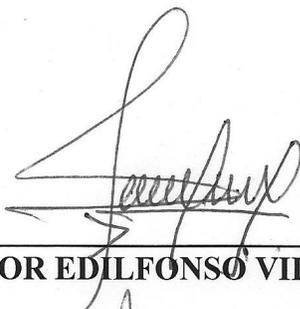
Michelli Karinne Barros da Silva

**Data de defesa:** 07-03-2014

**Programa de Pós-Graduação:** Estatística

**Dissertação de Mestrado defendida em 07 de março de 2014 e aprovada**

**Pela Banca Examinadora composta pelos Profs. Drs.**



---

**Prof(a). Dr(a). FILIDOR EDILFONSO VILCA LABRA**



---

**Prof(a). Dr(a). VÍCTOR HUGO LACHOS DÁVILA**

*Michelli Karinne Barros da Silva*

---

**Prof(a). Dr(a). MICHELLI KARINNE BARROS DA SILVA**



## Abstract

The family of Sinh-normal distributions is a class of symmetric distributions with three parameters, and due to presence of these parameters it is a very flexible distribution. When the Sinh-normal distribution is unimodal, it distribution could be used in place of the normal distribution and consequently in regression model. A subclass de distribution of Sinh-normal distributions is the log-transformation of the Birnbaum-Saunders fatigue-time distribution. So, several properties of the Birnbaum-Saunders distribution and some generalization can be obtained.

The main objective of work is to study some aspect of estimation and analysis of diagnostics in the Sinh-Normal regression model. The analysis of diagnostics is based on the Cook (1986) approach. Two data analysis is performed to see how the proposed model can be used in practice.

Furthermore, we investigate a test of homogeneity for shape parameters in the Sin-Normal regression model. We obtain the score statistics for such test. Finally, a numerical example is given to illustrate our methodology and the properties of the score statistics is investigated through Monte Carlo simulations.

## Resumo

A família de distribuições Sinh-Normal é uma classe de distribuições simétricas com três parâmetros, e devido à presença destes parâmetros esta família é flexível. Quando a distribuição Sinh-Normal é unimodal, esta distribuição pode ser utilizada em lugar da distribuição normal, e conseqüentemente nos modelos de regressão. Uma subclasse das distribuições é o log-transformação da distribuição de tempo de fadiga Birnbaum-Saunders. Assim, várias propriedades da distribuição Birnbaum-Saunders e algumas generalizações podem ser obtidas.

O principal objetivo deste trabalho é estudar alguns aspectos de estimação e análise de diagnóstico no modelo de regressão Sinh-Normal. A análise de diagnóstico baseia-se na metodologia de Cook (1986). Duas análises de dados são realizadas para ver como o modelo proposto pode ser utilizado na prática. Além disso, investigamos um teste de homogeneidade dos parâmetros de forma no modelo de regressão Sinh-Normal. Obtemos as estatísticas de escore para este teste. Finalmente, um exemplo numérico é apresentado para ilustrar a metodologia e as propriedades das estatísticas escore são investigadas através de simulações de Monte Carlo.



# Sumário

<b>Lista de Figuras</b>	<b>xix</b>
<b>Lista de Tabelas</b>	<b>xxi</b>
<b>Abreviações</b>	<b>xxiii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 A distribuição Sinh-Normal . . . . .	1
1.2 Modelos de regressão log-Birnbaum-Saunders . . . . .	3
1.3 Modelos de regressão heteroscedástico . . . . .	3
1.4 Proposta do trabalho e objetivos . . . . .	4
<b>2 A distribuição Sinh-Normal</b>	<b>7</b>
2.1 Introdução . . . . .	7
2.1.1 Distribuições relacionadas . . . . .	13
2.2 Momentos . . . . .	17
<b>3 Modelo de regressão Sinh-Normal</b>	<b>21</b>
3.1 Introdução . . . . .	21
3.2 Descrição do modelo . . . . .	21
3.3 Matriz de Informação Observada . . . . .	22
3.3.1 Derivadas parciais de primeira ordem . . . . .	23
3.3.2 Derivadas parciais de segunda ordem . . . . .	23
3.4 Estimação de parâmetros . . . . .	24
3.4.1 Seleção de modelos . . . . .	25
3.4.2 Análise de resíduos . . . . .	25
3.5 Análise de Diagnóstico . . . . .	25
3.5.1 Modelo de eliminação de casos . . . . .	26
3.5.2 Influência Local . . . . .	28
3.5.3 Esquemas de perturbação . . . . .	32
3.6 Estudo de simulação . . . . .	34
3.7 Aplicação 1: Fadiga biaxial . . . . .	36
3.7.1 Ajuste do modelo . . . . .	36
3.7.2 Análise de diagnóstico . . . . .	43

3.8	Aplicação 2: Produção de cebola . . . . .	52
3.8.1	Ajuste do modelo . . . . .	52
3.8.2	Análise de diagnóstico . . . . .	57
<b>4</b>	<b>Modelos lineares heteroscedásticos Sinh-Normal</b>	<b>65</b>
4.1	Introdução . . . . .	65
4.2	O modelo de regressão heteroscedástico . . . . .	65
4.2.1	Teste escore para homogeneidade do parâmetro de forma . . . . .	67
4.2.2	Teste escore para a homogeneidade do parâmetro de escala . . . . .	69
4.3	Estudo de simulação . . . . .	72
4.4	Aplicação 1: Pesquisa salarial . . . . .	79
4.4.1	Teste de heteroscedasticidade . . . . .	79
4.5	Aplicação 2: Fadiga biaxial . . . . .	84
<b>5</b>	<b>Considerações Finais</b>	<b>89</b>
<b>A</b>	<b>Programas em R</b>	<b>91</b>
A.1	Distribuição Sinh-Normal . . . . .	91
A.2	Distribuição Birbaum-Saunders generalizada . . . . .	91
<b>I</b>	<b>Conjuntos de dados</b>	<b>93</b>
<b>II</b>	<b>Licença</b>	<b>97</b>
II.1	Sobre a licença dessa obra . . . . .	97
	<b>Referência Bibliográfica</b>	<b>103</b>

*A Jesus, meu Senhor e Salvador. . . .*



E temos, mui firme, a palavra dos profetas, à qual bem fazeis em estar atentos, como a uma luz que alumia em lugar escuro, até que o dia amanheça, e a estrela da alva apareça em vossos corações. (2 Pedro 1:19)



# Agradecimentos

- A Jesus, meu Senhor e Salvador, pela sua presença constante na minha vida, pelo auxílio nas minhas escolhas e me confortar nas horas difíceis.
- Ao meu esposo Luis, pelo amor e paciência nos meus “maus” momentos. Amo você demais! Graças a sua presença foi mais fácil transpor os dias de desânimo e cansaço!
- Aos meus pais, Víctor e Doris, por todo amor, carinho e apoio incondicionais, Amo muito vocês!
- Ao meu irmão Yoshitomi, pelo apoio e compreensão.
- Ao meu orientador, Professor Filidor Edilfonso Vilca Labra, pelos importantes ensinamentos, pela paciência pela amizade e apoio, e pelo conforto nas horas difíceis.
- Ao Professor Victor Hugo Lachos Dávila, pelo apoio nos dias de desânimo.
- Ao Reginaldo Ramos Mendes, pela sua amável colaboração neste trabalho.
- A CAPES, pelo apoio financeiro para a execução desse projeto.



# Lista de Figuras

1.1	Comparação da função de densidade de probabilidade da distribuição Sinh-Normal e Normal. . . . .	2
2.1	Função de densidade de probabilidade para a distribuição SN (a) $0 < \alpha \leq 2$ e $\sigma = 1$ (b) $\alpha > 2$ e $\sigma = 1$ (c) $0 < \alpha \leq 2$ e $\sigma = 2$ (d) $\alpha > 2$ e $\sigma = 2$ . . . . .	11
2.2	Função de densidade de probabilidade para a distribuição SN (a) $0 < \alpha \leq 2$ e $\sigma = 3$ (b) $\alpha > 2$ e $\sigma = 3$ (c) $0 < \alpha \leq 2$ e $\sigma = 4$ (d) $\alpha > 2$ e $\sigma = 4$ . . . . .	12
2.3	Densidade de probabilidade da distribuição BSG para $\alpha < 2$ (a) $\sigma = 1$ (b) $\sigma = 2$ (c) $\sigma = 3$ (d) $\sigma = 4$ . . . . .	15
2.4	Função de densidade de probabilidade para a distribuição Birnbaum Saunders (a) $\alpha = 0,2$ e $\sigma \geq 2$ (b) $\alpha = 0,2$ e $\sigma \leq 2$ (c) $\alpha = 0,8$ e $\sigma \geq 2$ (d) $\alpha = 0,8$ e $\sigma \leq 2$ . . . . .	16
2.5	Comparação das funções de densidade de probabilidade das distribuições Normal e Sinh-Normal. . . . .	20
3.1	Representação gráfica do enfoque de influência local. . . . .	29
3.2	Gráficos para a perturbação na variável explicativa. . . . .	35
3.3	Diagrama de dispersão entre $\log(trabalho)$ e $\log(ciclos)$ com o valor médio previsto do $\log(ciclos)$ para conjunto de dados de fadiga. À esquerda usando preditor linear e ligação identidade e à direita preditor linear e ligação recíproca. . . . .	37
3.4	Diagrama de dispersão entre $\log(trabalho)$ e $\log(ciclos)$ com o valor médio previsto do $\log(ciclos)$ para conjunto de dados de fadiga. À esquerda usando preditor linear e ligação logarítmica e à direita preditor quadrático e ligação identidade. . . . .	38
3.5	Diagrama de dispersão entre $\log(trabalho)$ e $\log(ciclos)$ com o valor médio previsto do $\log(ciclos)$ para conjunto de dados de fadiga. À esquerda usando preditor quadrático e ligação recíproca e à direita preditor quadrático e ligação logarítmica. . . . .	38
3.6	Banda de Confiança de 95% para o valor da média prevista de $\log(ciclos)$ para conjunto de dados de fadiga. À esquerda usando preditor linear e ligação identidade e à direita preditor linear e ligação recíproca. . . . .	39
3.7	Banda de Confiança de 95% para o valor da média prevista de $\log(ciclos)$ para conjunto de dados de fadiga. À esquerda usando preditor linear e ligação logarítmica e à direita preditor quadrático e ligação identidade. . . . .	40

3.8	Banda de Confiança de 95% para o valor da média prevista de $\log(\text{ciclos})$ para conjunto de dados de fadiga. À esquerda usando quadrático e ligação recíproca e à direita preditor quadrático e ligação logarítmica. . . . .	40
3.9	Gráfico de probabilidade normal com envelope simulado para o ajuste do modelo com preditor linear e ligação logarítmica para os dados de fadiga. À esquerda usando resíduo de Pearson e à direita usando outro resíduo. . . . .	42
3.10	Envelope simulado para os dados de fadiga. . . . .	42
3.11	Gráfico de resíduos para os dados de fadiga. À esquerda usando resíduo de Pearson e à direita usando outro resíduo. . . . .	43
3.12	Distância Generalizada de Cook para os parâmetros $\alpha$ e $\beta$ para o conjunto de dados de fadiga. . . . .	44
3.13	Distância Generalizada de Cook para o parâmetro $\sigma$ e para $\theta$ para o conjunto de dados de fadiga. . . . .	44
3.14	Distância de verossimilhança para conjunto de dados de fadiga. . . . .	45
3.15	Gráfico $M(0)$ sob perturbação de ponderação de casos para conjunto de dados de fadiga. . . . .	46
3.16	Ponderação de casos para conjunto de dados de fadiga. À esquerda gráfico de influência total $C_i$ vs observações e à direita gráfico das médias de influência local $C_i(\alpha)$ vs $C_i(\beta)$ . . . . .	47
3.17	Gráfico $M(0)$ sob perturbação na variável explicativa para conjunto de dados de fadiga. . . . .	48
3.18	Perturbação na variável explicativa para conjunto de dados de fadiga. Na esquerda gráfico de influência total $C_i$ vs observações; na direita gráfico das medias de influência local $C_i(\alpha)$ vs $C_i(\beta)$ . . . . .	49
3.19	Gráfico $M(0)$ sob perturbação na variável resposta para conjunto de dados de fadiga. . . . .	50
3.20	Perturbação na variável resposta. À esquerda gráfico de influência total $C_i$ vs Observações; à direita gráfico das médias de influência local $C_i(\alpha)$ vs $C_i(\beta)$ . . . . .	51
3.21	Diagrama de dispersão entre $\log(\text{produção de cebola})$ e a densidade da planta com o valor médio previsto do $\log(\text{produção de cebola})$ para os dados da produção de cebola (a) $(\mathbf{x}_i^\top \beta)^{-1}$ e (b) $-\log(\mathbf{x}_i^\top \beta)$ . . . . .	53
3.22	Banda de confiança de 95% para o valor da média prevista de $\log(\text{produção de cebola})$ para os dados da produção de cebola (a) $(\mathbf{x}_i^\top \beta)^{-1}$ e (b) $-\log(\mathbf{x}_i^\top \beta)$ . . . . .	54
3.23	Gráfico de probabilidade normal com envelope simulado para os dados da produção de cebola para $-\log(\mathbf{x}_i^\top \beta)$ (a) Resíduo de Pearson (b) Outro resíduo. . . . .	55
3.24	Envelope simulado para os dados da produção de cebola para $-\log(\mathbf{x}_i^\top \beta)$ . . . . .	56
3.25	Gráfico de resíduos para os dados da produção de cebola para $-\log(\mathbf{x}_i^\top \beta)$ (a) Resíduo de Pearson (b) Outro resíduo. . . . .	56
3.26	Distância Generalizada de Cook para os parâmetros $\alpha$ e $\beta$ para os dados da produção de cebola. . . . .	57
3.27	Distância Generalizada de Cook para os diferentes para o parâmetro $\sigma$ e $\theta$ para os dados da produção de cebola. . . . .	58
3.28	Distância de verossimilhança para os dados da produção de cebola. . . . .	58

3.29	Gráfico $M(0)$ sob perturbação de ponderação de casos para os dados da produção de cebola. . . . .	59
3.30	Gráfico $M(0)$ sob perturbação na variável explicativa para os dados da produção de cebola. . . . .	60
3.31	Gráfico $M(0)$ sob perturbação na variável resposta para os dados da produção de cebola. . . . .	61
3.32	Ponderação de casos para os dados da produção de cebola. À esquerda gráfico de influência total $C_i$ vs observações e à direita gráfico das médias de influência local $C_i(\boldsymbol{\alpha})$ vs $C_i(\boldsymbol{\beta})$ . . . . .	62
3.33	Perturbação na variável explicativa. À esquerda gráfico de influência total $C_i$ vs observações e à direita Gráfico das médias de influência local $C_i(\boldsymbol{\alpha})$ vs $C_i(\boldsymbol{\beta})$ . . . . .	63
3.34	Perturbação na variável resposta para os dados da produção de cebola. À esquerda gráfico de influência total $C_i$ vs observações e à direita gráfico das médias de influência local $C_i(\boldsymbol{\alpha})$ vs $C_i(\boldsymbol{\beta})$ . . . . .	64
4.1	Distribuição empírica e teórica da estatística escore, para testar $H_0$ para o caso (a) sobre a função potência, para diferente tamanhos de amostra. . . . .	75
4.2	Distribuição empírica e teórica da estatística escore, para testar $H_0$ para o caso (a) sobre a função exponencial, para diferente tamanhos de amostra. . . . .	76
4.3	Distribuição empírica e teórica da estatística escore, para testar $H_0$ para o caso (b) sobre a função potência, para diferente tamanhos de amostra. . . . .	77
4.4	Distribuição empírica e teórica da estatística de escore, para testar $H_0$ para o caso (b) sobre a função exponencial, para diferente tamanhos de amostra. . . . .	78
4.5	Gráfico de probabilidade normal com envelope simulado para os dados da pesquisa salarial. Usando em (a) $\alpha_i = \alpha \exp(x_i\lambda)$ , (b) $\alpha_i = \alpha x_i^\lambda$ , (c) $\sigma_i = \sigma \exp(x_i\lambda)$ e (d) $\sigma_i = \sigma x_i^\lambda$ . . . . .	81
4.6	Gráfico de resíduos para os dados da pesquisa salarial usando o resíduo de Pearson. Usando em. (a) $\alpha_i = \alpha \exp(x_i\lambda)$ , (b) $\alpha_i = \alpha x_i^\lambda$ , (c) $\sigma_i = \sigma \exp(x_i\lambda)$ e (d) $\sigma_i = \sigma x_i^\lambda$ . . . . .	82
4.7	Envelope simulado para os dados da pesquisa salarial. Usando em (a) $\alpha_i = \alpha \exp(x_i\lambda)$ , (b) $\alpha_i = \alpha x_i^\lambda$ , (c) $\sigma_i = \sigma \exp(x_i\lambda)$ e (d) $\sigma_i = \sigma x_i^\lambda$ . . . . .	83
4.8	Gráfico de probabilidade normal com envelope simulado para os dados da fadiga biaxial sob modelo homocedástico. . . . .	85
4.9	Gráfico de resíduos de Pearson para os dados da fadiga biaxial sob modelo homocedástico. . . . .	86
4.10	Envelope simulado para os dados da fadiga biaxial sob modelo homocedástico. . . . .	86



# Lista de Tabelas

3.1	Estimadores de máxima verossimilhança com seus respectivos erros padrões. . . . .	34
3.2	Estimativas de máxima verossimilhança para a distribuição SN, com seus respectivos erros padrões, para os dados de fadiga biaxial. . . . .	36
3.3	Critérios para seleção de modelos para o conjunto de dados da fadiga biaxial. . . . .	37
3.4	Mudança nas estimativas dos parâmetros (em %), excluindo as observações potencialmente influentes. . . . .	52
3.5	Estimativas de máxima verossimilhança para a distribuição Sinh-Normal, com seus respectivos erros padrão, para os dados da produção de cebola. . . . .	52
3.6	Critérios para seleção de modelos para os dados da produção de cebola. . . . .	53
4.1	Nível de significância e poderes, para o caso (a) sob a função de potência (sem parênteses) e exponencial (com parênteses). . . . .	73
4.2	Tamanhos simulados e poderes para o caso (b) em função de potência (sem parênteses) e a função exponencial (com parênteses). . . . .	74
4.3	Estatística de escore (SC) e (p-valor) para os modelos heteroscedásticos dos parâmetros de forma e escala para dados da pesquisa salarial. . . . .	80
4.4	As estimativas de máxima verossimilhança (erros padrão aproximados) para os modelos heteroscedásticos dos parâmetros de forma e escala para dados da pesquisa salarial. . . . .	80
4.5	AIC e BIC para os testes heteroscedasticidade dos parâmetros de forma e escala para dados da pesquisa salarial. . . . .	80
4.6	As estimativas de máxima verossimilhança (erros padrão aproximados) para os modelos heteroscedásticos dos parâmetros de forma e escala para dados da fadiga biaxial. . . . .	84
4.7	Estatística de escore (E), p-valor, AIC e BIC para os testes heteroscedasticidade dos parâmetros de forma e escala para dados da fadiga biaxial. . . . .	85
I.1	Fadiga biaxial . . . . .	93
I.2	Produção de cebola . . . . .	94
I.3	Pesquisa salarial . . . . .	95



# Abreviações

**BS** Birnbaum-Saunders

**EMV** Estimativas de máxima verossimilhança

**FDA** Função de distribuição acumulada

**fdp** Função de densidade de probabilidade

**log-BS** Distribuição log-Birnbaum Saunders

**RSN** Modelo de regressão Sinh-Normal

**SQE** Soma dos quadrados dos resíduos (erros) [SN] Sinh-Normal



# Capítulo 1

## Introdução

### 1.1 A distribuição Sinh-Normal

A distribuição normal tem desempenhado um papel dominante na teoria e estatística aplicada. No entanto, é evidente que a curva normal não fornece uma representação adequada em muitas situações na prática. Patel & Read (1996) apontou "a distribuição normal desempenhou um papel fundamental na análise estatística ao longo dos anos, mas desde os anos 1960 consideráveis trabalhos tem chamado a atenção questionando esses pressupostos". De fato, Johnson (1949) apresentou uma discussão das transformadas de distribuições normais, e este trabalho deu espaço a muitas distribuições que são transformações da distribuição normal, como por exemplo a distribuição simétrica Sinh-Normal proposto por Rieck (1989), a distribuição Birnbaum-Saunders proposto por Birnbaum & Saunders (1969a) e Birnbaum & Saunders (1969b). Alguns resultados complementares podem ser encontrados num recente trabalho de Athayde *et al.* (2012), Leiva *et al.* (2010).

Em situações práticas, especialmente durante as análises descritivas dos dados existem grandes dificuldades na aplicação dos resultados com base no pressuposto de uma distribuição normal, porque em muitas situações os dados não provêm de uma distribuição normal. É usual por muitos pesquisadores propor uma transformação nos dados, com intuito de obter um melhor ajuste dos dados, mas este procedimento pode levar a conclusões errôneas ou as conclusões resultantes podem ser de difícil interpretação. Para alguns determinados conjunto de dados, o sistema de Johnson, originalmente apresentado por Johnson (1949) fornece uma possível distribuição cobrindo muitas combinações possíveis de assimetria e curtoses.

Johnson (1949) mencionou que é natural e também conveniente construir distribuições não normais, transformando uma variável aleatória que segue uma distribuição normal. Ele usou o método de transformação para gerar distribuições estatísticas que assumem uma ampla variedade de formas, a transformação

$$Z = \nu + \delta g\left(\frac{Y - \mu}{\sigma}\right), \quad (1.1.1)$$

em que  $Z \sim N(0, 1)$  e  $g(\cdot)$  é uma função monótona. Estas distribuições não normais possuem

quatro parâmetros  $\nu$ ,  $\delta$ ,  $\mu$  e  $\sigma$ , em que  $\nu$  e  $\delta$  correspondem aos parâmetros de forma e  $\mu$  e  $\sigma$  são os parâmetros de localização e de escala, respectivamente. É possível supor que  $\delta$  e  $\sigma$  são positivos. Baseado em (1.1.1) com  $\delta = 2/\alpha$  e  $g = \sinh(\cdot)$ , Rieck (1989) definiu o seguinte modelo

$$Z = \nu + \frac{2}{\alpha} \sinh\left(\frac{Y - \mu}{\sigma}\right) \sim N(0, 1). \quad (1.1.2)$$

Neste caso é dito que  $Y$  segue uma distribuição Sinh-Normal (SN) de quatro parâmetros. A notação  $Y \sim SN(\alpha, \mu, \sigma, \nu)$  é utilizada, e é reduzida a  $Y \sim SN(\alpha, \mu, \sigma)$ , quando  $\nu = 0$  em (1.1.2), para mais detalhes e aplicações da distribuição SN, consulte Rieck (1989) e Rieck & Nedelman (1991).

Na Figura 1.1 apresentamos os gráficos das funções de densidade das distribuições Normal e Sinh-Normal. Observe que, as distribuições apresentam um comportamento similar.

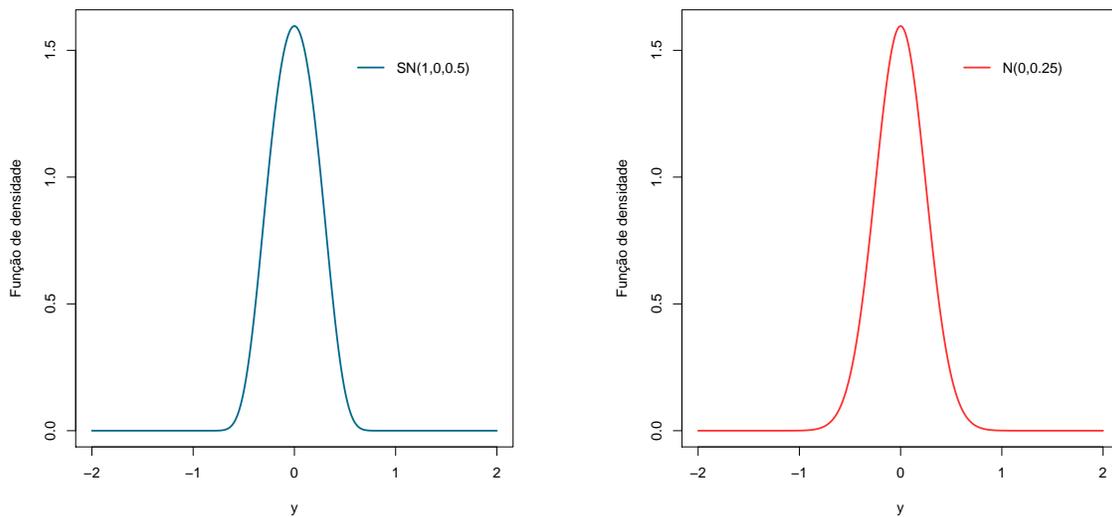


Figura 1.1: Comparação da função de densidade de probabilidade da distribuição Sinh-Normal e Normal.

Devido a sua aplicação e estreita relação com a distribuição normal, a distribuição SN tem sido objeto de estudos por muitos autores. As extensões da distribuição SN, consideradas na literatura, acompanha as extensões que a distribuição normal tem sofrido, Leiva *et al.* (2010) em lugar da distribuição normal, eles consideraram a distribuição skew-elíptica, Vilca *et al.* (2010) consideraram também extensões baseado em distribuições com aplicações em estudo ambiental.

Se  $Y \sim SN(\alpha, \mu, \sigma = 2)$ , então  $T = \exp(Y)$  segue uma distribuição Birnbaum-Saunders (BS), com parâmetro de forma  $\alpha > 0$  e parâmetro de escala  $\beta = \exp(\mu) > 0$ . A notação  $T \sim BS(\alpha, \beta)$  é usada neste caso; veja Birnbaum & Saunders (1969a), Jonhson *et al.* (1995), e Sanhueza *et al.*

(2008). Por este motivo, o modelo SN com  $\sigma = 2$  é também chamado de distribuição log-Birnbaum-Saunders (log-BS); veja Rieck & Nedelman (1991), Galea *et al.* (2004) e Leiva *et al.* (2007). Para a distribuição BS, Birnbaum & Saunders (1969b) discutiram as estimativas de máxima verossimilhança (EMV) dos parâmetros; Achar (1993) desenvolveram alguns métodos bayesianos aproximados com base no método de Laplace para integrais; Ng *et al.* (2003) obtiveram estimadores através do método de momentos modificado e Wang *et al.* (2006) exploraram uma estimativa de momentos modificado censurado.

Sobre distribuição BS, esta distribuição surgiu no contexto de engenharia de materiais, mas a sua aplicabilidade tem sido estendida para outros campos da engenharia. Além disso, devido aos argumentos teóricos no processo de construção da distribuição BS, que permite descrever processos de degradação acumulativa, tem transformado a distribuição BS em um modelo que pode ser aplicado em outras áreas, como por exemplo, em ciências da saúde, na área ambiental e florestal, em demografia, na área atuarial e financeira, entre outras. Todos estes aspectos mencionados têm permitido considerar a distribuição BS como um modelo de probabilidade, em vez de um modelo considerado apenas para descrever tempos de vida.

## 1.2 Modelos de regressão log-Birnbaum-Saunders

A distribuição da log-BS corresponde à uma distribuição SN. Esta relação tem permitido considerar modelos de regressão quando a resposta envolve logaritmos dos tempos de vida. Esta ideia foi proposta por Rieck & Nedelman (1991), e posteriormente muitos trabalhos tem sido discutidos na literatura. Para os modelos de regressão log-BS, Rieck & Nedelman (1991) investigaram as EMV e estimadores de mínimos quadrados dos parâmetros; Tisonas (2001) discutiu a estimadores Bayesianos para o modelo.

Galea *et al.* (2004) apresentaram métodos de diagnóstico para modelos de regressão log-BS. A parte central dos seus resultados baseia-se na metodologia de influência local proposto por Cook (1986). A teoria desenvolvida sob influencia local é baseada no trabalho de Lachos (2002). É sabido que, para avaliar a influência da  $i$ -ésima observação nas estimativas de parâmetros, uma abordagem direta é calcular o diagnóstico de um único caso com o  $i$ -ésimo caso excluído. Desde o trabalho pioneiro de Cook (1977), o modelo de eliminação de casos, como distância de Cook ou o afastamento pela verossimilhança tem sido aplicados com sucesso em diversos modelos estatísticos, ver, por exemplo, Christensen *et al.* (1992), Davison & Tsai (1992), Wei (1998), Tang *et al.* (2000), Galea *et al.* (2005), entre outros.

## 1.3 Modelos de regressão heteroscedástico

A modelagem de dados simétricos é frequentemente baseada na suposição de variância constante para os erros. Contudo, em muitas situações práticas essa suposição é dificilmente verificada.

Então, procura-se uma transformação na variável resposta para estabilizar a variância, e nem sempre é o caminho mais apropriado. Para contornar estes problemas, um parâmetro de dispersão é atribuído para cada observação, sendo relacionado linearmente através de combinações lineares de variáveis explicativas, por meio de uma função de ligação conhecida. Estes modelos são conhecidos na literatura como modelos de regressão heteroscedásticos, que tem sido largamente discutidos na área de econometria, veja por exemplo Park (1966) e Harvey (1976). Alguns procedimentos para verificar a presença heteroscedasticidade podem ser encontrados em Cook & Weisberg (1983) e Atkinson (1985), que apresentam alguns métodos gráficos para detectar heteroscedasticidade, Aitkin (1987) desenvolveu rotinas computacionais no GLIM para a estimação de máxima verossimilhança para modelagem da variância sob erros normais.

A suposição padrão na análise de regressão é a homogeneidade de variâncias do erro. A violação de tal suposição pode ter consequências negativas para a eficiência dos estimadores portanto, é importante para verificar se há heteroscedasticidade sempre que é considerado uma possibilidade (Cysneiros *et al.* (2010)). Este problema tem sido amplamente discutido na literatura estatística. Por exemplo, Barroso & Cordeiro (2005) consideraram correções de Bartlett em modelos de regressão heteroscedásticos com distribuições de *t*-Student; Cysneiros *et al.* (2007) discutiram métodos de diagnóstico em modelos lineares heteroscedásticos com erros simétricos e Cysneiros *et al.* (2010) introduziram a classe de modelos de regressão não-lineares heteroscedásticos simétricos onde um processo iterativo conjunto para estimar a média e os parâmetros de dispersão é proposto.

O teste de homogeneidade do parâmetro de forma em modelos de regressão log-BS. Este problema foi referido por ambos Rieck & Nedelman (1991) e Galea *et al.* (2004). Na verdade, se o parâmetro de forma, não é homogêneo nos modelos de regressão log-BS, digamos  $\alpha_i$  para cada observação  $i$ , então a inferência seria muito difícil lidar com a regressão log-BS.

## 1.4 Proposta do trabalho e objetivos

Devido à estrutura simétrica da distribuição SN é natural pensar que os erros num modelo de regressão possam ser distribuídos de acordo a uma distribuição SN. Um caso especial é o modelo de regressão log-BS, onde é assumido que os erros seguem uma distribuição  $SN(\alpha, 0, 2)$ .

Seguindo a ideia desenvolvida nos modelos de regressão, vamos considerar modelos de regressão onde os erros seguem uma distribuição SN. Especial atenção será dedicado a estimação, análise de diagnóstico e teste de hipóteses.

O principal objetivo deste trabalho é estudar alguns aspectos de estimação e análise de diagnóstico no modelo de regressão Sinh-Normal. O estudo de diagnóstico é desenvolvido seguindo a metodologia de Cook (1986). O teste de homogeneidade dos parâmetros de forma e escala no modelo de regressão Sinh-Normal é baseado na estatística de escore. Os objetivos específicos podem ser resumidos como segue:

- 1** Descrever o modelo de regressão SN, e modelar a média da variável resposta que relaciona um preditor linear, baseado em um conjunto de variáveis explicativas conhecidas e parâmetros desconhecidos, por meio de uma função de ligação. Um caso especial é o modelo de regressão linear proposto por Rieck & Nedelman (1991);
- 2** Realizar um estudo de diagnóstico no modelo de regressão linear SN, seguindo a metodologia Cook (1986) e as ideias de Villegas *et al.* (2013);
- 3** Apresentar um teste de homogeneidade dos parâmetros de forma e escala no modelo de regressão linear SN usando a estatística de score. Finalmente, um exemplo numérico é apresentado para ilustrar a metodologia e as propriedades da estatística score é investigado através de simulações de Monte Carlo.



# Capítulo 2

## A distribuição Sinh-Normal

### 2.1 Introdução

Nem todos os fenômenos aleatórios na natureza podem ser descritos pela distribuição normal. Assim, outras distribuições com comportamento similar são necessárias para modelar o comportamento aleatório de certos fenômenos. Sistemas de curvas de frequência devem ser construídos de modo a representar uma grande variedade de distribuições. Historicamente, era natural e conveniente considerar a construção de distribuições não normais, assumindo que uma transformação da variável aleatória tem uma distribuição normal, uma vez que esta distribuição é bem tabulada. Esta técnica é referida como o método de transformação. Edgeworth (1898) considerou transformações que podem ser representadas por polinômios, Wicksell (1917) e Rietz (1922) consideraram transformações mais gerais, incluindo a transformação logarítmica. No entanto, em comparação com o sistema de curvas de Pearson, as curvas propostas pelos referidos autores, somente cobrem uma variedade limitada de formas.

Johnson (1949) apresentou um método de transformação para gerar três famílias de distribuições que podem assumir grandes variedades de formas. As três famílias de distribuições são geradas por transformações da forma

$$Z = \nu + \delta f(Y; \mu, \sigma), \quad (2.1.1)$$

em que,  $Z$  é uma variável aleatória (v.a.) que segue uma distribuição normal padrão e  $f(\cdot)$  é uma função monótona da v.a.  $Y$ . Existem quatro parâmetros para essas distribuições que são indicados por  $\nu$ ,  $\delta$ ,  $\mu$  e  $\sigma$ , que são os parâmetros de não centralidade, forma, locação e escala respectivamente. A função de distribuição acumulada (FDA) de  $Y$  pode ser escrita por

$$F_Y(y) = F(y; \delta, \mu, \sigma, \nu) = \Phi \left[ \nu + \delta g \left( \frac{y - \mu}{\sigma} \right) \right], \quad y \in R,$$

em que  $\Phi(\cdot)$  é FDA da distribuição  $N(0, 1)$ .

A seguir apresentamos alguns exemplos que tem sido considerados na literatura. As três famílias de distribuições e as funções utilizadas para gerar essas famílias:

1. A distribuição  $S_L$ , com  $f(y; \mu, \sigma) = \log\left(\frac{y - \mu}{\sigma}\right)$ . Neste caso  $g(u) = \log(u)$ .
2. A distribuição  $S_B$ , com  $f(y; \mu, \sigma) = \log\left(\frac{y - \mu}{\sigma + \mu - y}\right)$ . Neste caso  $g(u) = \log\left(\frac{u}{1-u}\right)$ .
3. A distribuição  $S_U$ , com  $f(y; \mu, \sigma) = \operatorname{arcsinh}\left(\frac{y - \mu}{\sigma}\right)$ . Neste caso  $g(u) = \operatorname{arcsinh}(u)$ .

**Lema 2.1.1.** *Seja  $g$  uma função ímpar e  $F(y; \delta, \mu, \sigma, \nu) = \Phi\left[\nu + \delta g\left(\frac{y - \mu}{\sigma}\right)\right]$  a FDA de  $Y$ . Então, a distribuição de  $Y$  é simétrica em torno de  $\mu$  se, e somente se,  $\nu = 0$ .*

**Demonstração.** Primeiramente suponha que  $Y$  é simétrica em torno de  $\mu$  se e somente se a distribuição  $Y - \mu$  é a mesma que a distribuição de  $-(Y - \mu)$ . Seja  $F(y; \delta, \mu, \sigma, \nu) = \Phi\left[\nu + \delta g\left(\frac{y - \mu}{\sigma}\right)\right]$  a FDA para a variável aleatória  $Y$ . Uma vez que a distribuição de  $Y$  é simétrica em torno de  $\mu$ , a distribuição de  $U = Y - \mu$  é igual à distribuição de  $V = -(Y - \mu)$ , que é,  $F_U(x) = F_V(x)$  para todo  $x$ . De acordo com Rieck (1989), vemos que

$$F_U(x; \delta, -\mu + \mu, \sigma, \nu) = F(x; \delta, 0, \sigma, \nu)$$

e

$$F_V(x; \delta, \mu - \mu, \sigma, -\nu) = F(x; \delta, 0, \sigma, -\nu).$$

Como  $F_U(x) = F_V(x)$  para todos os  $x$ , temos que  $\nu = -\nu$ , que só pode ocorrer quando  $\nu = 0$ . Portanto, se a distribuição de  $Y$  é simétrica em torno de  $\mu$ , então  $\nu = 0$ .

Agora, suponha que  $\nu = 0$ . Então é claro que  $F_U(x) = F_V(x)$ . Portanto, se  $\nu = 0$ , então a distribuição de  $Y$  é simétrica em torno de  $\mu$ . ■

No decorrer deste trabalho concentraremos nossa atenção na família de distribuições que são gerados pela transformação dada por

$$Z = \nu + \frac{2}{\alpha} \sinh\left(\frac{Y - \mu}{\sigma}\right),$$

em que  $\delta$  em (2.1.1) é substituída por  $2/\alpha$ . Este reparametrização é conveniente para relacionar essa distribuição a outras distribuições. Será mostrado, mais adiante que a log transformação da distribuição de tempo de vida de fadiga BS, é um membro particular da família de distribuições definidas em (2.1.1), com  $\nu = 0$ .

A seguir apresentamos, formalmente, a classe de distribuições descritas acima, que foi discutida no trabalho de Rieck (1989). Seja  $Y$  uma v.a. satisfazendo a relação

$$Z = \nu + \frac{2}{\alpha} \sinh\left(\frac{Y - \mu}{\sigma}\right) \sim N(0, 1). \quad (2.1.2)$$

Então, dizemos que  $Y$  tem distribuição SN não central, em que  $\nu$  é o parâmetro de não-centralidade. Os parâmetros  $\mu$  e  $\sigma$  são os parâmetros de locação e escala, respectivamente, e os parâmetros  $\alpha$  e  $\nu$  são parâmetros de forma. De acordo a Johnson (1949),  $\nu$  afeta principalmente assimetria e  $\alpha$  afeta principalmente curtose. Se  $\nu = 0$ , então dizemos que  $Y$  tem uma distribuição SN central.

A distribuição de  $Y$  está relacionada com o modelo normal através da representação

$$Y = \mu + \sigma \operatorname{arcsinh}\left(\frac{\alpha(Z - \nu)}{2}\right), \quad (2.1.3)$$

em que  $Z \sim N(0, 1)$ , que é obtida de (2.1.2).

No decorrer deste trabalho, a distribuição SN não central de  $Y$  será denotada por  $Y \sim SN(\alpha, \mu, \sigma, \nu)$ . Por outro lado, quando  $\nu = 0$ , usamos a notação  $Y \sim SN(\alpha, \mu, \sigma)$ , que corresponde à distribuição SN central, ou simplesmente, distribuição SN.

Para  $Y$  com distribuição SN, a FDA é dada por

$$F(y) = \Phi\left[\frac{2}{\alpha} \sinh\left(\frac{y - \mu}{\sigma}\right)\right], \quad \alpha > 0, \sigma > 0, \mu \in \mathbb{R}, y \in \mathbb{R}, \quad (2.1.4)$$

em que  $\Phi(\cdot)$  é FDA da normal padrão. A função de densidade de probabilidade (fdp) de  $Y$  é dada por

$$\begin{aligned} f_Y(y) &= \frac{2}{\alpha\sigma\sqrt{2\pi}} \cosh\left(\frac{y - \mu}{\sigma}\right) \exp\left[-\frac{2}{\alpha^2} \sinh^2\left(\frac{y - \mu}{\sigma}\right)\right], \\ &= \sigma^{-1} \phi(\xi_{2y}) \xi_{1y}, \quad y \in \mathbb{R}, \end{aligned} \quad (2.1.5)$$

em que  $\xi_{1y} = \xi_1(\alpha, \beta, \sigma) = \frac{2}{\alpha} \cosh\left(\frac{y - \mu}{\sigma}\right)$ ,  $\xi_{2y} = \xi_2(\alpha, \beta, \sigma) = \frac{2}{\alpha} \sinh\left(\frac{y - \mu}{\sigma}\right)$  e  $\phi(\cdot)$  denota a fdp da distribuição normal padrão.

**Lema 2.1.2.** *A distribuição SN é fortemente unimodal se, e somente se,  $\alpha \leq 2$ .*

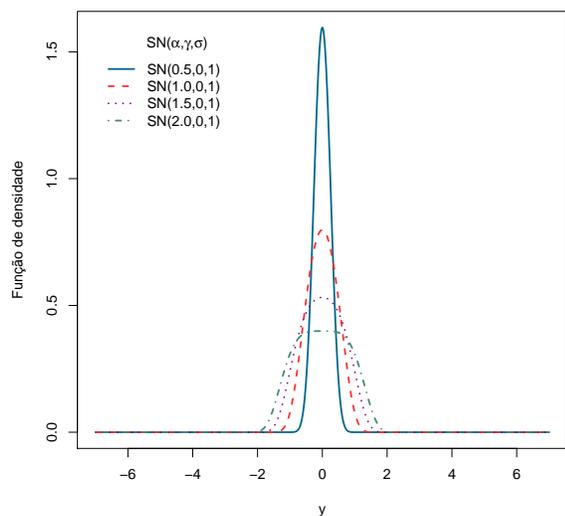
**Demonstração.** A distribuição é fortemente unimodal se  $g(y) = \log[f(y)]$  é côncava. Para mostrar que  $g(y)$  é côncava para  $\alpha \leq 2$ , vamos mostrar que  $g''(y) \leq 0$  para todo  $y$ . A segunda derivada de  $g(y)$  é

$$g''(y) = \frac{\operatorname{sech}^2(x) \left[1 - \frac{4}{\alpha^2} \cosh^2(x)\right] - \frac{8}{\alpha^2} \sinh^2(x)}{\sigma^2},$$

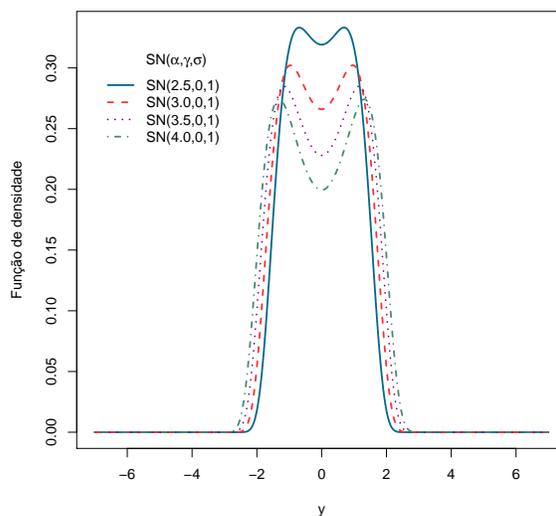
em que  $x = \frac{y - \mu}{\sigma}$ . Para  $\alpha \leq 2$ , ambos termos são menores que ou iguais a zero para todos os valores de  $y$ . Portanto,  $g''(y) \leq 0$  para todos os  $y$  e portanto a distribuição é fortemente unimodal. Se  $\alpha > 2$  em seguida, quando  $x = 0$ , então o primeiro termo é positivo e o segundo termo é zero. Portanto,  $g''(y)$  não é côncava e a distribuição não é fortemente unimodal. ■

Se  $Y \sim SN(\alpha, \mu, \sigma = 2)$ , então  $T = e^Y$  segue a distribuição BS com parâmetro de forma  $\alpha > 0$  e parâmetro de escala  $\beta = e^\mu > 0$ . A notação  $T \sim BS(\alpha, \beta)$  é utilizada neste caso. Por esta razão, o modelo SN é também chamado de distribuição log-BS. Como apontado por Rieck & Nedelman (1991), a função densidade de SN é simétrica em torno do parâmetro de locação  $\mu$ , fortemente unimodal para  $\alpha < 2$ , bimodal para  $\alpha > 2$  e platicúrtica para  $\alpha = 2$ .

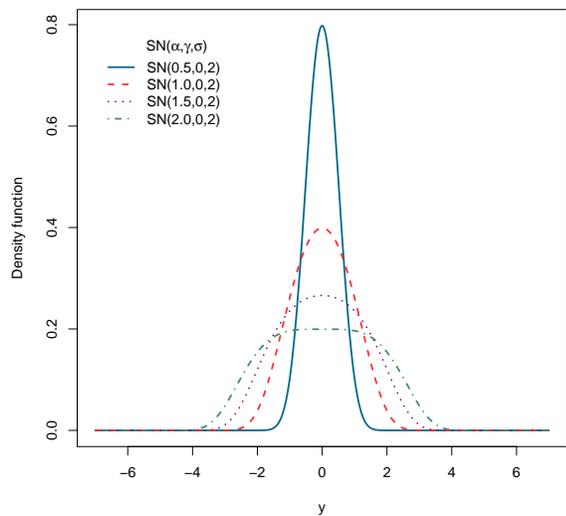
Nas Figuras 2.1 e 2.2, são apresentados gráficos da fdp da distribuição SN para diferentes valores de  $\alpha$ , considerando  $\sigma = 1$ ,  $\sigma = 2$  e  $\sigma = 3$ . Observe que, para valores de  $\alpha \leq 2$  a distribuição é fortemente unimodal e para valores de  $\alpha > 2$  é bimodal. Note-se que a medida que  $\sigma$  cresce a altura da distribuição diminui e sua amplitude tende a aumentar.



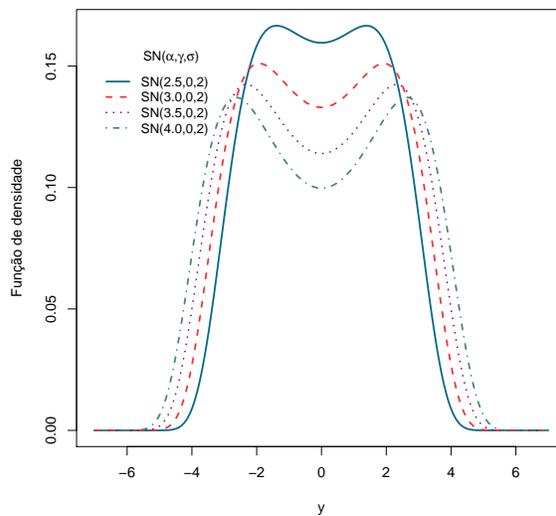
(a)



(b)

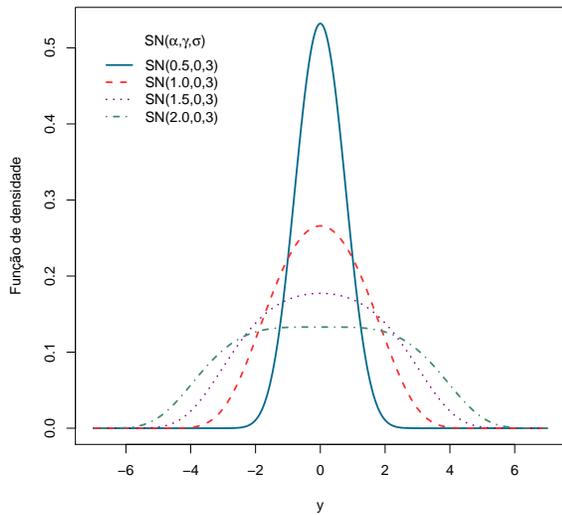


(c)

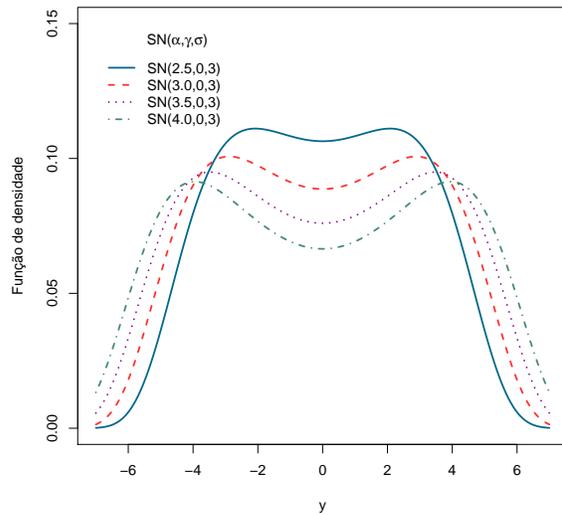


(d)

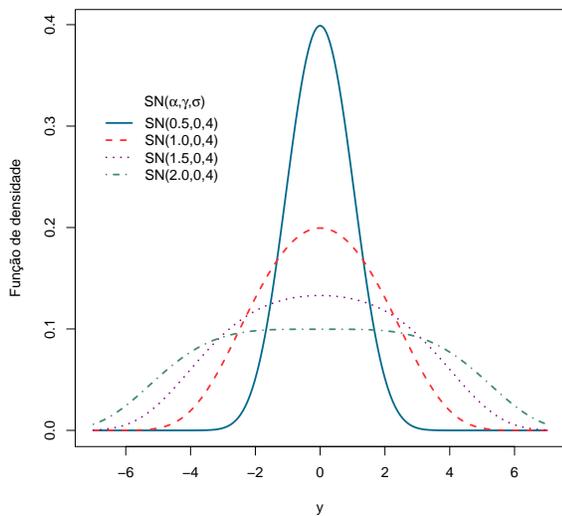
Figura 2.1: Função de densidade de probabilidade para a distribuição SN (a)  $0 < \alpha \leq 2$  e  $\sigma = 1$  (b)  $\alpha > 2$  e  $\sigma = 1$  (c)  $0 < \alpha \leq 2$  e  $\sigma = 2$  (d)  $\alpha > 2$  e  $\sigma = 2$ .



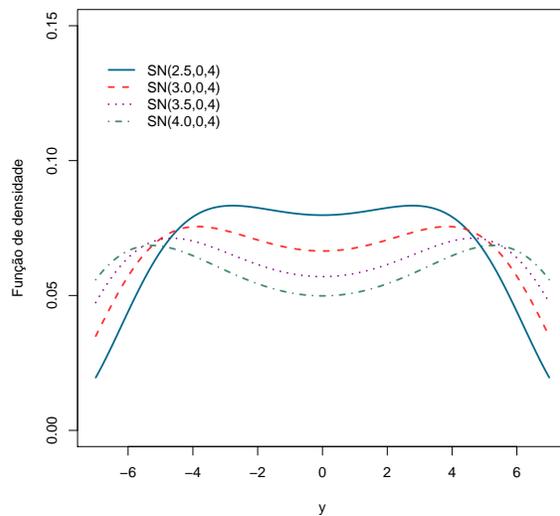
(a)



(b)



(c)



(d)

Figura 2.2: Função de densidade de probabilidade para a distribuição SN (a)  $0 < \alpha \leq 2$  e  $\sigma = 3$  (b)  $\alpha > 2$  e  $\sigma = 3$  (c)  $0 < \alpha \leq 2$  e  $\sigma = 4$  (d)  $\alpha > 2$  e  $\sigma = 4$ .

### 2.1.1 Distribuições relacionadas

Nesta subseção, apresentamos algumas propriedades da distribuição SN e a relação com a distribuição BS.

- 1) Seja  $Y \sim SN(\alpha, \mu, \sigma)$ . Então  $a + bY \sim SN(\alpha, a + b\mu, |b|\sigma)$ , para todo  $a, b \in \mathbb{R}$ ;
- 2) Seja  $Y_\alpha \sim SN(\alpha, \mu, \sigma)$  e  $S_\alpha = \frac{Y_\alpha - \mu}{0.5\alpha\sigma}$ . Então  $\lim S_\alpha \rightarrow Z$ , em que  $Z \sim N(0, 1)$ ;
- 3) A distribuição SN é unimodal para  $\alpha \leq 2$  e bimodal para  $\alpha > 2$ .

**Teorema 2.1.3.** *Seja  $Y \sim SN(\alpha, \mu, \sigma)$ . Então, a fdp de  $T = \exp(Y)$  é dada por*

$$f_T(t) = \phi(a_t)A_t, \quad t > 0, \quad (2.1.6)$$

em que

$$a_t = \left(\frac{1}{\alpha}\right) \left[ \left(\frac{t}{\beta}\right)^{\frac{1}{\sigma}} - \left(\frac{\beta}{t}\right)^{\frac{1}{\sigma}} \right] \quad e \quad A_t = \frac{t^{-(\frac{1}{\sigma}+1)}}{\sigma\alpha\beta^{\frac{1}{\sigma}}} \left(t^{\frac{2}{\sigma}} + \beta^{\frac{2}{\sigma}}\right),$$

com  $\beta = \exp(\mu)$ .

**Demonstração.** Se  $Y \sim SN(\alpha, \mu, \sigma)$ , então a FDA é como em (2.1.4). Assim

$$\begin{aligned} P(T \leq t) &= P(\exp(Y) \leq t) \\ &= P(Y \leq \log(t)) \\ &= \Phi \left[ \frac{2}{\alpha} \sinh \left( \frac{\log(t) - \mu}{\sigma} \right) \right] \\ &= \Phi \left[ \frac{1}{\alpha} \left\{ \exp \left( \frac{\log(t) - \mu}{\sigma} \right) - \exp \left( \frac{\mu - \log(t)}{\sigma} \right) \right\} \right] \\ &= \Phi \left[ \frac{1}{\alpha} \left\{ (e^{\log(t/e^\mu)})^{\frac{1}{\sigma}} - (e^{\log(e^\mu/t)})^{\frac{1}{\sigma}} \right\} \right] \\ &= \Phi \left[ \frac{1}{\alpha} \left\{ \left(\frac{t}{e^\mu}\right)^{\frac{1}{\sigma}} - \left(\frac{e^\mu}{t}\right)^{\frac{1}{\sigma}} \right\} \right]. \end{aligned}$$

Derivamos a FDA de T em relação à t, obtemos

$$f_T(t) = \frac{d\Phi(s)}{dt} = \Phi'(s) \frac{ds}{dt} = \phi(s) \frac{ds}{dt},$$

em que

$$s = \frac{1}{\alpha} \left[ \left(\frac{t}{e^\mu}\right)^{\frac{1}{\sigma}} - \left(\frac{e^\mu}{t}\right)^{\frac{1}{\sigma}} \right] \quad e \quad \frac{ds}{dt} = \frac{t^{-(\frac{1}{\sigma})}}{\sigma\alpha\beta^{\frac{1}{\sigma}}} \left(t^{\frac{2}{\sigma}} + \beta^{\frac{2}{\sigma}}\right).$$

Portanto, a fdp de  $T$  é

$$\begin{aligned} f_T(t) &= \phi(s) \frac{t^{-\left(\frac{1}{\sigma}\right)}}{\sigma \alpha \beta^{\frac{1}{\sigma}}} \left( t^{\frac{2}{\sigma}} + \beta^{\frac{2}{\sigma}} \right) \\ &= \phi \left( \frac{1}{\alpha} \left[ \left( \frac{t}{e^\mu} \right)^{\frac{1}{\sigma}} - \left( \frac{e^\mu}{t} \right)^{\frac{1}{\sigma}} \right] \right) \frac{t^{-\left(\frac{1}{\sigma}\right)}}{\sigma \alpha \beta^{\frac{1}{\sigma}}} \left( t^{\frac{2}{\sigma}} + \beta^{\frac{2}{\sigma}} \right). \end{aligned}$$

Em vista que para  $\sigma = 2$  no Teorema 2.1.3, a distribuição de  $T$  corresponde à bem conhecida distribuição Birnbaum-Saunders. Assim, do resultado acima, podemos dizer que a distribuição de  $T$  em (2.1.6) corresponde a uma generalização da distribuição Birnbaum-Saunders usual. Neste caso,  $\alpha$  e  $\sigma$  são parâmetros de forma e o parâmetro  $\beta$  é a mediana da distribuição de  $T$ . Dizemos que a v.a.  $T$ , com pdf como em 2.1.6 segue uma distribuição Birnbaum-Saunders generalizada, que denotaremos por  $T \sim BSG(\alpha, \beta, \sigma)$ . ■

**Teorema 2.1.4.** *Se  $T \sim BS(\alpha, \beta)$ , então  $Y = \log(T)$  é distribuída como uma  $SN(\alpha, \log(\beta), 2)$ .*

**Demonstração.** Seja  $T \sim BS(\alpha, \beta)$ . Então, a FDA de  $Y = \log(T)$  é dada por

$$\begin{aligned} P(Y \leq y) &= P(T \leq \exp(y)) \\ &= \Phi \left[ \frac{1}{\alpha} \sqrt{\frac{\exp(y)}{\beta}} - \sqrt{\frac{\beta}{\exp(y)}} \right] \\ &= \Phi \left[ \frac{1}{\alpha} \sqrt{\exp(y - \log(\beta))} - \sqrt{\frac{1}{\exp(y - \log(\beta))}} \right] \\ &= \Phi \left[ \frac{1}{\alpha} \frac{\exp(y - \log(\beta))}{2} - \frac{\exp(-(y - \log(\beta)))}{2} \right] \\ &= \Phi \left[ \frac{2}{\alpha} \sinh \left( \frac{y - \log(\beta)}{2} \right) \right], \end{aligned}$$

que é a FDA da distribuição  $SN(\alpha, \log(\beta), 2)$ . ■

Na Figura 2.3, apresentamos os gráficos da densidade de  $T$  para valores de  $\alpha = 0.2, 0.5, 1$ ;  $\beta = 1$  e considerando  $\sigma = 1, 2, 3, 4$ . Observe que, a distribuição torna-se assimétrica à medida que  $\alpha$  cresce, e simétrica em torno de  $\beta$  à medida que  $\alpha$  se aproxima de zero. Nota-se que quando o valor de  $\sigma$  cresce a distribuição tende a ser assimétrica.

Na Figura 2.4 apresentamos gráficos de densidades de  $T$  para diferentes valores de  $\sigma$  considerando  $\alpha = 0.2$ ,  $\alpha = 0.8$  e  $\beta = 1$ . Observe que, para valores de  $\sigma \geq 2$  a medida que o valor de  $\sigma$  cresce a distribuição torna-se mais assimétrica.

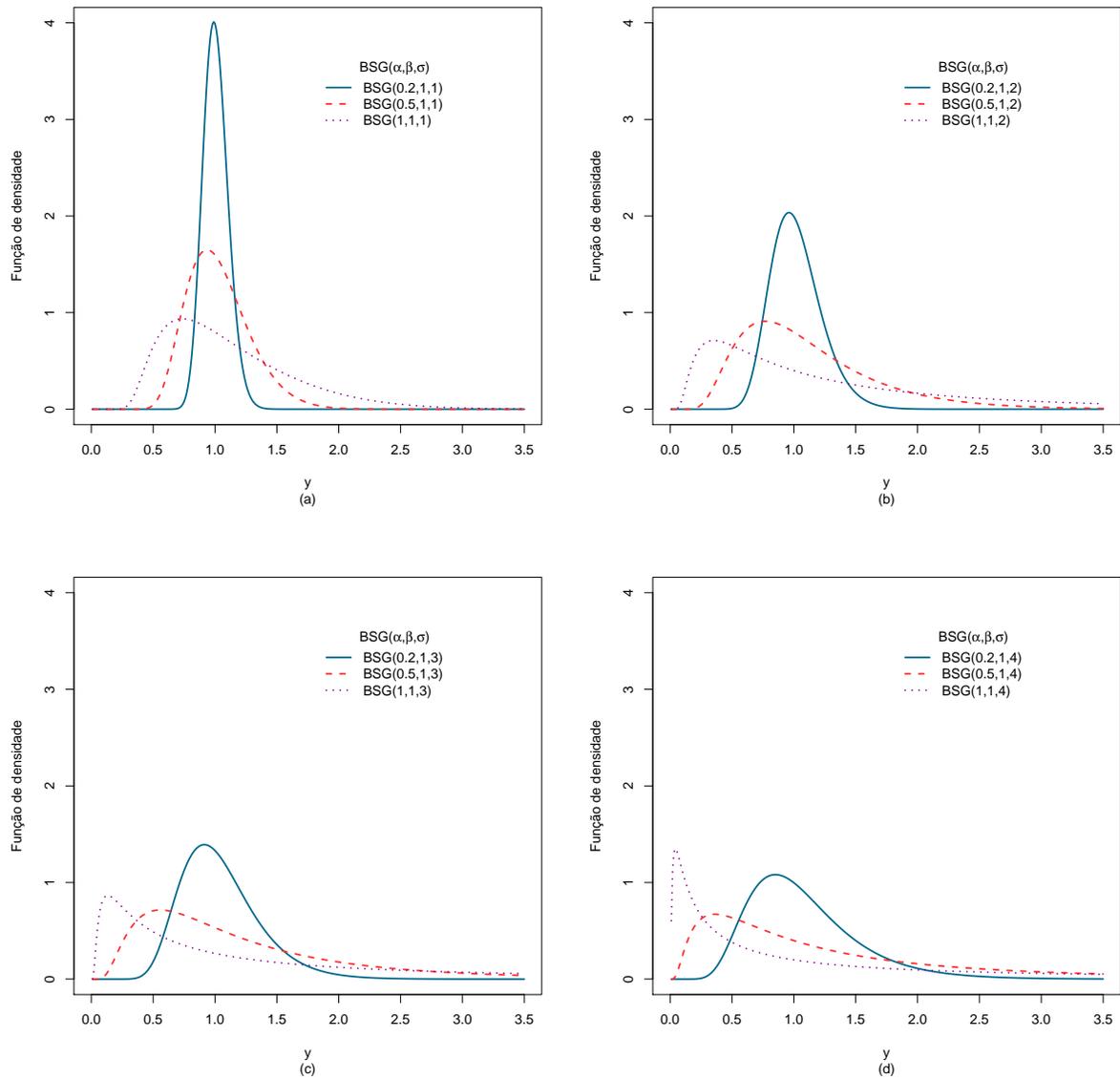


Figura 2.3: Densidade de probabilidade da distribuição BSG para  $\alpha < 2$  (a)  $\sigma = 1$  (b)  $\sigma = 2$  (c)  $\sigma = 3$  (d)  $\sigma = 4$ .

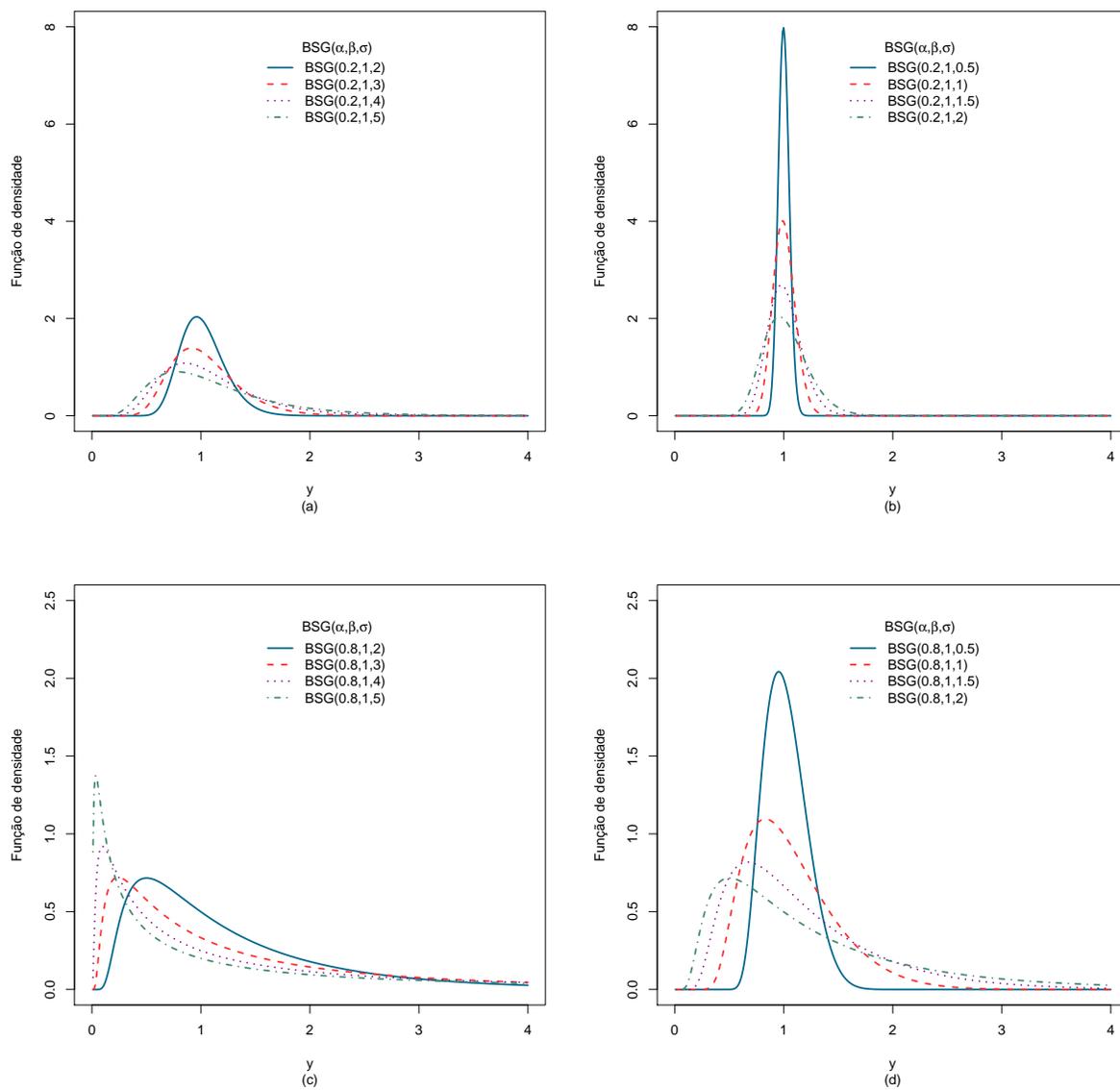


Figura 2.4: Função de densidade de probabilidade para a distribuição Birnbaum Saunders (a)  $\alpha = 0,2$  e  $\sigma \geq 2$  (b)  $\alpha = 0,2$  e  $\sigma \leq 2$  (c)  $\alpha = 0,8$  e  $\sigma \geq 2$  (d)  $\alpha = 0,8$  e  $\sigma \leq 2$ .

## 2.2 Momentos

Nesta subseção, apresentamos a função geradora momentos para a distribuição SN. Em seguida, usamos expansões assintóticas da função geradora para obter aproximações dos momentos Rieck (1989). Finalmente, proporciona-se um quadro para a variância da distribuição Sinh-Normal com valores de  $\alpha$  que são prováveis de ocorrer em aplicações.

**Teorema 2.2.1.** *Seja  $Y \sim SN(\alpha, \mu, \sigma)$ . Então, a função geradora de momentos de  $Y$  é*

$$m(t) = \exp(\mu t) \frac{K_a\left(\frac{1}{\alpha^2}\right) + K_b\left(\frac{1}{\alpha^2}\right)}{2K_{1/2}\left(\frac{1}{\alpha^2}\right)},$$

em que  $a = \frac{\sigma t + 1}{2}$ ,  $b = \frac{\sigma t - 1}{2}$ , e  $K_a(\cdot)$  é a função de Bessel modificada de terceiro tipo.

**Demonstração.** Seja  $S$  uma variável aleatória distribuída como  $SN(\alpha, 0, 1)$ . Então, a função de densidade de  $S$  pode ser escrita como

$$\begin{aligned} f(s) &= \frac{2}{\alpha\sqrt{2\pi}} \cosh(s) \exp\left[-\frac{2}{\alpha^2} \sinh^2(s)\right] \\ &= \frac{\exp\left(\frac{1}{\alpha^2}\right)}{\alpha\sqrt{2\pi}} [\exp(s) + \exp(-s)] \exp\left(-\frac{1}{\alpha^2} \cosh(2s)\right). \end{aligned}$$

Agora, a função geradora de momentos pode ser escrita como

$$\begin{aligned} E[\exp(tS)] &= \left[\frac{\exp\left(\frac{1}{\alpha^2}\right)}{\alpha\sqrt{2\pi}}\right] \int \exp\left[(t+1)s - \frac{1}{\alpha^2} \cosh(2s)\right] ds \\ &\quad + \left[\frac{\exp\left(\frac{1}{\alpha^2}\right)}{\alpha\sqrt{2\pi}}\right] \int \exp\left[(t-1)s - \frac{1}{\alpha^2} \cosh(2s)\right] ds, \end{aligned}$$

em que os limites de integração são de  $-\infty$  para  $+\infty$ . Usando uma mudança de variável de  $x = 2s$ , a integral torna-se

$$\begin{aligned} E[\exp(tS)] &= \left[\frac{\exp\left(\frac{1}{\alpha^2}\right)}{\alpha\sqrt{2\pi}}\right] \int \exp\left[\frac{(t+1)x}{2} - \frac{1}{\alpha^2} \cosh(x)\right] dx + \\ &\quad \left[\frac{\exp\left(\frac{1}{\alpha^2}\right)}{\alpha\sqrt{2\pi}}\right] \int \exp\left[\frac{(t-1)x}{2} - \frac{1}{\alpha^2} \cosh(x)\right] dx \\ &= \frac{K_a\left(\frac{1}{\alpha^2}\right) + K_b\left(\frac{1}{\alpha^2}\right)}{2K_{1/2}\left(\frac{1}{\alpha^2}\right)}, \end{aligned}$$

em que  $a = \frac{t+1}{2}$ ,  $b = \frac{t-1}{2}$ , e  $K$  é a função de Bessel modificada do terceiro tipo. Para uma descrição da função de Bessel modificada de terceiro tipo veja Abramowitz & Stegun (1972). A

função de geração de momentos para uma SN central geral pode ser encontrado usando a transformação  $Y = \mu + \sigma S$ . A função geradora de momentos resultante é a indicada no teorema. ■

Os momentos inteiros avaliados da distribuição BS são dadas em Jonhson & Kotz (1970). Como consequência imediata do Teorema 2.2.1, podemos encontrar todos os momentos, fracionados assim como integrais, para uma variável aleatória  $X$  de uma distribuição BS. Pelo Teorema 2.1.4, temos que  $Y = \log(X)$  é distribuído como uma  $SN(\alpha, \log(\beta), 2)$ . Portanto a função de geração de momentos para  $Y$  é dada por  $m(t)$  com  $a = \frac{2t+1}{2}$ ,  $b = \frac{2t-1}{2}$  e  $\mu = \log(\beta)$ . Desde que

$$m(t) = E[\exp(tY)] = E(X^t),$$

temos que o  $t$ -ésimo momento da distribuição Birnbaum-Saunders é  $m(t)$  com os parâmetros acima indicados.

Em princípio, os momentos podem ser encontrados pela diferenciação da função geradora de momentos em relação a  $t$  e, em seguida, definindo  $t = 0$ . No entanto, para esta distribuição é difícil de obter expressões significativas para as derivadas. A seguir, apresentamos expressões assintóticas para a função geradora de momentos quando  $\alpha$  é pequeno e grande.

Valores pequenos de  $\alpha$  correspondem aos grandes valores para o argumento  $1/\alpha^2$  da função de Bessel acima. Para um grande argumento  $z$ , Abramowitz & Stegun (1972) fornecem as seguintes expansões assintóticas da função Bessel de terceiro tipo

$$K_\nu(z) \simeq \exp(-z) \frac{\sqrt{\pi}}{2z} \left[ 1 + \frac{\mu-1}{8} + \frac{(\mu-1)(\mu-9)}{2!(8z)^2} + \dots \right],$$

em que  $\mu = 4\nu^2$ . Usando os três primeiros termos desta expansão, temos que a função geradora de momentos é

$$m(t) \simeq \exp(\mu t) \left[ 1 + \frac{\alpha^2 \sigma^2 t^2}{8} + \frac{\alpha^4 (\sigma^4 t^4 - 4\sigma^2 t^2)}{128} \right]. \quad (2.2.1)$$

Para grandes valores de  $\alpha$ , os momentos podem ser aproximados utilizando o seguinte resultado respeito da função de Bessel modificada. A aproximação de primeiro ordem para  $K_a\left(\frac{1}{\alpha^2}\right)$ , para  $a > 0$ , podem ser encontradas em Abramowitz & Stegun (1972), e é dada pela

$$K_a\left(\frac{1}{\alpha^2}\right) \simeq \Gamma(a) 2^{a-1} \alpha^{2a}.$$

usando essa expansão assintótica, temos que a função geradora de momentos é

$$m(t) \simeq \frac{1}{2\sqrt{\pi}} \exp(t\mu) \left\{ \Gamma\left[\frac{\sigma t + 1}{2}\right] (\alpha\sqrt{2})^{\sigma t} + \Gamma\left[\frac{\sigma(1 - \sigma t)}{2}\right] (\alpha\sqrt{2})^{-\sigma t} \right\}, \quad (2.2.2)$$

para  $|t| < \frac{1}{\sigma}$

A média e variância da distribuição SN são

$$\mathbb{E}[Y] = \mu \quad \text{e} \quad \text{Var}(Y) = \sigma^2 \eta(\alpha), \quad (2.2.3)$$

respectivamente. Em que  $\eta(\alpha) = \frac{\alpha^2}{4} \left(1 - \frac{\alpha^2}{4}\right)$  para  $0 < \alpha < 2$  baseado na aproximação em (2.2.1) e  $\eta(\alpha) = 2.197543451 - 1.963510026 \log(\alpha\sqrt{2}) + [\log(\alpha\sqrt{2})]^2$  para  $\alpha > 2$  baseado na aproximação (2.2.2).

Na Figura 2.5, apresentamos gráficos para comparar a distribuição Normal com a distribuição SN. Observe que, a medida que  $\sigma$  cresce a altura de ambas distribuições diminui e sua amplitude tende a aumentar. Portanto, ambas distribuições apresentam um comportamento similar.

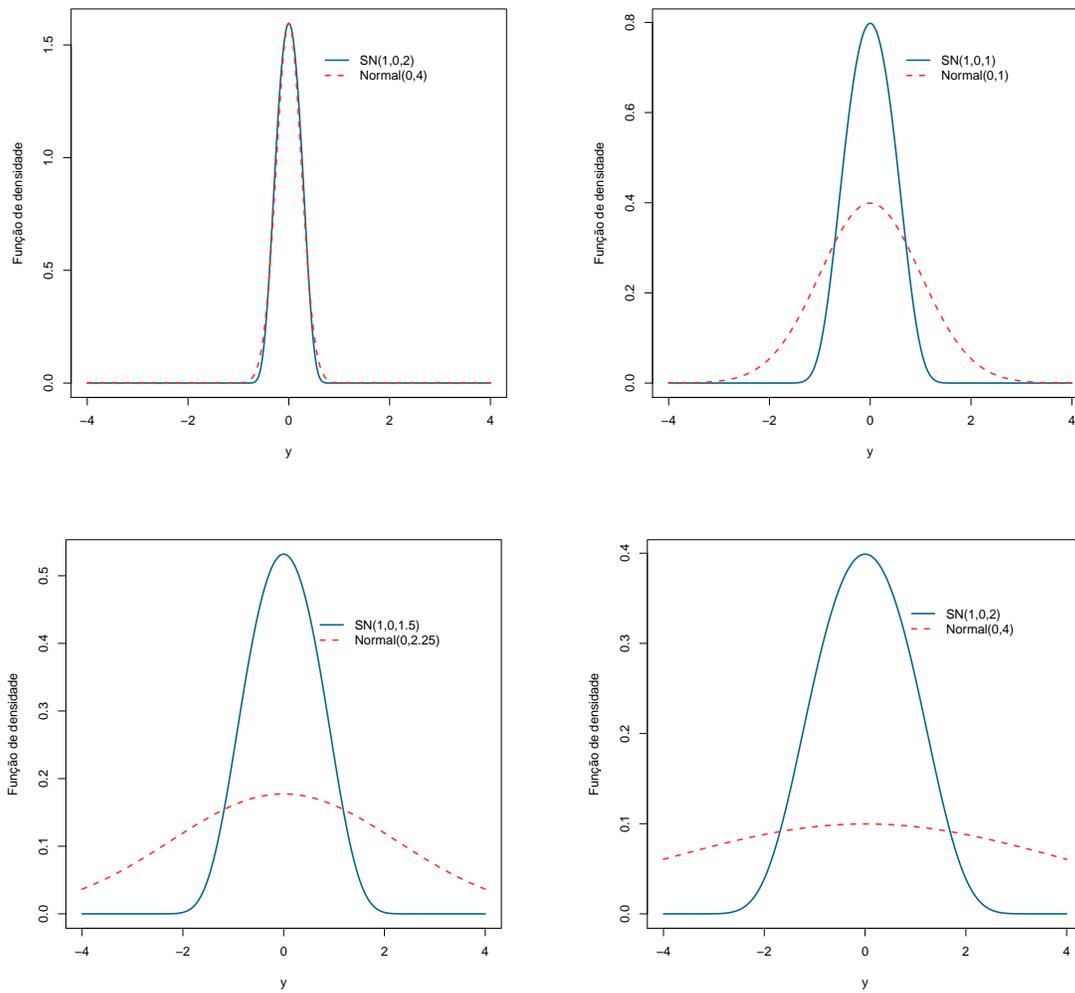


Figura 2.5: Comparação das funções de densidade de probabilidade das distribuições Normal e Sinh-Normal.

# Capítulo 3

## Modelo de regressão Sinh-Normal

### 3.1 Introdução

Rieck & Nedelman (1991) apresentaram um modelo de regressão log-BS baseado na distribuição SN. Esse modelo de regressão tem sido estudado por vários autores. Algumas referências importantes são Tisionas (2001), Galea *et al.* (2004), Leiva *et al.* (2007), Lemonte & Cordeiro (2010), Xiao *et al.* (2010) e Cancho *et al.* (2010), entre outros. Generalizações do modelo de regressão log-BS introduzido por Rieck & Nedelman (1991) são apresentados em Xie & Wei (2007) e Lemonte & Cordeiro (2009).

### 3.2 Descrição do modelo

O modelo de regressão sob a distribuição SN, é definido por

$$Y_i = \mu_i + \epsilon_i, \quad i = 1, \dots, n, \quad (3.2.1)$$

em que  $Y_i$  é a variável resposta,  $\epsilon_i \sim SN(\alpha, 0, \sigma)$  e  $\mu_i$  depende de covariáveis através da relação  $g(\mu_i) = \mathbf{x}_i^\top \boldsymbol{\beta}$ , em que  $\mathbf{x}_i^\top = (x_{i1}, \dots, x_{ip})$  é um vetor de covariáveis e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  é o vetor de parâmetros. Sendo  $g(\cdot)$  a função de ligação, que é monotona e pelo menos duas vezes diferenciável como em modelos lineares generalizados. Suponha que  $g_0(\cdot)$  é a função inversa de  $g$ . Então podemos escrever  $\mu_i$  da seguinte forma  $\mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta}) = g_0(\mathbf{x}_i^\top \boldsymbol{\beta})$ . De acordo com as propriedades da distribuição SN, vamos ter que

$$\mathbb{E}[Y_i] = \mu_i \quad \text{e} \quad \text{Var}(Y_i) = \sigma^2 \eta(\alpha), \quad \alpha > 0,$$

em que  $\eta(\alpha)$  é como em (2.2.3). Sob as condições descritas, o modelo definido em (3.2.1) é chamado de modelo de regressão Sinh-Normal (RSN).

Como a distribuição SN é simétrica, o modelo de regressão SN representa uma alternativa de aqueles já estudados na literatura, veja, por exemplo, Villegas *et al.* (2013). Em vista que a distribuição SN está fortemente relacionada com a distribuição normal, poderíamos pensar que o

modelo de regressão SN é uma alternativa ao modelo de regressão normal, para  $\alpha < 2$ , pois neste caso a distribuição é unimodal.

Do um ponto de vista de tempo de vida sabemos que se  $T \sim BSG(\alpha, \beta, \sigma)$ , então  $Y = \log(T) \sim SN(\alpha, \mu, \sigma)$ . Assim o modelo de regressão definido em (3.2.1), pode ser visto como uma generalização do modelo de regressão log-BS, amplamente discutido em Rieck & Nedelman (1991), Galea *et al.* (2004), Xie & Wei (2007) e Lemonte & Cordeiro (2009).

A fdp do modelo de regressão SN pode ser escrito como

$$\begin{aligned} f(y_i; \boldsymbol{\theta}) &= \frac{2}{\alpha\sigma\sqrt{2\pi}} \exp\left(-\frac{2}{\alpha^2} \sinh^2\left(\frac{y_i - \mu_i}{\sigma}\right)\right) \cosh\left(\frac{y_i - \mu_i}{\sigma}\right), \\ &= \sigma^{-1} \phi(\xi_{2i}) \xi_{1i}, \end{aligned} \quad (3.2.2)$$

em que  $\xi_{1i} = \xi_{1i}(\alpha, \boldsymbol{\beta}, \sigma) = \frac{2}{\alpha} \cosh\left(\frac{y_i - \mu_i}{\sigma}\right)$ ,  $\xi_{2i} = \xi_{2i}(\alpha, \boldsymbol{\beta}, \sigma) = \frac{2}{\alpha} \sinh\left(\frac{y_i - \mu_i}{\sigma}\right)$  e  $\phi(\cdot)$  denota a função de densidade da distribuição normal padrão.

Para uma amostra  $y_1, \dots, y_n$  do modelo de RSN, a função de verossimilhança é dada por

$$L(\boldsymbol{\theta}; y_1, \dots, y_n) = \prod_{i=1}^n \left\{ \frac{2}{\alpha\sigma\sqrt{2\pi}} \exp\left(-\frac{2}{\alpha^2} \sinh^2\left(\frac{y_i - \mu_i}{\sigma}\right)\right) \cosh\left(\frac{y_i - \mu_i}{\sigma}\right) \right\},$$

em que  $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top, \sigma)^\top$

e a função de log-verossimilhança pode ser escrita como

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_j(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(y_j; \boldsymbol{\theta}), \quad (3.2.3)$$

em que

$$\ell_j(\boldsymbol{\theta}) = -\frac{1}{2} \log(2\pi) - \log(\sigma) + \log(\xi_{1i}) - \frac{1}{2} \xi_{2i}^2.$$

Na seguinte seção apresentamos, as primeiras e segundas derivadas da função de log-verossimilhança que serão usadas na implementação do método de diagnóstico.

### 3.3 Matriz de Informação Observada

Nesta subsecção, vamos apresentar a matriz de informação observada usando as notações de (3.2.2). Esta matriz será utilizada na parte de estimação e na aplicação do método de influência local.

### 3.3.1 Derivadas parciais de primeira ordem

Como  $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_j(\boldsymbol{\theta})$  então  $\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \frac{\partial \ell_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ ,  $D$  - matriz diagonal,  $\mathbf{1}_n^T = (1, 1, 1, \dots, 1)$ ,

$\mathbf{U}(\boldsymbol{\theta}) = \left( \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}, \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha}, \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} \right)^\top$  em que os elementos são

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= \frac{1}{\sigma} \mathbf{D}_\beta [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)] \boldsymbol{\xi}_2, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \alpha} &= \frac{1}{\alpha} [\mathbf{1}_n^T D(\boldsymbol{\xi}_2) \boldsymbol{\xi}_2 - n], \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma} &= \frac{1}{\sigma} [\boldsymbol{\epsilon}^T (D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)) \boldsymbol{\xi}_2 - n], \end{aligned}$$

em que

$$\begin{aligned} \mathbf{D}_\beta &= \frac{\partial \boldsymbol{\mu}^\top}{\partial \boldsymbol{\beta}} = \left( \frac{\partial \mu_1}{\partial \boldsymbol{\beta}}, \dots, \frac{\partial \mu_n}{\partial \boldsymbol{\beta}} \right), \quad \boldsymbol{\xi}_1 = (\xi_{11}, \dots, \xi_{1n})^\top, \quad \boldsymbol{\xi}_2 = (\xi_{21}, \dots, \xi_{2n})^\top \quad \text{e} \\ \boldsymbol{\epsilon} &= (\epsilon_1, \dots, \epsilon_n)^\top \quad \text{com} \quad \epsilon_i = (y_i - \mu_i)/\sigma. \end{aligned}$$

### 3.3.2 Derivadas parciais de segunda ordem

Seja  $\ell(\boldsymbol{\theta})$  a função de log-verossimilhança definida em (3.2.3). Então, a matriz de derivadas parciais de segunda ordem com respeito de  $\boldsymbol{\theta}$ , é dada por

$$\ddot{\mathbf{L}} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \sum_{j=1}^n \frac{\partial^2 \ell_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}. \quad (3.3.1)$$

Assim, considerando as propriedades das derivadas vetoriais e após algumas manipulações algébricas, temos os seguintes elementos para a matriz de informação observada

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{1}{\sigma} \left[ -\frac{1}{\sigma} \mathbf{D}_\beta \mathbf{A}_{12} \mathbf{D}_\beta^\top + \mathbf{D}_{\beta\beta} \mathbf{B}_{12} \right], \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \alpha} &= -\frac{2}{\alpha \sigma} \mathbf{D}_\beta D(\boldsymbol{\xi}_1) \boldsymbol{\xi}_2, \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma} &= -\frac{1}{\sigma^2} \mathbf{D}_\beta [\mathbf{A}_{12} \boldsymbol{\epsilon} + [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)] \boldsymbol{\xi}_2], \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \alpha} &= \frac{1}{\alpha^2} [n - 3 \mathbf{1}_n^\top D(\boldsymbol{\xi}_2) \boldsymbol{\xi}_2], \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma \partial \alpha} &= -\frac{2}{\alpha \sigma} \boldsymbol{\epsilon}^\top D(\boldsymbol{\xi}_1) \boldsymbol{\xi}_2, \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma \partial \sigma} &= \frac{1}{\sigma^2} [n - \boldsymbol{\epsilon}^\top \mathbf{A}_{12} \boldsymbol{\epsilon} - 2 \boldsymbol{\epsilon}^\top [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)] \boldsymbol{\xi}_2], \end{aligned}$$

em que

$$\begin{aligned}\mathbf{D}_{\beta\beta} &= \frac{\partial \text{Vec}(\mathbf{D}_\beta)}{\partial \beta} = \left( \frac{\partial^2 \mu_1}{\partial \beta \partial \beta^\top}, \dots, \frac{\partial^2 \mu_n}{\partial \beta \partial \beta^\top} \right), \\ \mathbf{A}_{12} &= [D^2(\boldsymbol{\xi}_1) + D^2(\boldsymbol{\xi}_2)] + D^2(\boldsymbol{\xi}_2)D^{-2}(\boldsymbol{\xi}_1) - \mathbf{I}_n \quad \text{e} \\ \mathbf{B}_{12} &= [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)]\boldsymbol{\xi}_2 \otimes \mathbf{I}_p,\end{aligned}$$

em que  $\text{Vec}(\mathbf{A})$  denota o vetor dado por  $\text{Vec}(\mathbf{A}) = (\mathbf{A}_{\cdot 1}^\top, \dots, \mathbf{A}_{\cdot s}^\top)^\top$ , onde  $\mathbf{A}_{\cdot s}$  indica a  $j$ -ésima coluna da matriz  $\mathbf{A}$  de ordem  $r \times s$  e  $\otimes$  é o produto de Kronecker.

### 3.4 Estimação de parâmetros

Dado que as equações resultantes para encontrar as EMV não tem forma fechada, métodos numéricos são requeridos para encontrar as EMV dos parâmetros do modelo. O método de Newton-Raphson pode ser utilizado, para solucionar as equações não lineares de verossimilhança, para determinar o ponto em que a função de verossimilhança é maximizada. O conjunto de equações iterativas do método de Newton Raphson é assim definido:

$$\hat{\boldsymbol{\theta}}^{(t+1)} = \hat{\boldsymbol{\theta}}^{(t)} + [\mathbf{I}(\hat{\boldsymbol{\theta}}^{(t)})]^{-1} \mathbf{U}(\boldsymbol{\theta}^{(t)}), \quad (3.4.1)$$

em que  $\mathbf{I}(\hat{\boldsymbol{\theta}}) = -\ddot{\mathbf{L}}$ , sendo que  $\hat{\boldsymbol{\theta}}^{(t)}$  e  $\hat{\boldsymbol{\theta}}^{(t+1)}$  são as estimativas de  $\boldsymbol{\theta}$  nos passos  $t$  e  $t + 1$ , respectivamente e  $\mathbf{U}(\boldsymbol{\theta})$  é a função escore. As iterações são repetidas até que uma regra de convergência adequada seja satisfeita. Como critério de convergência, podemos utilizar  $\|\hat{\boldsymbol{\theta}}^{(t+1)} - \hat{\boldsymbol{\theta}}^{(t)}\| < \epsilon$ , onde  $\|\mathbf{a}\|$  indica a norma do vetor  $\mathbf{a}$  e  $\epsilon > 0$ . Valores iniciais são necessários para implementar este algoritmo. Neste caso  $\beta^{(0)}$  é obtido por mínimos quadrados, nos outros casos  $\alpha^{(0)}$  e  $\sigma^{(0)}$  são respectivamente dados por

$$\begin{aligned}\alpha^{(0)} &= \sqrt{\frac{4}{n} \sum_{i=1}^n [\sinh(\Delta_i)]^2} \quad \text{e} \\ \sigma^{(0)} &= 2 \left[ \frac{\sum_{i=1}^n \sinh(\Delta_i) \cosh(\Delta_i) \Delta_i}{\sum_{i=1}^n [\sinh(\Delta_i)]^2} - \frac{1}{n} \sum_{i=1}^n \frac{\sinh(\Delta_i) \Delta_i}{\cosh(\Delta_i)} \right],\end{aligned} \quad (3.4.2)$$

em que  $\Delta_i = (y_i - g_0(\mathbf{x}_i^\top \boldsymbol{\beta}^{(0)})) / 2$  e  $g_0(\cdot)$  é a inversa da função de ligação  $g(\cdot)$ .

O método de Newton Raphson poder gerar uma sequencia de valores que convergem para um ponto onde a função tem um máximo local, isto acontece quando os valores iniciais não são apropriados. Em consequência, é recomendável repetir algumas vezes o processo considerando valores iniciais distintos, para determinar o ponto onde a função alcança seu valor máximo.

### 3.4.1 Seleção de modelos

Para fazer a escolha do modelo apropriado, alguns critérios geralmente utilizados serão considerados: O Critério de Informação de Akaike (AIC), Akaike (1973) e o Critério de Informação Bayesiano (BIC), Schwarz (1978), que são definidos respectivamente por

$$AIC = -2\ell(\hat{\boldsymbol{\theta}}) + 2(p + 2) \quad \text{e} \quad BIC = -2\ell(\hat{\boldsymbol{\theta}}) + (p + 2)\log(n), \quad (3.4.4)$$

em que  $\ell(\cdot)$  é a log-verossimilhança,  $p$  é o número de elementos do parâmetro  $\beta$  estimado. Tanto o AIC quanto o BIC aumentam conforme a soma dos quadrados dos erros (SQE) aumenta. Além disso, ambos critérios penalizam modelos com muitas variáveis sendo que valores menores de AIC e BIC são preferíveis. Como modelos com mais variáveis tendem a produzir menor SQE, mas usam mais parâmetros, a melhor escolha é balancear o ajuste com a quantidade de variáveis.

### 3.4.2 Análise de resíduos

A análise de resíduos tem como finalidade identificar observações atípicas, assim como verificar a correta especificação do modelo ajustado. Para isso, neste trabalho consideramos os seguintes resíduos, o primeiro é o comumente chamado de resíduo de Pearson,

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\widehat{\text{Var}}(y_i)}}, \quad (3.4.5)$$

em que  $\widehat{\text{Var}}(y_i) = \hat{\sigma}^2 \frac{\hat{\alpha}^2}{4} (1 - \frac{\hat{\alpha}^2}{4})$ , para valores pequenos de  $\alpha$  e  $\widehat{\text{Var}}(y_i) = \hat{\sigma}^2 [2.197543451 - 1.963510026 \ln(\hat{\alpha}\sqrt{2}) + [\ln(\hat{\alpha}\sqrt{2})]^2]$ , para valores grandes de  $\alpha$ ,  $\hat{\mu}_i = g_0(\mathbf{x}_i^\top \hat{\boldsymbol{\beta}})$ ,  $i = 1, \dots, n$ , sendo  $g_0(\cdot)$  a inversa da função de ligação, com  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\alpha}$  e  $\hat{\sigma}$  sendo os EMV de  $\boldsymbol{\beta}$ ,  $\alpha$  e  $\sigma$ , respectivamente.

Um segundo resíduo é considerado seguindo a relação da distribuição SN com a distribuição normal padrão discutidos na Seção 2.1. A saber,

$$r_i = \frac{2}{\hat{\alpha}} \sinh\left(\frac{y - \hat{\mu}_i}{\hat{\sigma}}\right), \quad i = 1, \dots, n. \quad (3.4.6)$$

Com estes resíduos construímos envelopes simulados, como foi sugerido por Atkinson (1981), para detectar uma incorreta especificação da distribuição do erro, da componente sistemática, assim como, a presença de observações atípicas.

## 3.5 Análise de Diagnóstico

As técnicas de diagnóstico de influência consistem em avaliar a sensibilidade das estimativas dos parâmetros em um determinado modelo, quando ocorre uma perturbação no conjunto de dados ou nos pressupostos do modelo. Existem, basicamente, duas abordagens para detectar observações atípicas. A primeira abordagem é a eliminação de casos, que é uma abordagem para analisar o modelo ajustado após a exclusão de uma ou mais observações que são avaliados diretamente

por algumas métricas, como a distância verossimilhança e distância do Cook discutidos em Cook (1977). A segunda abordagem é o método de influência local proposta por Cook (1986), que tem o objetivo de avaliar as mudanças nos resultados da análise, quando é incorporada uma pequena perturbação no modelo.

De acordo com Xie & Wei (2007), vamos apresentar o modelo de eliminação de casos e de acordo com Galea *et al.* (2004) as medidas de influência local na base da seguinte função  $\ell(\boldsymbol{\theta})$ . Consideremos primeiramente o modelo de eliminação de casos.

### 3.5.1 Modelo de eliminação de casos

A seguir, vamos discutir o efeito da eliminação de uma determinada observação na EMV de  $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}^\top, \widehat{\alpha}, \widehat{\sigma})^\top$ . Se a  $i$ -ésima observação é deletada, denotaremos por  $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\alpha}_{[i]}, \widehat{\sigma}_{[i]})^\top$  a EMV de  $\boldsymbol{\theta}$  que maximiza a função de log-verossimilhança  $\ell_{[i]}(\boldsymbol{\theta}) = \sum_{j \neq i} \ell_j(\boldsymbol{\theta})$ .

Para avaliar que a influência do  $i$ -ésimo caso na estimativa de máxima verossimilhança  $\widehat{\boldsymbol{\theta}}$ , a idéia básica é comparar a diferença entre  $\widehat{\boldsymbol{\theta}}_{[i]}$  e  $\widehat{\boldsymbol{\theta}}$ . Assim, se  $\widehat{\boldsymbol{\theta}}_{[i]}$  está muito afastado de  $\widehat{\boldsymbol{\theta}}$ , o  $i$ -ésimo caso é considerado como influente.

Como  $\widehat{\boldsymbol{\theta}}_{[i]}$  é necessário para todos os casos, a carga computacional envolvida pode ser bastante pesada, especialmente quando uma amostra é grande. Sendo assim, a seguinte, pseudo aproximação um passo  $\widehat{\boldsymbol{\theta}}_{[i]}^1$  é utilizada para reduzir a carga, veja Cook & Weisberg (1982) e Xie & Wei (2007)

$$\widehat{\boldsymbol{\theta}}_{[i]}^1 = \widehat{\boldsymbol{\theta}} + (-\ddot{\mathbf{L}})^{-1} \dot{\ell}_{[i]}(\widehat{\boldsymbol{\theta}}), \quad (3.5.1)$$

em que  $\ddot{\mathbf{L}} = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$  e  $\dot{\ell}_{[i]}(\widehat{\boldsymbol{\theta}}) = \frac{\partial \ell_{[i]}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}$ , tem os seus elementos dados por

$$\dot{\ell}_{[i]\beta}(\widehat{\boldsymbol{\theta}}) = \partial \ell_{[i]}(\widehat{\boldsymbol{\theta}}) / \partial \boldsymbol{\beta} = \frac{1}{\widehat{\sigma}} \sum_{j \neq i} \left( \widehat{\xi}_{1j} \widehat{\xi}_{2j} - \frac{\widehat{\xi}_{2j}}{\widehat{\xi}_{1j}} \right) \frac{\partial \mu_j}{\partial \boldsymbol{\beta}}, \quad (3.5.2)$$

$$\dot{\ell}_{[i]\alpha}(\widehat{\boldsymbol{\theta}}) = \partial \ell_{[i]}(\widehat{\boldsymbol{\theta}}) / \partial \alpha = \frac{1}{\widehat{\alpha}} \sum_{j \neq i} \widehat{\xi}_{2j}^2 - \frac{n-1}{\widehat{\alpha}} e \quad (3.5.3)$$

$$\dot{\ell}_{[i]\sigma}(\widehat{\boldsymbol{\theta}}) = \partial \ell_{[i]}(\widehat{\boldsymbol{\theta}}) / \partial \sigma = -\frac{n-1}{\widehat{\sigma}} + \frac{1}{\widehat{\sigma}} \sum_{j \neq i} \left( \widehat{\xi}_{1j} \widehat{\xi}_{2j} - \frac{\widehat{\xi}_{2j}}{\widehat{\xi}_{1j}} \right) \widehat{\epsilon}_j, \quad (3.5.4)$$

em que  $\epsilon_j = (y_j - \mu_j) / \sigma$ .

Em seguida, vamos obter a aproximação de um passo de  $\widehat{\boldsymbol{\theta}}_{[i]} = (\widehat{\boldsymbol{\beta}}_{[i]}^\top, \widehat{\alpha}_{[i]}, \widehat{\sigma}_{[i]})^\top$ ,  $i = 1, \dots, n$  baseado em (3.5.1), ou seja, a relação entre as estimativas dos parâmetros de dados completos e dos parâmetros com o  $i$ -ésimo caso excluído é obtido seguindo a ideia de estimativa dos parâmetros Xie & Wei (2007).

**Teorema 3.5.1.** *Para a regressão Sinh-Normal, a relação entre estimativas de parâmetros de dados completos e os dados com o  $i$ -ésimo caso excluído pode ser expressada como*

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{[i]}^1 &= \widehat{\boldsymbol{\beta}} + [\ddot{\mathbf{P}}_{\beta.}] \dot{\ell}_{[i]}(\widehat{\boldsymbol{\theta}}), \\ \widehat{\alpha}_{[i]}^1 &= \widehat{\alpha} + [\ddot{\mathbf{P}}_{\alpha.}] \dot{\ell}_{[i]}(\widehat{\boldsymbol{\theta}}), \\ \widehat{\sigma}_{[i]}^1 &= \widehat{\sigma} + [\ddot{\mathbf{P}}_{\sigma.}] \dot{\ell}_{[i]}(\widehat{\boldsymbol{\theta}}),\end{aligned}$$

em que  $[\ddot{\mathbf{P}}_{\beta.}]$ ,  $[\ddot{\mathbf{P}}_{\alpha.}]$  e  $[\ddot{\mathbf{P}}_{\sigma.}]$  são as linhas  $(p+2) \times (p+2)$  da matriz  $\ddot{\mathbf{P}} = [-\ddot{\mathbf{L}}]^{-1}$ .

Do Teorema 3.5.1 pode ser desenvolvido um método de eliminação de casos para avaliar observações influentes, baseado na distância generalizada de Cook, e a distância de verosimilhança. Para avaliar a influência do  $i$ -ésimo caso na estimativa de máxima verosimilhança  $\widehat{\boldsymbol{\theta}}$ , a ideia básica é a de comparar a diferença entre  $\widehat{\boldsymbol{\theta}}_{[i]}$  e  $\widehat{\boldsymbol{\theta}}$ . Assim, se  $\widehat{\boldsymbol{\theta}}_{[i]}$  está afastado de  $\widehat{\boldsymbol{\theta}}$  em algum sentido, então o  $i$ -ésimo caso é considerado como influente. Inspirado nas métricas para medir a distância de  $\widehat{\boldsymbol{\theta}}_{[i]}$  e  $\widehat{\boldsymbol{\theta}}$  proposto por Cook (1977) e Xie & Wei (2007), consideramos o seguinte *distância de Cook generalizada* que é dado da seguinte forma:

$$GD_i = (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}})^\top \{-\ddot{\mathbf{L}}\} (\widehat{\boldsymbol{\theta}}_{[i]} - \widehat{\boldsymbol{\theta}}), \quad i = 1, \dots, n. \quad (3.5.5)$$

Substituindo (3.5.1) em (3.5.5), obtemos a seguinte aproximação:

$$GD_i^1 = \dot{\ell}_{[i]}(\widehat{\boldsymbol{\theta}})^\top \{-\ddot{\mathbf{L}}\}^{-1} \dot{\ell}_{[i]}(\widehat{\boldsymbol{\theta}}), \quad i = 1, \dots, n.$$

Por outro lado, se o interesse principal é avaliar a influência das observações de um subconjunto particular de  $\boldsymbol{\theta}$ , por exemplo,  $\boldsymbol{\beta}$ ,  $\alpha$  ou  $\sigma$ , pode ser facilmente considerado a influência do  $i$ -ésimo caso em alguns subconjuntos de parâmetros. Seguindo o mesmo procedimento usado por Xie & Wei (2007), podemos ter a distância de Cook generalizada para o parâmetro subconjunto  $\boldsymbol{\beta}$ ,  $\alpha$  e  $\sigma$  que são, respectivamente, dados por

$$\begin{aligned}GD_i^1(\boldsymbol{\beta}) &= \dot{\ell}_{[i]\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}})^\top \ddot{\mathbf{P}}_{\boldsymbol{\beta}\boldsymbol{\beta}} \dot{\ell}_{[i]\boldsymbol{\beta}}(\widehat{\boldsymbol{\theta}}), \\ GD_i^1(\alpha) &= \ddot{\mathbf{P}}_{\alpha\alpha} \dot{\ell}_{[i]\alpha}(\widehat{\boldsymbol{\theta}})^2, \\ GD_i^1(\sigma) &= \ddot{\mathbf{P}}_{\sigma\sigma} \dot{\ell}_{[i]\sigma}(\widehat{\boldsymbol{\theta}})^2,\end{aligned}$$

em que  $\ddot{\mathbf{P}}_{\gamma\gamma}$  indica que a entrada da matriz  $\ddot{\mathbf{P}} = \{-\ddot{\ell}(\widehat{\boldsymbol{\theta}})\}^{-1}$  correspondentes a  $\gamma = \boldsymbol{\beta}, \alpha, \sigma$ .

Outra medida para avaliar a influência do  $i$ -ésimo caso é a distância de verosimilhança  $LD_i$  que é definida como

$$LD_i = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{[i]})\}. \quad (3.5.6)$$

Podemos obter uma aproximação de  $LD_i$  substituindo (3.5.1) em (3.5.6), assim obtemos a seguinte aproximação  $LD_i^1$  de  $LD_i$  dada por

$$LD_i^1 = 2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{[i]}^1)\}. \quad (3.5.7)$$

### 3.5.2 Influência Local

Nesta seção vamos apresentar em forma resumida o método de influência local proposto por Cook (1986), e alguns métodos alternativos que tem sido utilizados para avaliar a influência de determinada observação.

Para um conjunto de dados seja  $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$  o logaritmo da função de log-verossimilhança do modelo postulado, em que  $\boldsymbol{\theta}$  é um vetor  $(p+2) \times 1$  de parâmetros desconhecidos e  $\ell_i(\boldsymbol{\theta})$  é a contribuição da  $i$ -ésima observação à log-verossimilhança  $\ell(\boldsymbol{\theta})$ . A perturbação no modelo é introduzida através de um vetor  $\boldsymbol{\omega}$  de dimensão  $q \times 1$ , em que  $\boldsymbol{\omega} \in \Omega \subseteq \mathbb{R}^q$ ,  $\Omega$  um aberto. Seja  $\ell(\boldsymbol{\theta}, \boldsymbol{\omega})$  a função de log-verossimilhança do modelo perturbado. E seja  $\boldsymbol{\omega}_0 \in \Omega$  tal que  $\ell(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}, \boldsymbol{\omega}_0)$ , para todo  $\boldsymbol{\theta}$ . Vamos supor que  $\ell(\boldsymbol{\theta}, \boldsymbol{\omega})$  seja duas vezes diferenciável em  $(\boldsymbol{\theta}^\top, \boldsymbol{\omega}^\top)^\top$ .

Sejam  $\widehat{\boldsymbol{\theta}}$  e  $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  as EMV de  $\boldsymbol{\theta}$  sob o modelo postulado e perturbado, respectivamente. O objetivo é comparar  $\widehat{\boldsymbol{\theta}}$  e  $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  quando  $\boldsymbol{\omega}$  varia em  $\Omega$ . Cook (1986) sugere que a comparação entre  $\widehat{\boldsymbol{\theta}}$  e  $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  seja feita através do afastamento pela verossimilhança  $LD(\boldsymbol{\omega})$ , definida por

$$LD(\boldsymbol{\omega}) = 2 \left[ \ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}) \right], \quad \boldsymbol{\omega} \in \Omega. \quad (3.5.8)$$

A função definida acima, pode ser vista como uma generalização do afastamento pela verossimilhança  $LD_i$ , considerado para avaliar a exclusão da  $i$ -ésima observação e definido por

$$LD_i(\boldsymbol{\theta}) = 2 \left[ \ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_{(i)}) \right], \quad (3.5.9)$$

em que  $\widehat{\boldsymbol{\theta}}_{(i)}$  é a EMV de  $\boldsymbol{\theta}$  obtido após se eliminar a  $i$ -ésima observação,  $i = 1, \dots, n$ .

O sentido da distância entre  $\widehat{\boldsymbol{\theta}}$  e  $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$ , baseado na função  $LD(\boldsymbol{\omega})$ , pode depender da concavidade da função de log-verossimilhança  $\ell(\boldsymbol{\theta})$ . Se  $\ell(\boldsymbol{\theta})$  é suficientemente achatada, pode-se dizer que  $\widehat{\boldsymbol{\theta}}$  e  $\widehat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  estão próximos entre si, enquanto que se  $\ell(\boldsymbol{\theta})$  for suficientemente concentrada em torno de  $\widehat{\boldsymbol{\theta}}$  estas estimativas podem estar distante entre si.

A ideia de Cook (1986) é investigar o comportamento da função  $LD(\boldsymbol{\omega})$  numa vizinhança de  $\boldsymbol{\omega}_0$ . Para isso considerou o gráfico de  $LD(\boldsymbol{\omega})$  versus  $\boldsymbol{\omega}$  que pode ser visto como a superfície geométrica formada pelos valores do vetor  $(q+1)$  dimensional.

$$\boldsymbol{\alpha}(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix},$$

em que  $\boldsymbol{\omega}$  varia em  $\Omega$ . A função  $\boldsymbol{\alpha}(\boldsymbol{\omega})$  é chamada de gráfico de influência local, esta é uma superfície em  $\mathbb{R}^{q+1}$  e pode ser usada para avaliar a influência ao variar  $\boldsymbol{\omega}$  através de  $\Omega$ . Uma ilustração gráfica é dada na Figura 3.1.

Aqui gostaríamos ter um gráfico da influência completa (isto é, gráfico de  $\boldsymbol{\alpha}(\boldsymbol{\omega})$  variando  $\boldsymbol{\omega}$ ) para avaliar a influência de um modelo particular e um conjunto particular de dados, no entanto, isto somente é possível nos casos em que  $q \leq 2$ . O método de influência local utiliza a curvatura

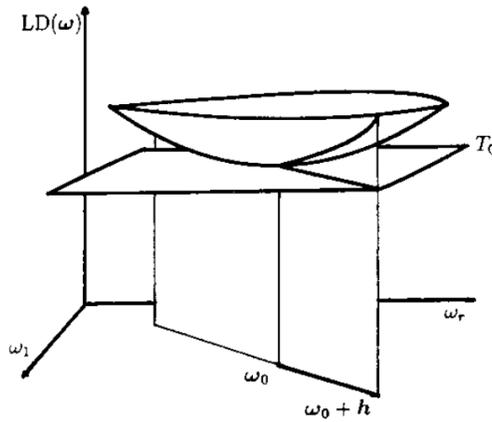


Figura 3.1: Representação gráfica do enfoque de influência local.

normal de  $\alpha(\omega)$  em  $\omega_0$  que é um ponto de mínimo local da função  $LD(\omega)$ .

A ideia principal do método de Cook (1986) baseado na superfície  $\alpha(\omega)$  em torno de  $\omega_0$  consiste em considerar o plano tangente ( $\mathbf{T}_0$ ) à superfície  $\alpha(\omega)$  em  $\omega_0$ . Como  $LD(\omega)$  atinge o mínimo em  $\omega_0$ , temos que  $\mathbf{T}_0$  é paralelo a  $\Omega \subseteq \mathbb{R}^q$ . Cada vetor  $\mathbf{h}$  em  $\Omega$ , de norma um, determina um plano que contém  $\mathbf{h}$  e que é ortogonal a  $\mathbf{T}_0$ . A intersecção, chamada de seção normal, deste plano com a superfície é chamada de linha levantada (tradução livre de “lifted line”) e pode ser obtida ao considerar o gráfico de  $LD(\omega_0 + a\mathbf{h})$  em função de  $a$ , onde  $a \in \mathbb{R}$ . A curvatura normal da linha levantada, denotada por  $C_{\mathbf{h}}$  é agora definida como a curvatura da curva plana  $(a, LD(\omega_0 + a\mathbf{h}))$  em  $a = 0$  e pode ser visualizado como a inversa do raio do círculo que melhor se ajusta em  $\omega_0$ . Valores “grandes” de  $C_{\mathbf{h}}$  indicam sensibilidade em relação à perturbação considerada na direção  $\mathbf{h}$ .

Cook (1986) mostra que a curvatura normal na direção  $\mathbf{h}$  pode ser expressa na seguinte forma:

$$C_{\mathbf{h}} = 2 \left| \mathbf{h}^{\top} \ddot{\mathbf{F}} \mathbf{h} \right|, \quad (3.5.10)$$

em que  $\|\mathbf{h}\| = 1$ ,  $\ddot{\mathbf{F}} = \mathbf{\Delta}^{\top} (-\ddot{\mathbf{L}})^{-1} \mathbf{\Delta}$ , com  $\mathbf{\Delta}$  sendo uma matriz  $(p+2) \times q$  com elementos  $\Delta_{ij} = \frac{\partial^2 \ell(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \theta_i \partial \omega_j}$ ,  $i = 1, \dots, (p+2)$ ,  $j = 1, \dots, q$  e  $-\ddot{\mathbf{L}}$  é a matriz de informação observada para o modelo postulado avaliada no ponto  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  e  $\boldsymbol{\omega} = \omega_0$ .

Existem muitas maneiras que a expressão em (3.5.10) pode ser usada para estudar  $\alpha(\omega)$ , cada uma correspondendo a uma específica escolha do vetor unitário  $\mathbf{h}$ . Seja  $C_{max} = \max_{\mathbf{h}: \|\mathbf{h}\|=1} C_{\mathbf{h}}$  que corresponde ao maior autovalor de  $-\mathbf{2F}$  (veja, Seber (1984); pag. 526) e seja  $\mathbf{h}_{max}$  o autovetor associado ao maior autovalor  $C_{max}$ . Neste caso  $\mathbf{h}_{max}$  é a direção que produz a maior mudança local nas estimativas dos parâmetros, sendo que os elementos mais influentes podem ser identificados através do “tamanho” relativo dos elementos em  $\mathbf{h}_{max}$ , enquanto que a sensibilidade em relação a perturbação causada pode ser indicada pelo valor de  $C_{max}$ .

O afastamento pela verossimilhança descrito acima é uma medida de influência da perturbação  $\omega$  na estimação de máxima verossimilhança do parâmetro  $\theta$ .

Quando o interesse for um subconjunto  $\theta_1$  de  $\theta^\top = (\theta_1^\top, \theta_2^\top)^\top$  o equivalente a (3.5.8) é dado por

$$LD_s(\omega) = 2 \left[ \ell(\hat{\theta}) - \ell(\hat{\theta}_{1\omega}, g(\hat{\theta}_{1\omega})) \right], \quad (3.5.11)$$

em que  $g$  é a função que maximiza  $\ell(\theta_1, \theta_2)$  para cada  $\theta_1$  fixo de  $\hat{\theta}_{1\omega}$ , e, é determinada pela partição de  $\hat{\theta}_\omega^\top = (\hat{\theta}_{1\omega}^\top, \hat{\theta}_{2\omega}^\top)^\top$ . Neste caso, a curvatura normal na direção  $\mathbf{h}$  é dada por

$$C_{\mathbf{h}}(\theta_1) = 2 \left[ \mathbf{h}^\top \Delta^\top (\ddot{\mathbf{L}}^{-1} - \mathbf{B}_{22}) \Delta \mathbf{h} \right], \quad (3.5.12)$$

em que  $\mathbf{B}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix}$  e  $\ddot{\mathbf{L}}_{22}$  é determinada pela partição de  $\ddot{\mathbf{L}} = \begin{pmatrix} \ddot{\mathbf{L}}_{11} & \ddot{\mathbf{L}}_{12} \\ \ddot{\mathbf{L}}_{21} & \ddot{\mathbf{L}}_{22} \end{pmatrix}$ .

Segundo Seber (1984) para qualquer vetor  $\mathbf{v}$ , pode-se mostrar que

$$0 \geq \mathbf{v}^\top \begin{pmatrix} 0 & 0 \\ 0 & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix} \mathbf{v} \geq \mathbf{v}^\top \ddot{\mathbf{L}}^{-1} \mathbf{v},$$

e, portanto,

$$C_{\mathbf{h}}(\theta_1) = -2\mathbf{h}^\top \Delta^\top \ddot{\mathbf{L}}^{-1} \Delta \mathbf{h} + 2\mathbf{h}^\top \Delta^\top \begin{pmatrix} 0 & 0 \\ 0 & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix} \Delta \mathbf{h} \quad (3.5.13)$$

$$= C_{\mathbf{h}} + 2\mathbf{h}^\top \Delta^\top \begin{pmatrix} 0 & 0 \\ 0 & \ddot{\mathbf{L}}_{22}^{-1} \end{pmatrix} \Delta \mathbf{h} \quad (3.5.14)$$

$$\leq C_{\mathbf{h}}. \quad (3.5.15)$$

Isto significa que a curvatura normal para  $\theta_1$  na direção  $\mathbf{h}$ , nunca pode ser maior que a curvatura normal para  $\theta$  na mesma direção.

Uma outra escolha evidente corresponde à perturbação na direção do  $i$ -ésimo caso que é obtida considerando à  $\mathbf{h}$  igual ao vetor  $\mathbf{h}_i$  que tem um na  $i$ -ésima posição e zero nas outras posições, dando como resultado a seguinte medida de influência local

$$C_i = C_{h_i} = 2 \left| \Delta_i \left( -\ddot{\mathbf{L}} \right)^{-1} \Delta_i \right|, \quad i = 1, \dots, q, \quad (3.5.16)$$

em que  $\Delta_i$  é a  $i$ -ésima coluna de  $\Delta$ . Verveke & Molenberghs (1997) sugere considerar a  $i$ -observação como influente se  $C_i$  é maior que o valor  $2 \sum_{i=1}^q C_i / q$ . Note que  $C_i$  é de interesse particular desde que representa a influência local de cada observação separadamente na estimação de  $\theta$ .

Finalmente, existem muitas formas possíveis de escolher o vetor  $\mathbf{h}$ . Por exemplo, a influência local na estimação de  $\boldsymbol{\theta}_1$  na direção do  $i$ -ésimo caso. A correspondente curvatura normal na direção  $\mathbf{h}_i$ , será denotada por  $C_i(\boldsymbol{\theta}_1)$ .

Embora, o método de influência local proposto por Cook (1986) tem sido aplicado com sucesso em muitos problemas estatísticos, alguns aspectos tem sido questionados por alguns autores (veja, discussão em Cook (1986)), por exemplo, a falta de um critério objetivo para avaliar a grandeza da curvatura normal e tamanho relativo das componentes da direção correspondente à maior curvatura normal.

Poon & Poon (1999) apresentaram um método de influência local similar ao método proposto por Cook (1986), que no lugar de usar curvatura normal recomenda o uso da curvatura normal conformalizada, sendo esta invariante sob reparametrização e assume valores entre 0 e 1.

Segundo Poon & Poon (1999), a curvatura normal conforme em  $\boldsymbol{\omega}_0$  do gráfico  $\boldsymbol{\alpha}(\boldsymbol{\omega})$  na direção  $\mathbf{h}$  é definida por

$$B_{\mathbf{h}} = - \frac{\mathbf{h}^\top \boldsymbol{\Delta}^\top (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta} \mathbf{h}}{\sqrt{\text{tr} \left( \boldsymbol{\Delta}^\top (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta} \right)^2}}, \quad (3.5.17)$$

avaliadas em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  e  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ . De (3.5.10) e (3.5.17) podemos observar que o trabalho para obter  $\mathbf{C}_{\mathbf{h}}$  e  $\mathbf{B}_{\mathbf{h}}$  são os mesmos, mas  $B_{\mathbf{h}}$  possui várias propriedades importantes, que podem ser encontradas em Poon & Poon (1999).

De acordo a Poon & Poon (1999), curvatura normal conformalizada total em  $\boldsymbol{\omega}_0$  da superfície  $\boldsymbol{\alpha}(\boldsymbol{\omega})$  na direção  $\mathbf{h}_i$  pode ser expressa como

$$B_{\mathbf{h}_i} = - \frac{\boldsymbol{\Delta}_i^\top (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta}_i}{\sqrt{\text{tr} \left( \boldsymbol{\Delta}^\top (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta} \right)^2}}, \quad (3.5.18)$$

em que  $\mathbf{h}_i$  e  $\boldsymbol{\Delta}_i$  são como em (3.5.16),  $i = 1, \dots, q$ . Se  $\lambda_1 \geq \dots \geq \lambda_k > 0$  denota os  $k$  autovalores diferentes de zero de

$$- \frac{\boldsymbol{\Delta}^\top (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta}}{\sqrt{\text{tr} \left( \boldsymbol{\Delta}^\top (\ddot{\mathbf{L}})^{-1} \boldsymbol{\Delta} \right)^2}}, \quad (3.5.19)$$

e se  $\mathbf{a}_1, \dots, \mathbf{a}_k$  são os correspondentes autovalores, então pode-se verificar que

$$B_{\mathbf{h}_i} = \sum_{j=1}^k \lambda_j a_{ji}^2, \quad (3.5.20)$$

em que  $a_{ji}$  é o  $i$ -ésimo componente do vetor  $\mathbf{a}_j$ ,  $i = 1, \dots, q$ . Isto mostra que casos individuais possa haver uma grande curvatura local total  $B_{\mathbf{h}_i}$ , sem ter uma grande componente  $|a_{1i}|$  na direção  $\mathbf{h}_{max}$ . De acordo a Poon & Poon (1999), a contribuição conjunta da  $j$ -ésima base de vetores de perturbações para todo autovetor  $r$ -influyente (o autovetor  $\mathbf{b}$  é  $r$ -influyente, se  $|B_{\mathbf{b}}| \geq r/\sqrt{q}$ ) é definido como

$$m(r)_j = \sum_{i=1}^s \lambda_i a_{ji}^2, \quad (3.5.21)$$

em que  $s$  é o número de autovalores maiores a  $r/\sqrt{q}$ ,  $j = 1, \dots, q$  e uma adequada marca de nível para julgar a grandeza da influência local é  $2\bar{m}(r)$  para  $r > 0$  e  $2b$  para  $r = 0$ , onde  $\bar{m}(r) = \sum_{i=1}^s \lambda_i/q$  e  $b = tr(\mathbf{A}) / (q\sqrt{tr(\mathbf{A}^2)})$ , com  $\mathbf{A} = -\mathbf{\Delta}^\top (\ddot{\mathbf{L}})^{-1} \mathbf{\Delta}$ . Assim, a avaliação de casos influentes é baseada em  $\{m(r)_l, l = 1, \dots, q\}$ . Zhu & Lee (2001) propõe como ponto de corte o valor  $\bar{m}(r) + c^* DP(m(r))$ , onde  $DP(m(r))$  é o desvio padrão de  $\{m(r)_l, l = 1, \dots, q\}$  e  $c^*$  é uma constante arbitrária. Neste trabalho consideramos  $c^* = 1.5$ .

Dos argumentos precedentes, encontramos que se  $s = 1$  (ou  $r$  suficientemente grande) as observações influentes identificadas são as mesmas que as encontradas por  $\mathbf{h}_{max}$ . Por outro lado  $r = 0$  identifica as mesmas observações que a medida de influência local  $C_i$ , isto é, os gráficos resultantes são proporcionais (veja, ilustração numérica) com a diferença que esta nova metodologia é mas objetiva, assim Poon & Poon (1999) propõe um método de diagnóstico unificado para avaliar a influência local de pequenas perturbações no conjunto de dados.

### 3.5.3 Esquemas de perturbação

Nesta seção, vamos avaliar a matriz  $\mathbf{\Delta}$  sob três esquemas de perturbação para o modelo definido em (3.2.1), a saber: (i) perturbação de ponderação de casos, (ii) perturbação na variável explicativa, e (iii) perturbação na variável resposta. Obtemos as matrizes apropriadas para avaliar a influência local para os esquemas considerados. Para cada esquema de perturbação, tem-se a forma particionada  $\mathbf{\Delta}_{\omega_o} = (\mathbf{\Delta}_\beta^\top, \mathbf{\Delta}_\alpha^\top, \mathbf{\Delta}_\sigma^\top)^\top$ , em que

$$\begin{aligned} \mathbf{\Delta}_\beta &= \frac{\partial^2 \ell(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\omega}_o, \hat{\boldsymbol{\theta}}} \in \mathbb{R}^{p \times n}, \\ \mathbf{\Delta}_\alpha &= \frac{\partial^2 \ell(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \alpha \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\omega}_o, \hat{\boldsymbol{\theta}}} \in \mathbb{R}^{1 \times n}, \\ \mathbf{\Delta}_\sigma &= \frac{\partial^2 \ell(\boldsymbol{\theta}, \boldsymbol{\omega})}{\partial \sigma \partial \boldsymbol{\omega}^\top} \Big|_{\boldsymbol{\omega}_o, \hat{\boldsymbol{\theta}}} \in \mathbb{R}^{1 \times n}, \end{aligned}$$

e o vetor de perturbação é dado por  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$ .

### Perturbação de ponderação de casos

Considerando uma atribuição arbitrária de ponderações (pesos) para as unidades experimentais, o logaritmo da função de verossimilhança pode ser representada por

$$\ell(\boldsymbol{\theta}, \boldsymbol{\omega}) = \sum_{i=1}^n \omega_i \ell_i(\boldsymbol{\theta}),$$

em que  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_n)^\top$  tem ordem  $n \times 1$ . Neste caso, o modelo não perturbado é obtido quando  $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$ . Para este esquema de perturbação, encontramos

$$\begin{aligned} \Delta_\beta &= \frac{1}{\sigma} \mathbf{D}_\beta [D(\boldsymbol{\xi}_1)D(\boldsymbol{\xi}_2) - D(\boldsymbol{\xi}_2)D^{-1}(\boldsymbol{\xi}_1)], \\ \Delta_\alpha &= -\frac{1}{\alpha} \mathbf{1}_n^\top + \frac{1}{\alpha} \mathbf{1}_n^\top D^2(\boldsymbol{\xi}_2), \\ \Delta_\sigma &= -\frac{1}{\sigma} \mathbf{1}_n^\top + \frac{1}{\sigma} \boldsymbol{\epsilon}^\top [D(\boldsymbol{\xi}_1)D(\boldsymbol{\xi}_2) - D(\boldsymbol{\xi}_2)D^{-1}(\boldsymbol{\xi}_1)]. \end{aligned}$$

### Perturbação na variável explicativa

Neste caso, estamos interessados em perturbar uma variável explicativa contínua específica. Sem perda de generalidade, assumimos que esta variável é  $x_p$ . A perturbação aditiva é obtida por  $x_{\omega_{ip}} = x_{ip} + \omega_i S_x$ , para  $i = 1, \dots, n$ , em que  $S_x$  é um fator de escala que pode ser o desvio padrão dos valores  $x_p$ . Assim, a função de log-verossimilhança perturbada é dada por  $\ell(\boldsymbol{\theta}, \boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}, \omega_i)$ , em que

$$\ell_i(\boldsymbol{\theta}, \omega_i) = c - \log(\sigma) - \frac{1}{2} \xi_{\omega_i 2}^2 + \log(\xi_{\omega_i 1}),$$

com  $\xi_{\omega_i 1}$  e  $\xi_{\omega_i 2}$  analogamente definido como em (3.2.2), com  $\mu_{i\omega} = g^{-1}(\mathbf{x}_{i\omega}^\top \boldsymbol{\beta}) = g_0(\mathbf{x}_{i\omega}^\top \boldsymbol{\beta})$  em lugar de  $\mu_i = g^{-1}(\mathbf{x}_i^\top \boldsymbol{\beta}) = g_0(\mathbf{x}_i^\top \boldsymbol{\beta})$ ,  $x_{\omega_{ip}}$  em vez de  $x_{ip}$  e  $\boldsymbol{\omega}_0 = \mathbf{0}$ . Definindo  $\boldsymbol{\mu}_\omega = (\mu_{1\omega}, \dots, \mu_{n\omega})^\top$ , temos que os elementos da matriz  $\Delta_{\boldsymbol{\omega}_0}$  são dados por

$$\begin{aligned} \Delta_\beta &= \frac{1}{\sigma} \left[ -\frac{1}{\sigma} \mathbf{D}_\beta \mathbf{A}_{12}(\hat{\mathbf{u}}) D(\dot{\boldsymbol{\mu}}_\omega) + \mathbf{D}_{\omega\beta} [D(\boldsymbol{\xi}_1)D(\boldsymbol{\xi}_2) - D(\boldsymbol{\xi}_2)D^{-1}(\boldsymbol{\xi}_1)] \right], \\ \Delta_\alpha &= -\frac{2}{\alpha\sigma} \dot{\boldsymbol{\mu}}_\omega^\top D(\boldsymbol{\xi}_1)D(\boldsymbol{\xi}_2), \\ \Delta_\sigma &= -\frac{1}{\sigma^2} [\boldsymbol{\epsilon}^\top \mathbf{A}_{12} + \boldsymbol{\xi}_2^\top [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)]] D(\dot{\boldsymbol{\mu}}_\omega), \end{aligned}$$

em que  $\dot{\boldsymbol{\mu}}_\omega = \left( \frac{\partial \mu_{1\omega}}{\partial \omega_1}, \dots, \frac{\partial \mu_{n\omega}}{\partial \omega_n} \right)^\top$ ,  $\mathbf{D}_\beta = \frac{\partial \boldsymbol{\mu}_\omega^\top}{\partial \boldsymbol{\beta}}$  e  $\mathbf{D}_{\beta\omega} = \frac{\partial \dot{\boldsymbol{\mu}}_\omega^\top}{\partial \boldsymbol{\beta}}$  todos avaliados em  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$  e em  $\hat{\boldsymbol{\theta}}$ .

### Perturbação na variável resposta

Neste caso estamos interessados em perturbar as variáveis de resposta  $y_1, \dots, y_n$  dado por  $y_{\omega_i} = y_i + S_y \omega_i$ ,  $i = 1, \dots, n$ , em que  $S_y$  é um fator de escala que pode ser o desvio padrão dos valores de  $y$ . A função de log-verossimilhança perturbada é dada por

$$\ell(\boldsymbol{\theta}, \boldsymbol{\omega}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta}, \omega_i),$$

em que  $\ell_i(\boldsymbol{\theta}, \omega_i) = c - \frac{1}{2}\xi_{\omega_i 2}^2 + \log(\xi_{\omega_i 1})$ , com  $\xi_{\omega_i 1}$  e  $\xi_{\omega_i 2}$  são dadas como em perturbações de uma variável explicativa, mas neste caso com  $y_{\omega_i}$  em lugar de  $y_i$  e  $\boldsymbol{\omega}_0 = \mathbf{0}$ . Os elementos da matriz  $\boldsymbol{\Delta}_{\boldsymbol{\omega}}$  são dados por

$$\begin{aligned} \boldsymbol{\Delta}_{\beta} &= \frac{S_y}{\sigma^2} \mathbf{D}_{\beta} \mathbf{A}_{12}, \\ \boldsymbol{\Delta}_{\alpha} &= \frac{2S_y}{\alpha\sigma} \mathbf{1}_n^{\top} D(\boldsymbol{\xi}_1) D(\boldsymbol{\xi}_2), \\ \boldsymbol{\Delta}_{\sigma} &= \frac{S_y}{\sigma^2} [\boldsymbol{\epsilon}^{\top} \mathbf{A}_{12} + \boldsymbol{\xi}_2^{\top} [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)]] . \end{aligned}$$

## 3.6 Estudo de simulação

Para demonstrar a robustez do algoritmo de Newton Raphson no sentido de estimação, foram simulados dados com a presença de uma variável regressora, da forma  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ , onde  $\epsilon_i \sim SN(\alpha, 0, \sigma)$ ,  $i = 1, \dots, 50$  e a variável  $x$  foi gerada da distribuição  $U(0, 1)$ . Os valores escolhidos dos parâmetros foram  $\beta_0 = 0.5$ ,  $\beta_1 = -2$ ,  $\alpha = 0.5$  e  $\sigma = 1.5$ . A Tabela (3.1) apresenta os valores escolhidos e estimados dos parâmetros do modelo. Os erros padrões são calculados utilizando a matriz de informação de Fisher observada, na forma  $EP(\hat{\boldsymbol{\theta}}) = \left[ \text{Diag} \left( I^{-1}(\hat{\boldsymbol{\theta}}) \right) \right]^{\frac{1}{2}}$ .

Tabela 3.1: Estimadores de máxima verossimilhança com seus respectivos erros padrões.

Parâmetros	Valores verdadeiros	Estimativas	Erros padrão
$\beta_1$	0,5	0,36	0,13
$\beta_2$	-2	-1,83	0,20
$\alpha$	0,5	0,44	0,83
$\sigma$	1,5	1,49	1,06

A seguir, será apresentado o mesmo estudo de simulação para estudar o comportamento do esquema de perturbação em uma particular variável explicativa. Seguindo o mesmo procedimento

sugerido por Ortega *et al.* (2003), utilizando os valores iniciais detalhados acima, substituímos o valor máximo de  $x_i$  por

$$x_{max} \leftarrow x_{max} + 0.2\sqrt{\|\mathbf{x}\|}, \tag{3.6.1}$$

em que  $\|\mathbf{x}\| = \mathbf{x}^\top \mathbf{x}$ . Então ajustamos o modelo de regressão SN e aplicamos o esquema de perturbação na variável explicativa. O gráfico de influência local e global é mostrado na Figura 3.2, respectivamente. Observamos que a observação 11 (perturbada através de (3.6.1)) exerce grande influência, confirmando a sensibilidade da metodologia proposta.

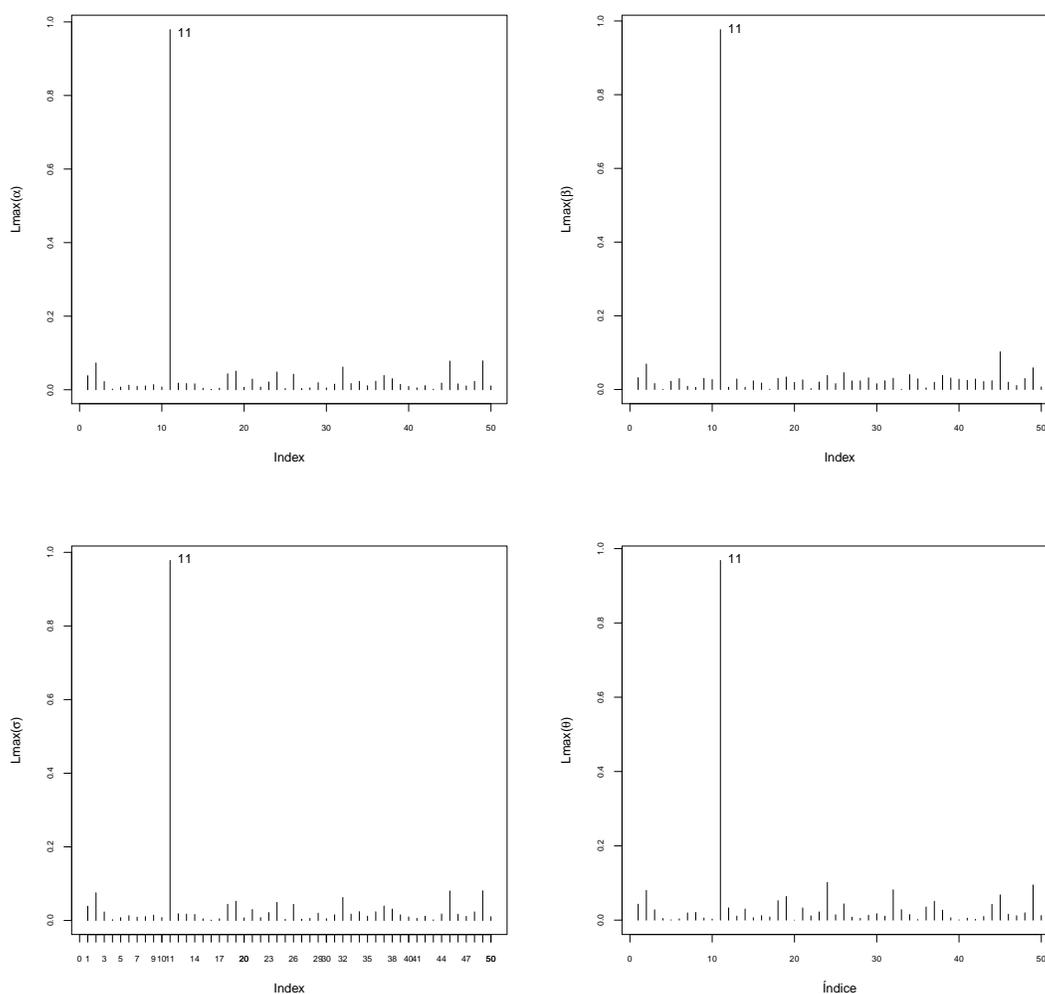


Figura 3.2: Gráficos para a perturbação na variável explicativa.

### 3.7 Aplicação 1: Fadiga biaxial

Para ilustrar a aplicação do modelo de regressão SN e a análise de diagnóstico, foi ajustado um modelo a um conjunto de dados previamente analisados em Rieck & Nedelman (1991), Galea *et al.* (2004) e Xie & Wei (2007), que correspondem a 46 peças de metal que foram submetidas à fadiga em ciclos. O objetivo do estudo foi verificar o número de ciclos até que ocorresse a falha. A variável resposta  $N$  é o número de ciclos até que ocorra a falha e a variável explicativa  $z$  é o trabalho por ciclo ( $mJ/m^3$ ). Os dados são apresentados no Anexo I, Tabela I.1. Este conjunto de dados foi analisado por Rieck & Nedelman (1991). Neste trabalho, vamos considerar o seguinte modelo de regressão

$$y_i = \mu_i + \epsilon_i,$$

em que  $y_i = \log N_i$ ,  $x_i = \log(z_i)$ ,  $\mu_i = g_0(\mathbf{x}_i^\top \boldsymbol{\beta})$  e  $\epsilon_i \sim SN(\alpha, 0, \sigma)$ ,  $i = 1, \dots, 46$ , para um modelo linear e vamos considerar  $x_i^2$  para um modelo quadrático.

#### 3.7.1 Ajuste do modelo

Com o intuito de ilustrar a metodologia proposta, vamos utilizar as seguintes funções de ligação: identidade ( $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ ), recíproca ( $\mu_i = (\mathbf{x}_i^\top \boldsymbol{\beta})^{-1}$ ) e logarítmica ( $\mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$ ) para um modelo linear e um modelo quadrático.

Na Tabela 3.2 são apresentados as EMV com seus respectivos erros padrões, que foram obtidos utilizando a matriz de informação de Fisher observada na Seção 3.6. De acordo com os critérios AIC e BIC, apresentados na Tabela 3.3, o modelo linear com função de ligação logarítmica apresenta melhor ajuste o qual também é avaliado pelos erros padrões das estimativas do mesmo.

Tabela 3.2: Estimativas de máxima verossimilhança para a distribuição SN, com seus respectivos erros padrões, para os dados de fadiga biaxial.

Preditor	Função de ligação	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\sigma$
Linear	Identidade	12,2229	-1,6641		2,2676	0,4665
		(0,2632)	(0,0756)		(1,0622)	(0,1385)
Linear	Logarítmica	2,7681	-0,2649		3,1920	0,3705
		(0,0369)	(0,0116)		(1,4167)	(0,0897)
Linear	Recíproca	0,0181	0,0406		3,6587	0,3476
		(0,0057)	(0,0019)		(1,5770)	(0,0772)
Quadrático	Identidade	15,9425	-3,8577	0,3128	3,3337	0,3610
		(1,7733)	(1,0545)	(0,1530)	(1,4749)	(0,0853)
Quadrático	Logarítmica	1,3549	-0,0245	0,0003	0,5561	0,8729
		(0,2767)	(0,1668)	(0,0245)	(1,5331)	(0,0835)
Quadrático	Recíproca	0,0519	0,0197	0,0032	3,4443	0,3555
		(0,0439)	(0,0268)	(0,0040)	(1,5568)	(0,0843)

Tabela 3.3: Critérios para seleção de modelos para o conjunto de dados da fadiga biaxial.

Preditor	Função de ligação	AIC	BIC
Linear	Identidade	51,2345	58,5490
Linear	Logarítmica	47,6021	54,9166
Linear	Recíproca	47,8473	55,1619
Quadrático	Identidade	49,1839	58,3271
Quadrático	Logarítmica	49,2259	58,3691
Quadrático	Recíproca	49,2328	58,3760

Nas Figuras 3.3, 3.4 e 3.5 temos os diagramas de dispersão entre o  $\log(\text{trabalho})$  e  $\log(\text{ciclos})$  com o valor médio previsto do  $\log(\text{ciclos})$ . Observe que, de acordo com linha ajustada os modelos lineares apresentam melhor ajuste, sendo melhores para aqueles com função de ligação logarítmica e recíproca em ordem de prioridade.

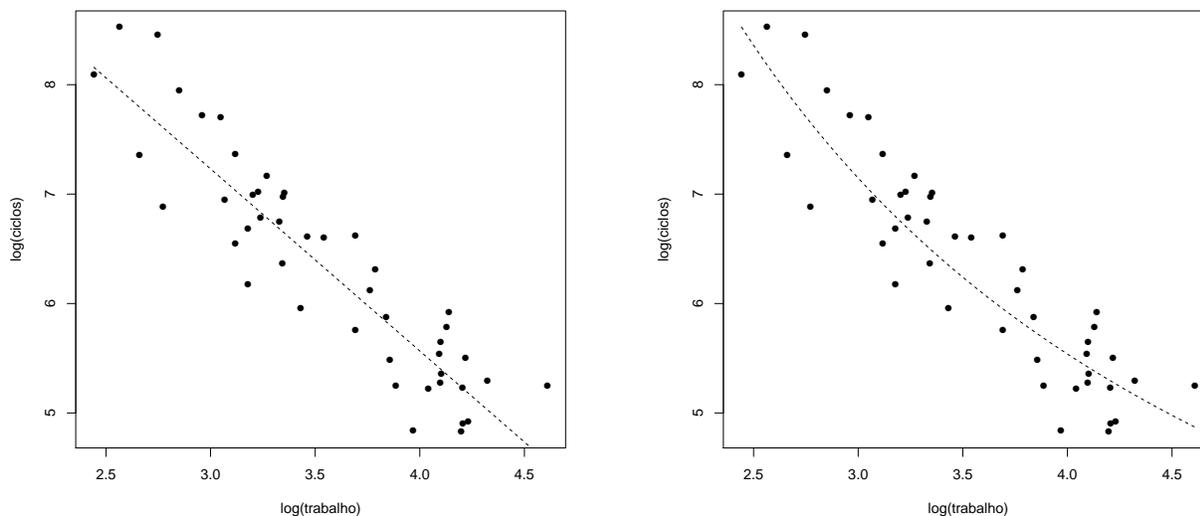


Figura 3.3: Diagrama de dispersão entre  $\log(\text{trabalho})$  e  $\log(\text{ciclos})$  com o valor médio previsto do  $\log(\text{ciclos})$  para conjunto de dados de fadiga. À esquerda usando preditor linear e ligação identidade e à direita preditor linear e ligação recíproca.

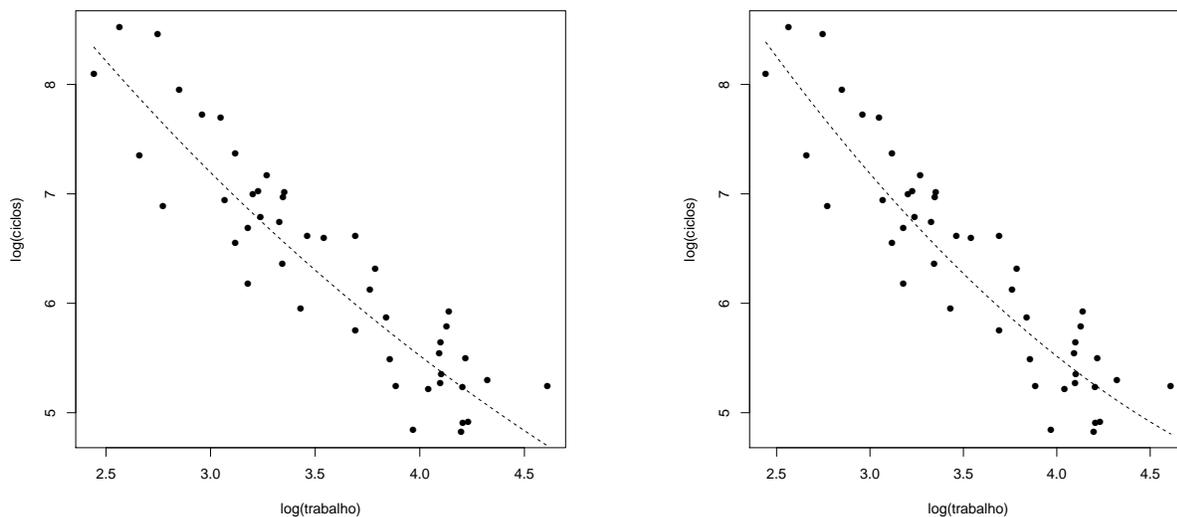


Figura 3.4: Diagrama de dispersão entre  $\log(\text{trabalho})$  e  $\log(\text{ciclos})$  com o valor médio previsto do  $\log(\text{ciclos})$  para conjunto de dados de fadiga. À esquerda usando preditor linear e ligação logarítmica e à direita preditor quadrático e ligação identidade.

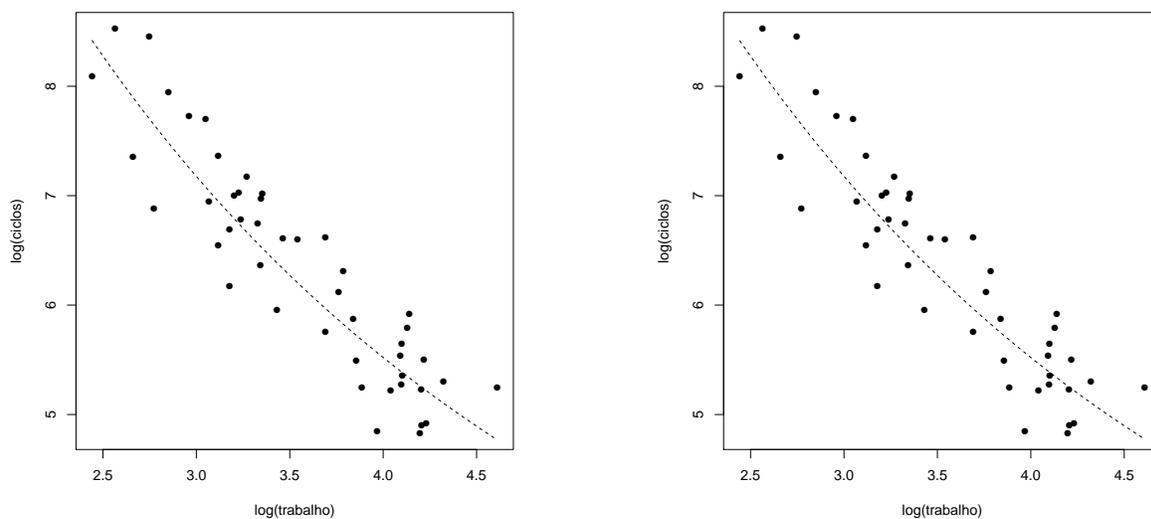


Figura 3.5: Diagrama de dispersão entre  $\log(\text{trabalho})$  e  $\log(\text{ciclos})$  com o valor médio previsto do  $\log(\text{ciclos})$  para conjunto de dados de fadiga. À esquerda usando preditor quadrático e ligação recíproca e à direita preditor quadrático e ligação logarítmica.

Nas Figuras 3.6, 3.7 e 3.8 temos as bandas de confiança de 95% para o valor da média prevista de  $\log(\text{ciclos})$ . Observe que, os modelos lineares com ligações logarítmica e recíproca têm bandas com intervalos de confiança com menor comprimento em comparação com os modelos quadráticos, sendo menor para o modelo linear com ligação logarítmica, o qual é avaliado pelos critérios AIC e BIC.

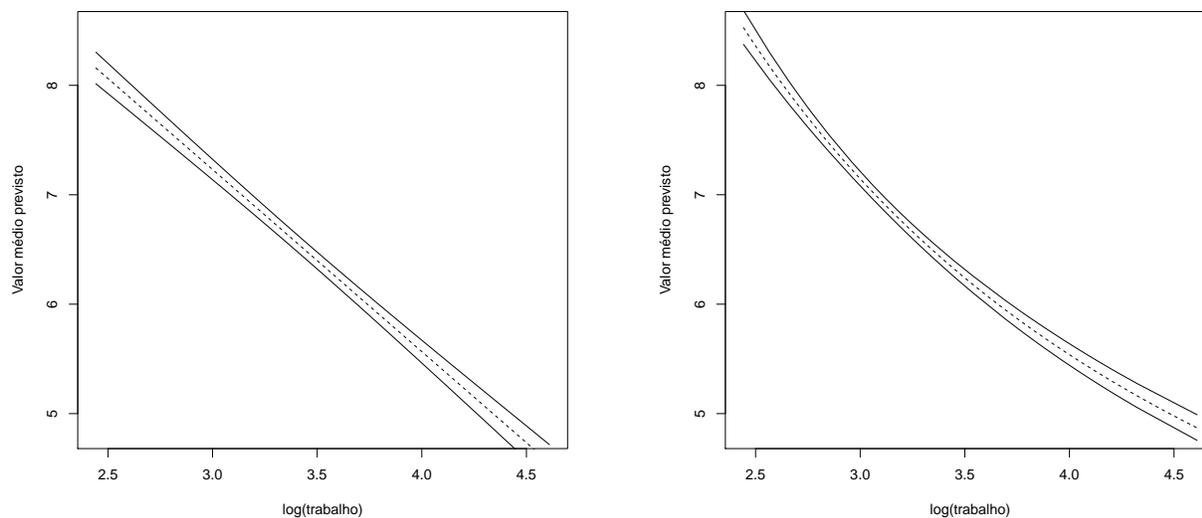


Figura 3.6: Banda de Confiança de 95% para o valor da média prevista de  $\log(\text{ciclos})$  para conjunto de dados de fadiga. À esquerda usando preditor linear e ligação identidade e à direita preditor linear e ligação recíproca.

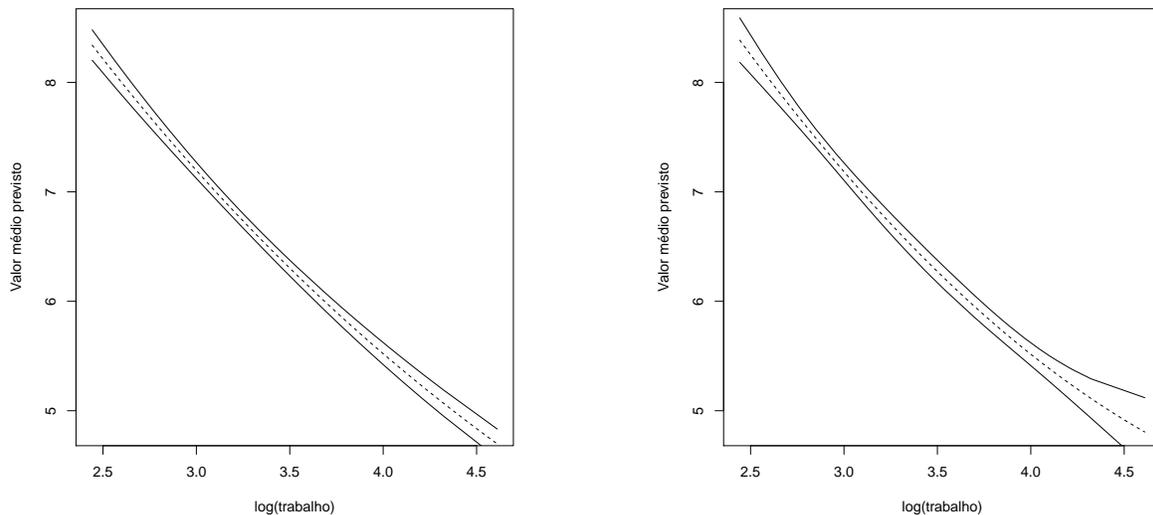


Figura 3.7: Banda de Confiança de 95% para o valor da média prevista de  $\log(\text{ciclos})$  para conjunto de dados de fadiga. À esquerda usando preditor linear e ligação logarítmica e à direita preditor quadrático e ligação identidade.

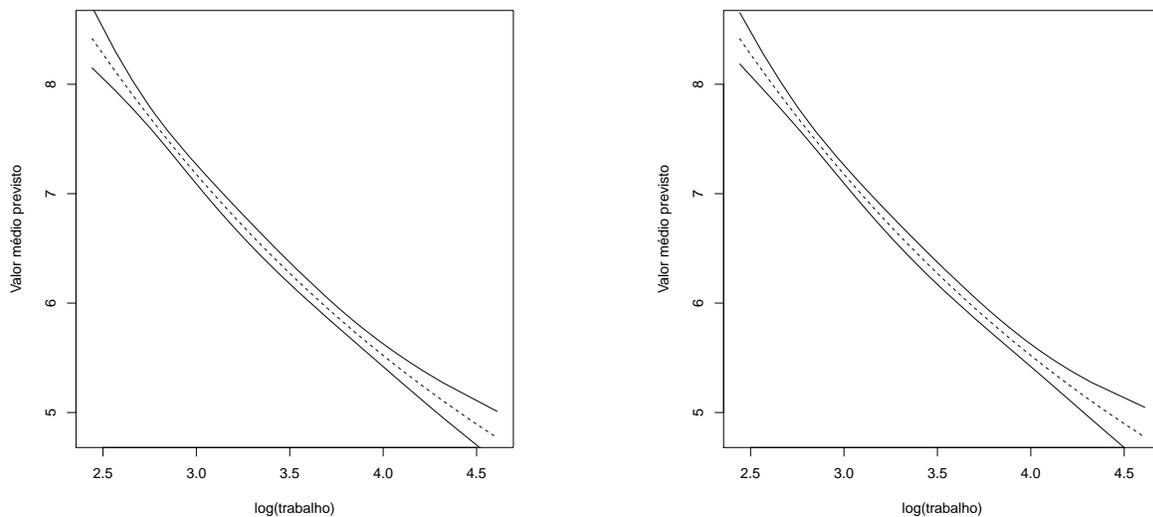


Figura 3.8: Banda de Confiança de 95% para o valor da média prevista de  $\log(\text{ciclos})$  para conjunto de dados de fadiga. À esquerda usando quadrático e ligação recíproca e à direita preditor quadrático e ligação logarítmica.

De acordo com o análise anterior, escolhemos o modelo linear com ligação logarítmica. Isto é

$$Y_i = \mu_i + \epsilon_i, \quad i = 1, \dots, 46,$$

em que  $\mu_i = \exp(\mathbf{x}_i^\top \boldsymbol{\beta})$ .

Observe que, podemos expressar  $\mu_i$  seguinte forma

$$\begin{aligned} \mu_i &= \exp(\beta_0 + \beta_1 \log(z_i)) \\ &= \exp(\beta_0) [\exp(\log(z_i))]^{\beta_1} \\ &= \gamma z_i^\beta \end{aligned}$$

Portanto, o modelo pode-se expressar da seguinte forma

$$Y_i = \gamma z_i^\beta + \epsilon_i, \quad i = 1, \dots, 46. \quad (3.7.1)$$

Este modelo é o mesmo ao obtido em Lemonte & Cordeiro (2010), em que o conjunto de dados tem melhor ajuste para um modelo não linear.

Antes de realizar a análise de diagnóstico, vamos avaliar o ajuste deste modelo ao conjunto de dados e fazer a análise de resíduos.

Na Figura 3.9 temos os gráficos de probabilidade normal com envelopes simulados para os resíduos apresentados na Subseção 3.4.2 e Na Figura 3.10 o envelope simulado. Observe que, estes gráficos fornecem evidencia, junto aos critérios mostrados anteriormente, de que o modelo proporciona um bom ajuste aos dados.

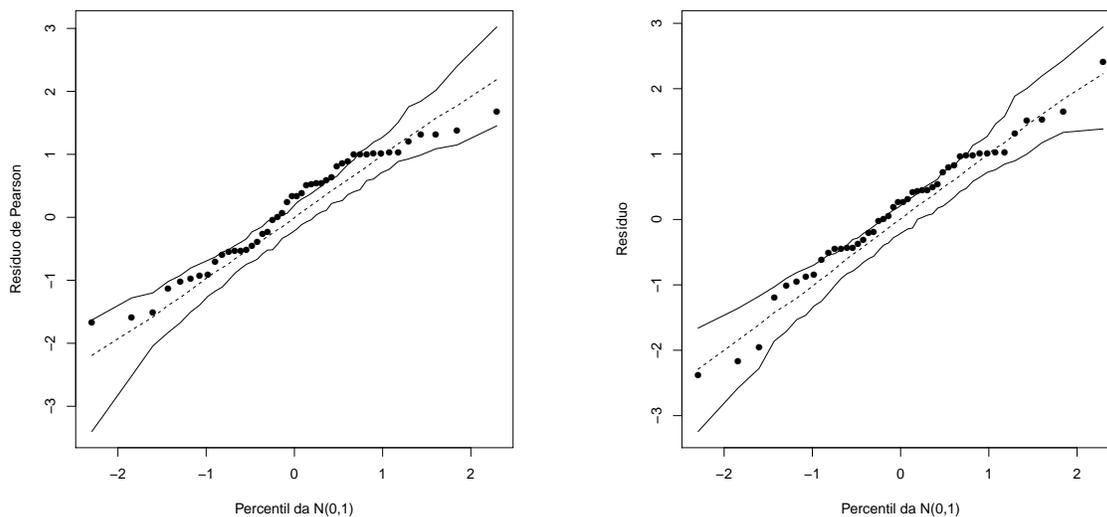


Figura 3.9: Gráfico de probabilidade normal com envelope simulado para o ajuste do modelo com preditor linear e ligação logarítmica para os dados de fadiga. À esquerda usando resíduo de Pearson e à direita usando outro resíduo.

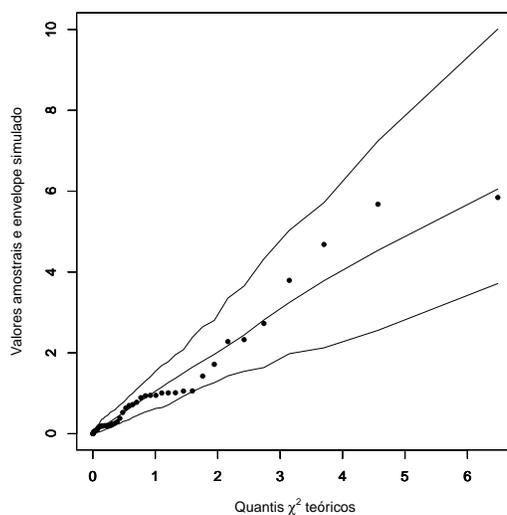


Figura 3.10: Envelope simulado para os dados de fadiga.

Na Figura 3.11 apresentamos os gráficos para os resíduos que foram abordados na Subseção 3.4.2. Observe que, as observações 4, 5, 12 e 32 são possíveis outliers.

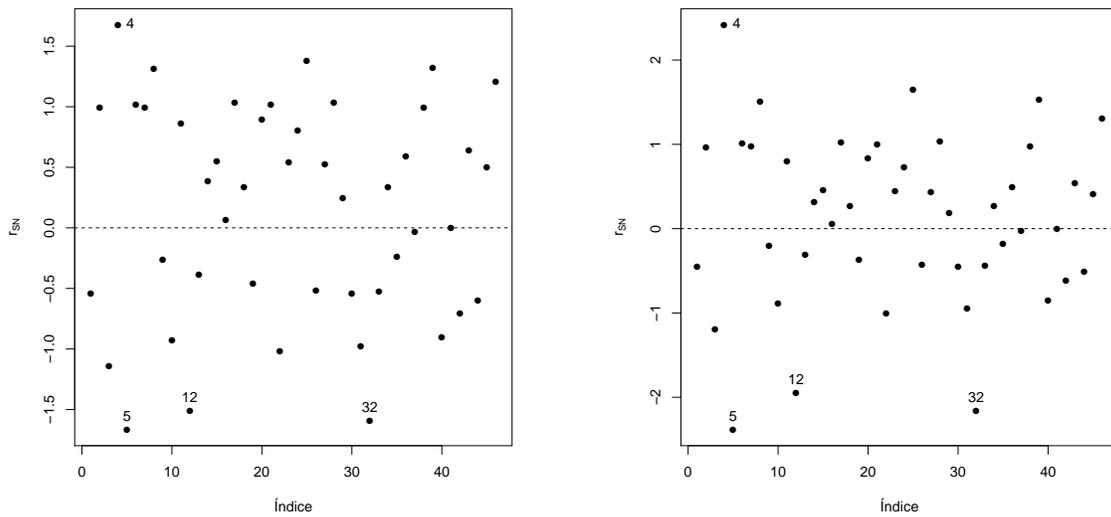


Figura 3.11: Gráfico de resíduos para os dados de fadiga. À esquerda usando resíduo de Pearson e à direita usando outro resíduo.

### 3.7.2 Análise de diagnóstico

Baseado no modelo linear com ligação logarítmica, apresentamos os resultados da análise de diagnóstico.

#### Modelo de eliminação de casos

Nas Figuras 3.12 e 3.13 apresentamos a distância de Cook ( $GD_i^1$ ) e na Figura 3.14 temos o afastamento pela log-verossimilhança ( $LD_i^1$ ), definidas na Subseção 3.5.1. Os resultados mostrados indicam que as observações 4, 5, 12 e 32 são identificadas como influentes nas estimações sob o parâmetro global  $\theta$  e também para a estimação dos parâmetros  $\alpha$ ,  $\beta$  e  $\sigma$ , em que as observações 4 e 5 destacam-se como as mais influentes.

Observe que, no caso do afastamento pela log-verossimilhança ( $LD_i^1$ ) temos um padrão similar (veja a Figura 3.14), como também, que as observações 4, 5, 12 e 32 são identificadas como influentes no modelo Sinh-Normal. Entre elas, as observações 4 e 5 destacam-se como as mais influentes.

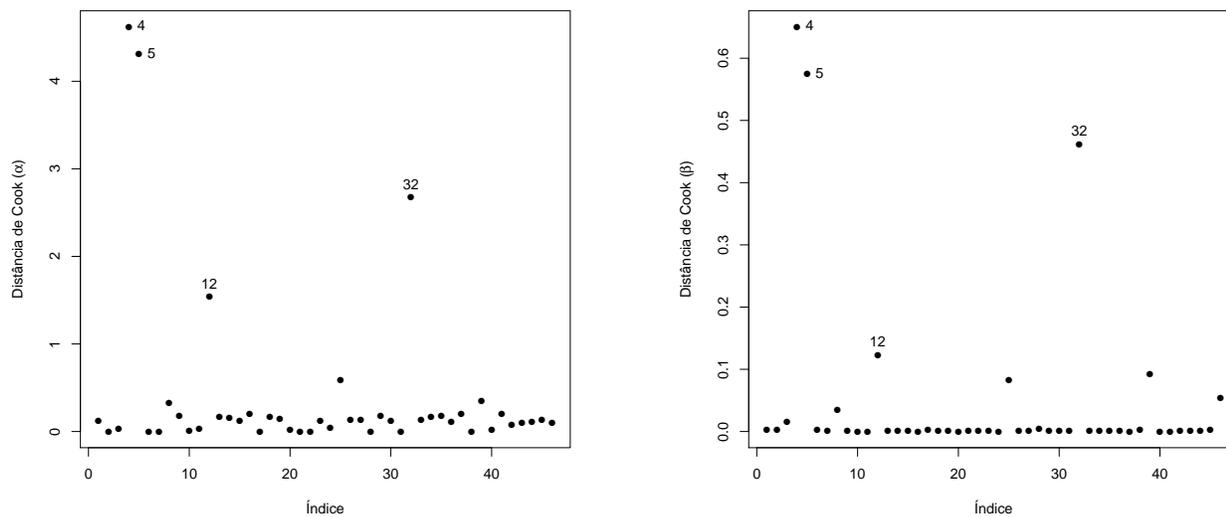


Figura 3.12: Distância Generalizada de Cook para os parâmetros  $\alpha$  e  $\beta$  para o conjunto de dados de fadiga.

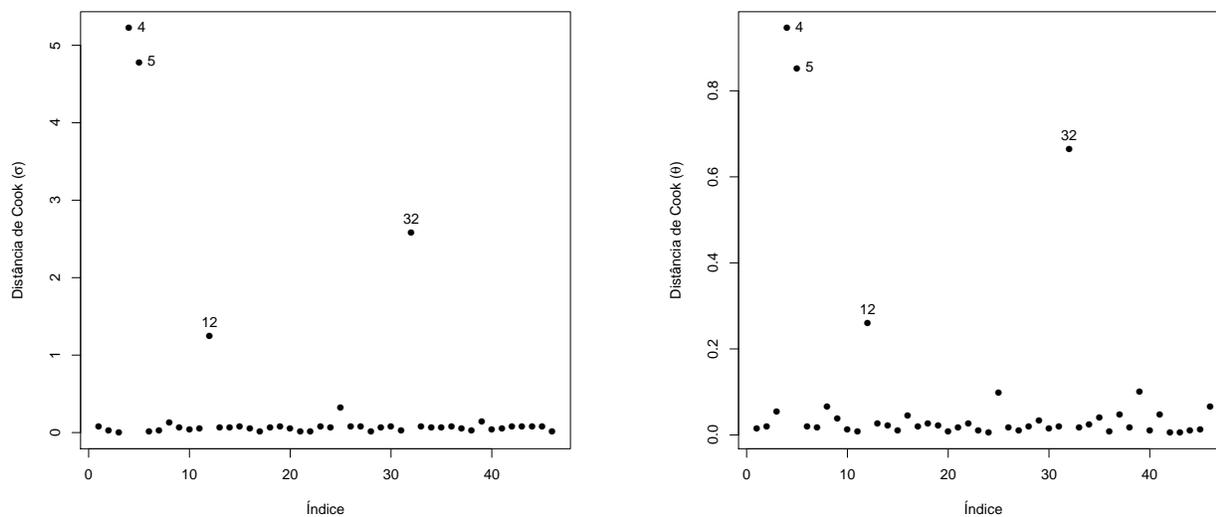


Figura 3.13: Distância Generalizada de Cook para o parâmetro  $\sigma$  e para  $\theta$  para o conjunto de dados de fadiga.

A seguir apresentamos um estudo de influência local baseado no esquema de perturbação de casos

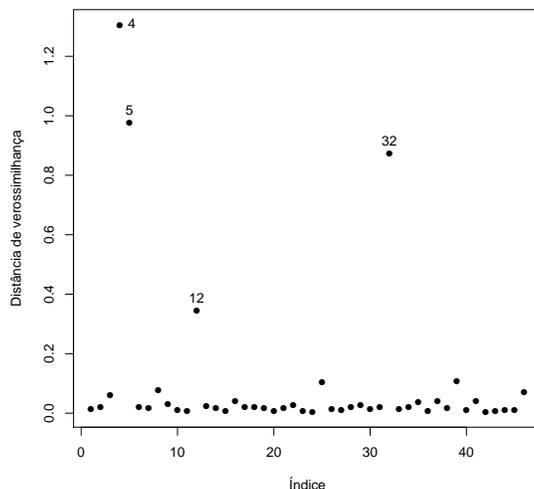


Figura 3.14: Distância de verossimilhança para conjunto de dados de fadiga.

### Perturbação de ponderação de casos

Sob o esquema de perturbação de ponderação de casos, na Figura 3.15 temos o gráfico  $M(0)$  para o parâmetro global e também para cada parâmetro, observe que de acordo com a marca de nível, indicada no gráfico por uma linha pontilhada, as observações 4, 5 e 32 são identificadas como influentes para todos os parâmetros.

Sob o esquema de perturbação de ponderação de casos, na Figura 3.16 temos o gráfico de influência total  $C_i$  vs observações e o gráfico das médias de influência local  $C_i(\boldsymbol{\alpha})$  vs  $C_i(\boldsymbol{\beta})$ . Observe que, de acordo com a marca de nível, indicada no gráfico por uma linha pontilhada, as observações 4, 5, 12 e 32 são identificadas como influentes em ambos gráficos.

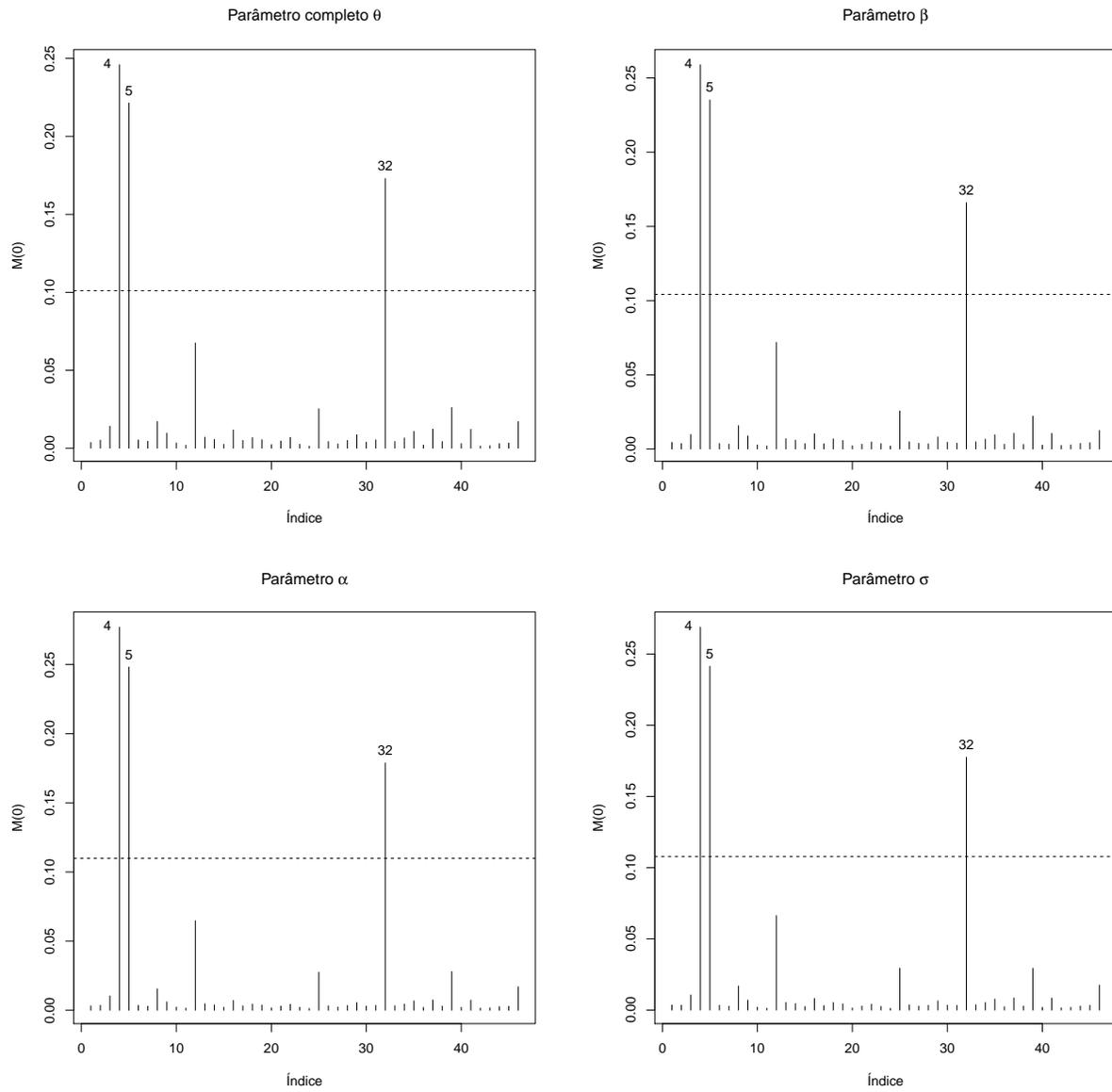


Figura 3.15: Gráfico  $M(0)$  sob perturbação de ponderação de casos para conjunto de dados de fadiga.

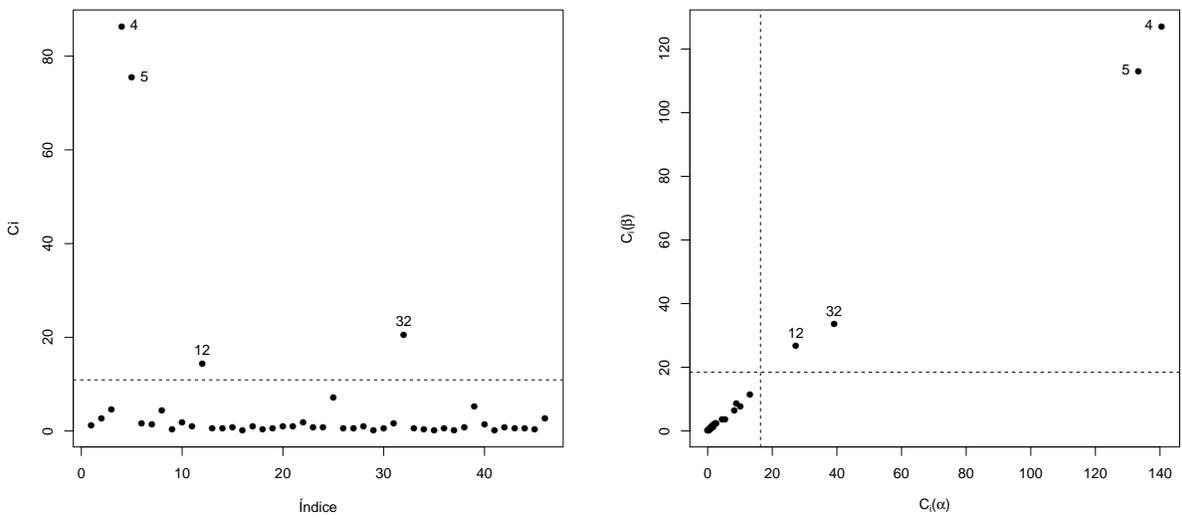


Figura 3.16: Ponderação de casos para conjunto de dados de fadiga. À esquerda gráfico de influência total  $C_i$  vs observações e à direita gráfico das médias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ .

### Perturbação na variável explicativa

Sob o esquema de perturbação na variável explicativa, na Figura 3.17 temos o gráfico  $M(0)$  para o parâmetro global e também para cada parâmetro, observe que de acordo com a marca de nível, indicada no gráfico por uma linha pontilhada, as observações 4 e 5 são identificadas como influentes para todos os parâmetros.

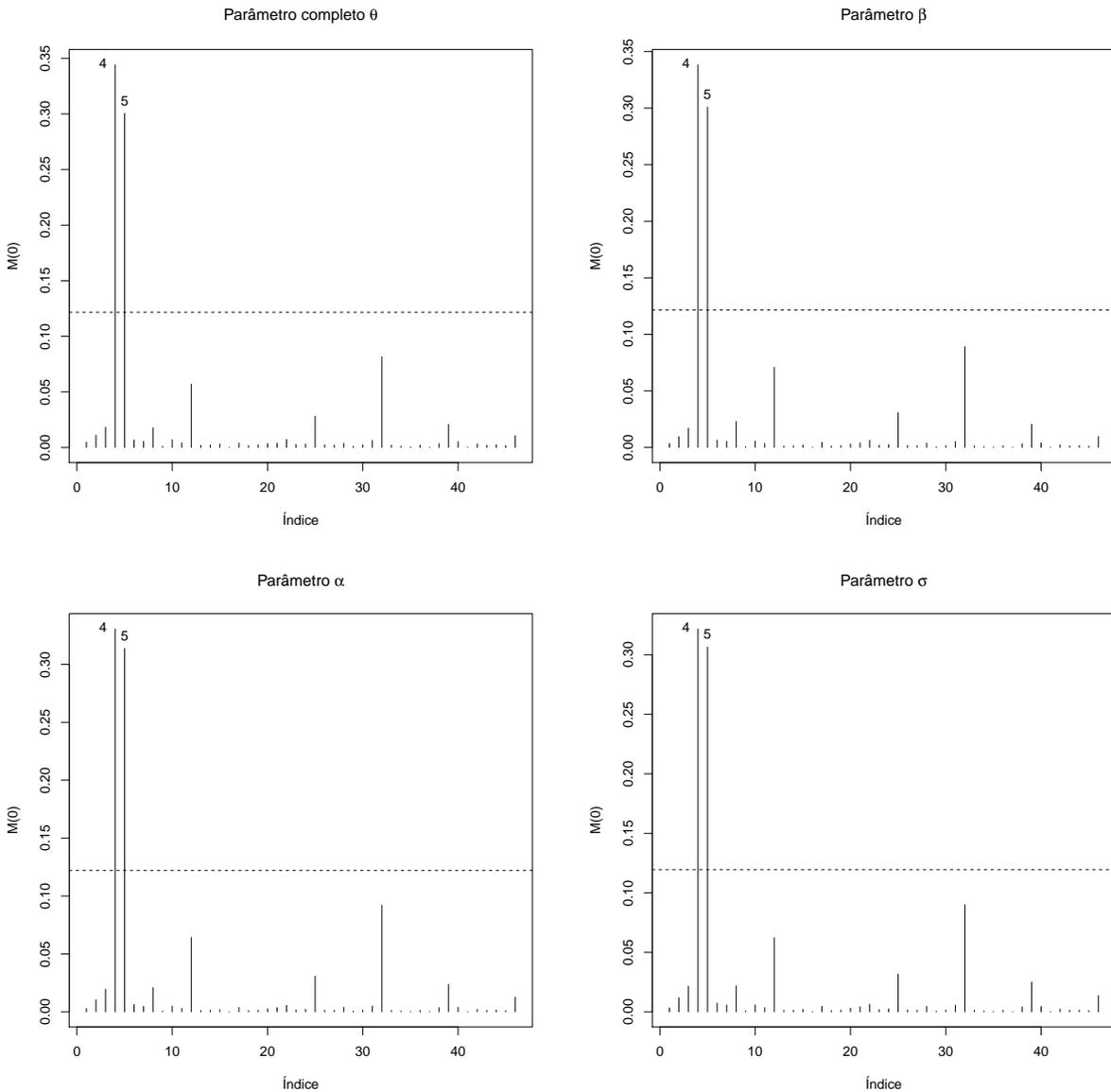


Figura 3.17: Gráfico  $M(0)$  sob perturbação na variável explicativa para conjunto de dados de fadiga.

Sob o esquema de perturbação na variável explicativa, na Figura 3.18 temos o gráfico de in-

fluência total  $C_i$  vs observações e o gráfico das medias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ . Observe que, de acordo com a marca de nível, indicada no gráfico por uma línea pontilhada, as observações 4, 5, 12 e 32 são identificadas como influentes em ambos gráficos.

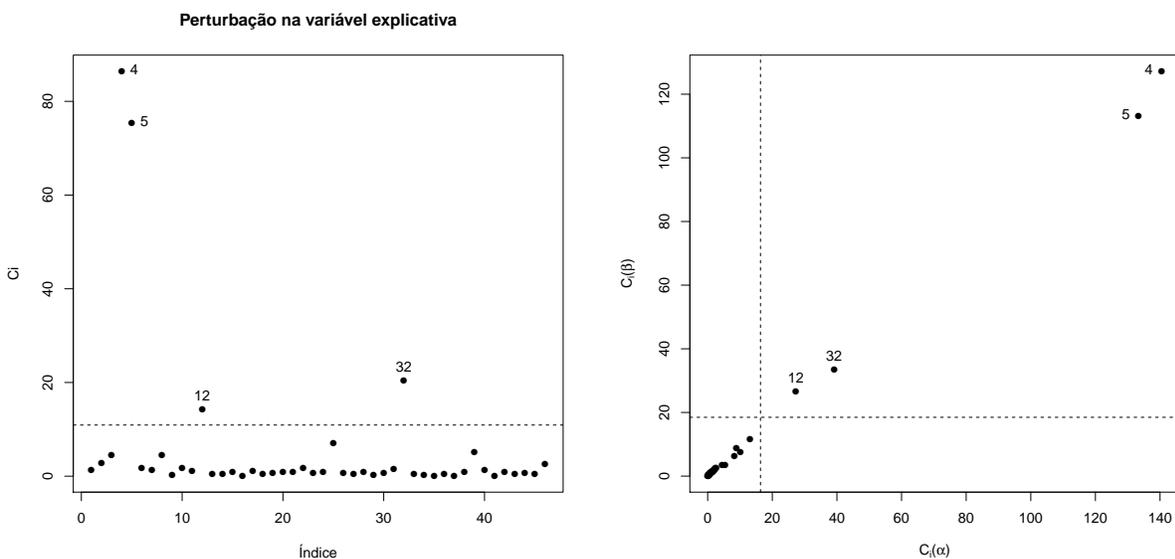


Figura 3.18: Perturbação na variável explicativa para conjunto de dados de fadiga. Na esquerda gráfico de influência total  $C_i$  vs observações; na direita gráfico das medias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ .

**Perturbação na variável resposta**

Sob o esquema de perturbação na variável resposta, na Figura 3.19 temos a gráfica  $M(0)$  para o parâmetro global e também para cada parâmetro, observe que de acordo com a marca de nível, indicada no gráfico por uma linha pontilhada, as observações 4, 5 e 32 são identificadas como influentes para todos os parâmetros.

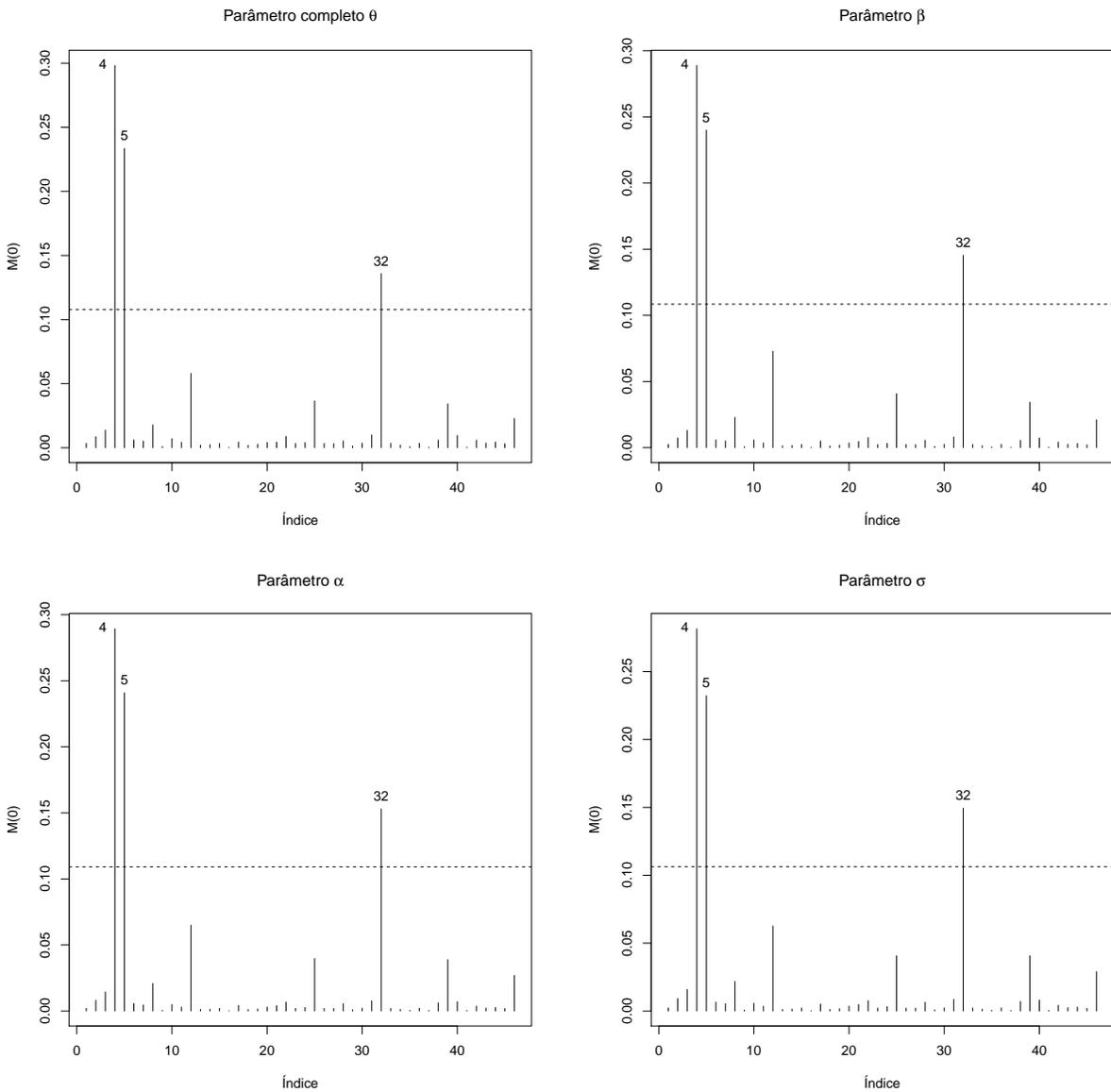


Figura 3.19: Gráfico  $M(0)$  sob perturbação na variável resposta para conjunto de dados de fadiga.

Sob o esquema de perturbação na variável explicativa, na Figura 3.20 temos o gráfico de in-

fluência total  $C_i$  vs observações e o gráfico das médias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ . Observe que, de acordo com a marca de nível, indicada no gráfico por uma línea pontilhada, as observações 4, 5, 12 e 32 são identificadas como influentes em ambos.

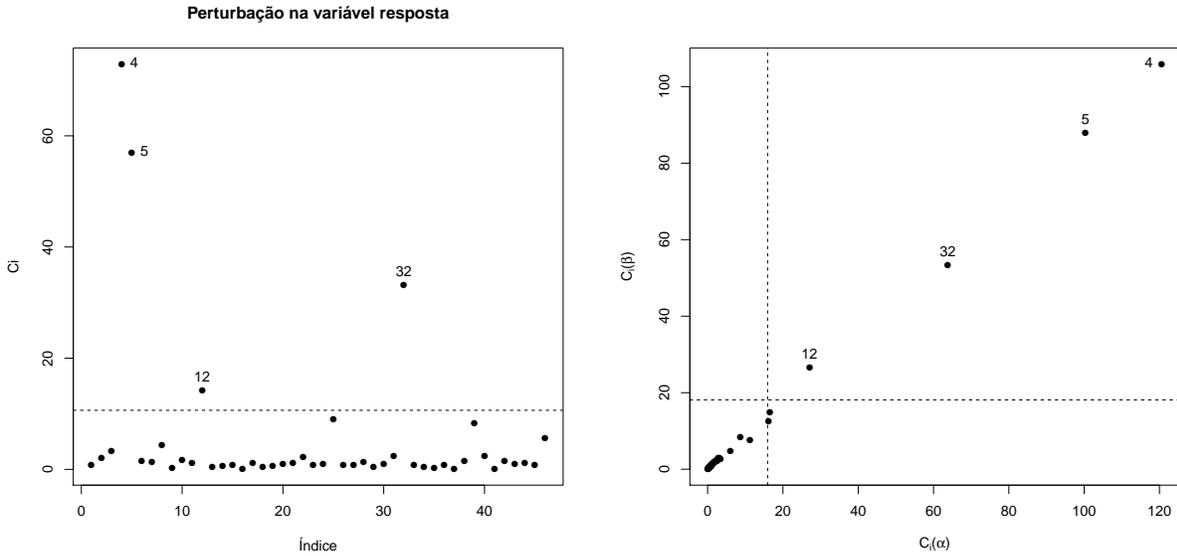


Figura 3.20: Perturbação na variável resposta. À esquerda gráfico de influência total  $C_i$  vs Observações; à direita gráfico das médias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ .

De acordo com o análise anterior temos que as observações com maior influência são 4, 5, 32 e 12 em ordem de prioridade. Logo, para avaliar o impacto das três mais influentes eliminamos as observações 4, 5 e 32 e reajustamos o modelo. Na Tabela 3.4 apresentamos a variação das estimativas  $\alpha$ ,  $\beta_1$ ,  $\beta_2$  e  $\sigma$  obtidas a través de

$$RC = \frac{|\hat{\theta}_{j(i)} - \hat{\theta}_j|}{\hat{\theta}_j} \times 100 \quad j = 0, \dots, 4 \quad e \quad i = 1, \dots, n,$$

em que  $\hat{\theta}_{j(i)}$  denota a estimativa de máxima verossimilhança de  $\theta_j$  após a retirada da  $i$ -ésima observação. Por exemplo, para  $j=1$  temos que  $\theta_1 = \beta_1$  e  $\hat{\beta}_{1(i)}$  é a estimativa de máxima verossimilhança para  $\beta_1$  após a retirada da  $i$ -ésima observação. De acordo com a Tabela 3.4 observamos que apenas o ponto 4 exerce maior mudança nas estimativas.

Tabela 3.4: Mudança nas estimativas dos parâmetros (em %), excluindo as observações potencialmente influentes.

Obs. Eliminada	$\hat{\beta}_1$	RC	$\hat{\beta}_2$	RC	$\hat{\alpha}$	RC	$\hat{\sigma}$	RC
4	2,7359	1,1614	-0,2568	3,0604	4,1125	28,8360	0,3110	16,0498
5	2,7976	1,0682	-0,2725	2,9133	3,4035	6,6263	0,3422	7,6191
32	2,7418	0,9498	-0,2555	3,5401	2,8977	9,2213	0,3777	1,9383

Os resultados obtidos após as análises de diagnóstico são similares aos obtidos por Rolim Santana (2009) e Xie & Wei (2007) os quais obtiveram as observações 4, 5, 12 e 32 como influentes.

### 3.8 Aplicação 2: Produção de cebola

Este conjunto de dados foi descrito por Ratkowsky (1983). A variável resposta,  $P$ , é a produção de cebola (g/planta) e a variável explanatória,  $x$ , é a densidade da planta (planta/ $m^2$ ), em uma amostra com 42 observações correspondentes à produção de cebola na localidade de Uraidla. Os dados são apresentados no Anexo I, Tabela I.2. Neste caso, vamos a considerar o seguinte modelo

$$y_i = \mu_i + \epsilon_i,$$

em que  $y_i = \log P_i$ ,  $\mu_i = g_0(\mathbf{x}_i^\top \boldsymbol{\beta})$  e  $\epsilon_i \sim SN(\alpha, 0, \sigma)$ ,  $i = 1, \dots, 42$ .

#### 3.8.1 Ajuste do modelo

Nesta seção ajustamos o modelo de acima usando as funções de ligação  $(\mathbf{x}_i^\top \boldsymbol{\beta})^{-1}$  e  $\mu_i = -\log(\mathbf{x}_i^\top \boldsymbol{\beta})$ , função de ligação que foi usada por Santos dos Reis (2013) para o mesmo conjunto de dados.

Na tabela 3.5 são apresentadas as EMV com seus respectivos erros padrão para os dois modelos ajustados. Observe que, para o modelo com função de ligação  $\mu_i = -\log(\mathbf{x}_i^\top \boldsymbol{\beta})$  temos erros padrão mais pequenos.

Tabela 3.5: Estimativas de máxima verossimilhança para a distribuição Sinh-Normal, com seus respectivos erros padrão, para os dados da produção de cebola.

Preditor	$\mu_i$	$\beta_1$	$\beta_2$	$\alpha$	$\sigma$
Linear	$(\mathbf{x}_i^\top \boldsymbol{\beta})^{-1}$	0,1886 (0,0023)	0,0005 (<0,001)	0,1401 (0,1112)	2,1180 (1,6370)
Linear	$-\log(\mathbf{x}_i^\top \boldsymbol{\beta})$	0,0035 (0,0005)	0,0001 (<0,001)	0,1398 (0,1065)	2,1289 (1,5753)

De acordo com os critérios AIC e BIC, apresentados na Tabela 3.6, o modelo linear com função de ligação  $\mu_i = -\log(\mathbf{x}_i^\top \boldsymbol{\beta})$ , apresenta um melhor ajuste o qual também é avaliado porque os erros padrões das estimativas deste modelo.

Tabela 3.6: Critérios para seleção de modelos para os dados da produção de cebola.

Preditor	$\mu_i$	AIC	BIC
Linear	$(\mathbf{x}_i^\top \boldsymbol{\beta})^{-1}$	-33,0564	-26,1057
Linear	$-\log(\mathbf{x}_i^\top \boldsymbol{\beta})$	-33,2961	-26,3454

Na Figura 3.21 apresentamos o diagrama de dispersão entre  $\log(\text{produção de cebola})$  e a densidade da planta com o valor médio previsto do  $\log(\text{produção de cebola})$ . Observe que, o modelo com função de ligação  $\mu_i = -\log(\mathbf{x}_i^\top \boldsymbol{\beta})$  apresenta um melhor ajuste ao conjunto de dados e que a observação 38 parece ser outlier.

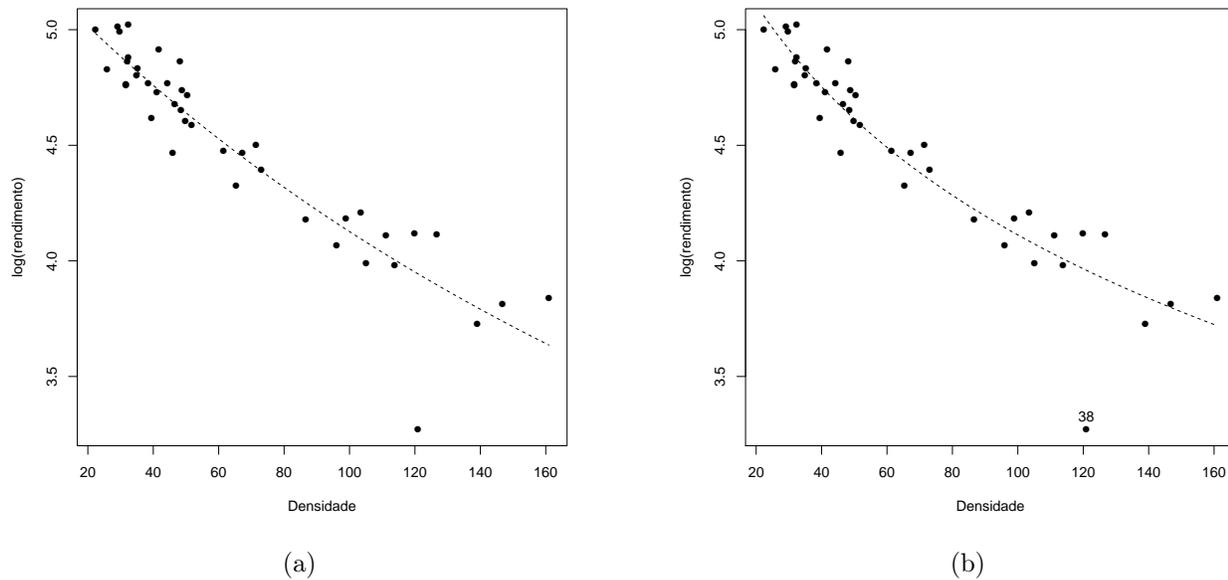


Figura 3.21: Diagrama de dispersão entre  $\log(\text{produção de cebola})$  e a densidade da planta com o valor médio previsto do  $\log(\text{produção de cebola})$  para os dados da produção de cebola (a)  $(\mathbf{x}_i^\top \boldsymbol{\beta})^{-1}$  e (b)  $-\log(\mathbf{x}_i^\top \boldsymbol{\beta})$ .

Na Figura 3.22 apresentamos a banda de confiança de 95% para o valor da média prevista de  $\log(\text{produção de cebola})$ . Observe que, a banda do modelo com função de ligação  $\mu_i = -\log(\mathbf{x}_i^\top \boldsymbol{\beta})$  tem intervalos de confiança com menor comprimento em comparação com a outra, o que fornece evidência de que o mesmo proporciona um melhor ajuste ao conjunto de dados.

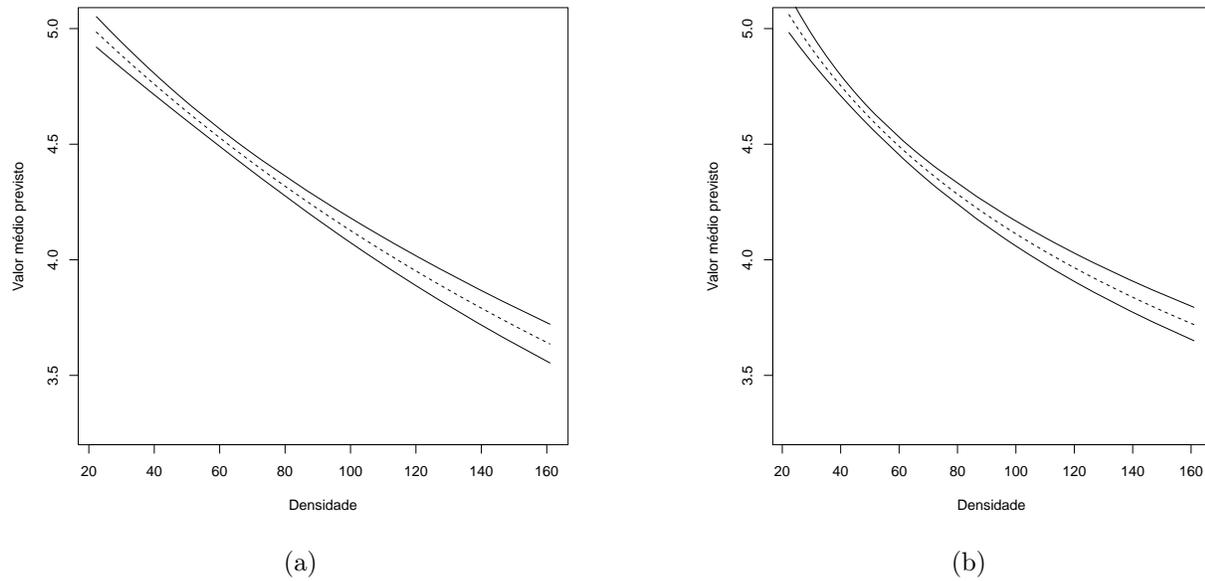


Figura 3.22: Banda de confiança de 95% para o valor da média prevista de  $\log(\text{produção de cebola})$  para os dados da produção de cebola (a)  $(\mathbf{x}_i^\top \boldsymbol{\beta})^{-1}$  e (b)  $-\log(\mathbf{x}_i^\top \boldsymbol{\beta})$ .

De acordo com o análise anterior, escolhemos o modelo com função de ligação  $\mu_i = -\log(\mathbf{x}_i^\top \boldsymbol{\beta})$ . Isto é

$$y_i = \mu_i + \epsilon_i,$$

em que  $y_i = \log P_i$ ,  $\mu_i = -\log(\mathbf{x}_i^\top \boldsymbol{\beta})$  e  $\epsilon_i \sim SN(\alpha, 0, \sigma)$ ,  $i = 1, \dots, 42$ .

Antes de realizar a análise de diagnóstico, iremos avaliar o ajuste deste modelo ao conjunto de dados e análise de resíduos.

Na Figura 3.23 temos os gráficos de probabilidade normal com envelopes simulados para os resíduos apresentados na Subseção 3.4.2 e na Figura 3.24 temos o envelope simulado, observe que, a observação 38 afeta o ajuste do modelo.

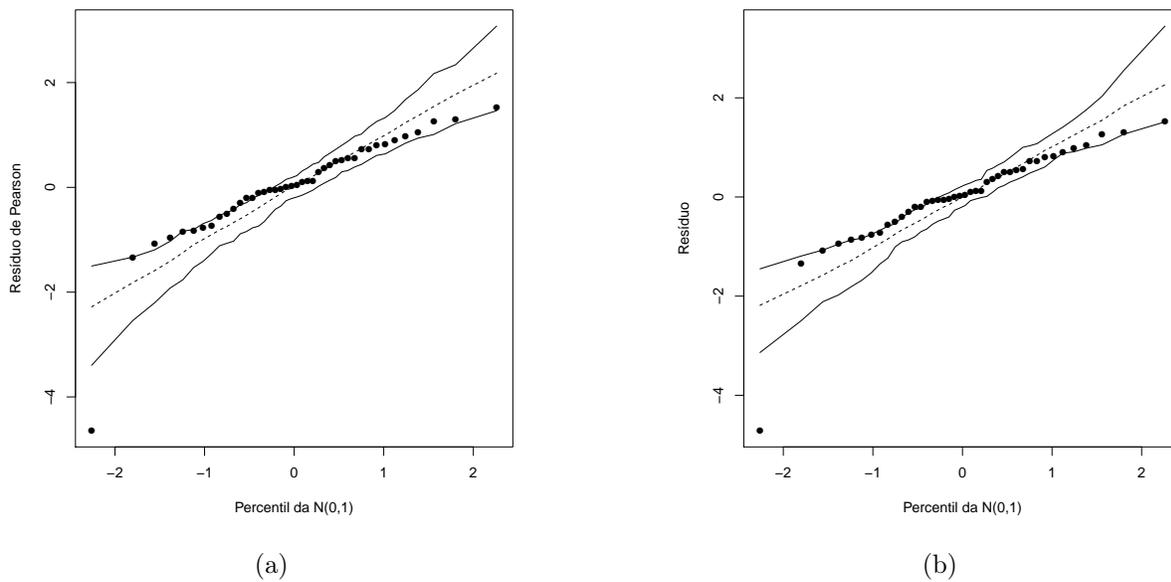


Figura 3.23: Gráfico de probabilidade normal com envelope simulado para os dados da produção de cebola para  $-\log(\mathbf{x}_i^\top \boldsymbol{\beta})$  (a) Resíduo de Pearson (b) Outro resíduo.

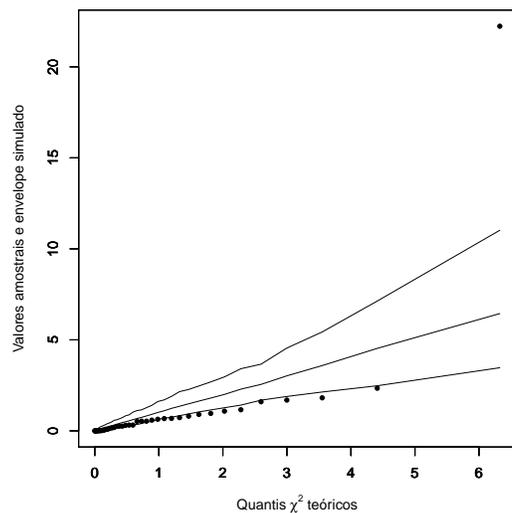


Figura 3.24: Envelope simulado para os dados da produção de cebola para  $-\log(\mathbf{x}_i^\top \boldsymbol{\beta})$ .

Na Figura 3.25 apresentamos os gráficos para os resíduos que foram abordados na Subseção 3.4.2. Observe que, a observação 38 é um possível outlier.

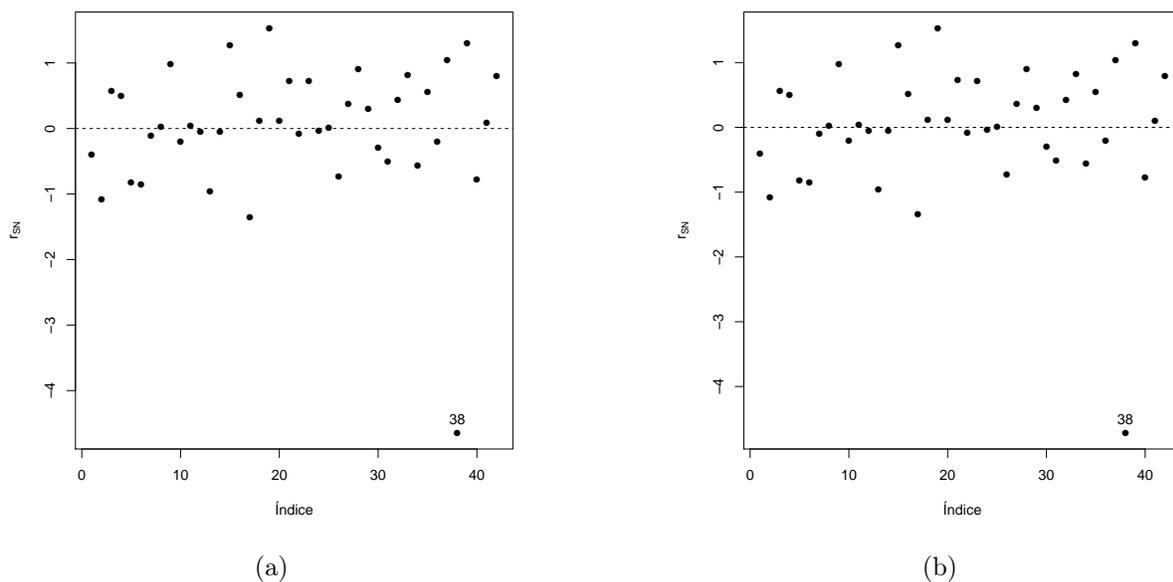


Figura 3.25: Gráfico de resíduos para os dados da produção de cebola para  $-\log(\mathbf{x}_i^\top \boldsymbol{\beta})$  (a) Resíduo de Pearson (b) Outro resíduo.

### 3.8.2 Análise de diagnóstico

Baseado nos resultados obtidos anteriormente, apresentamos os resultados da análise de diagnóstico.

#### Modelo de eliminação de casos

Nas Figuras 3.26 e 3.27 são apresentadas a distância de Cook ( $GD_i^1$ ) e na Figura 3.28 temos o afastamento pela log-verossimilhança ( $LD_i^1$ ), definidas na Subseção 3.5.1. Os resultados mostrados indicam que a observação 38 é identificada como influente nas estimações dos parâmetros.

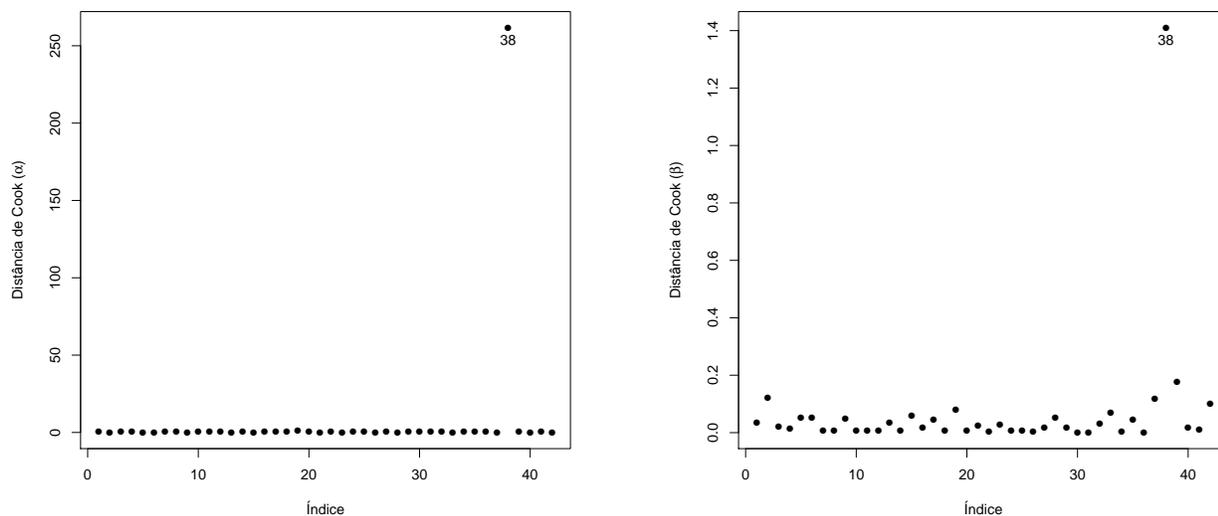


Figura 3.26: Distância Generalizada de Cook para os parâmetros  $\alpha$  e  $\beta$  para os dados da produção de cebola.

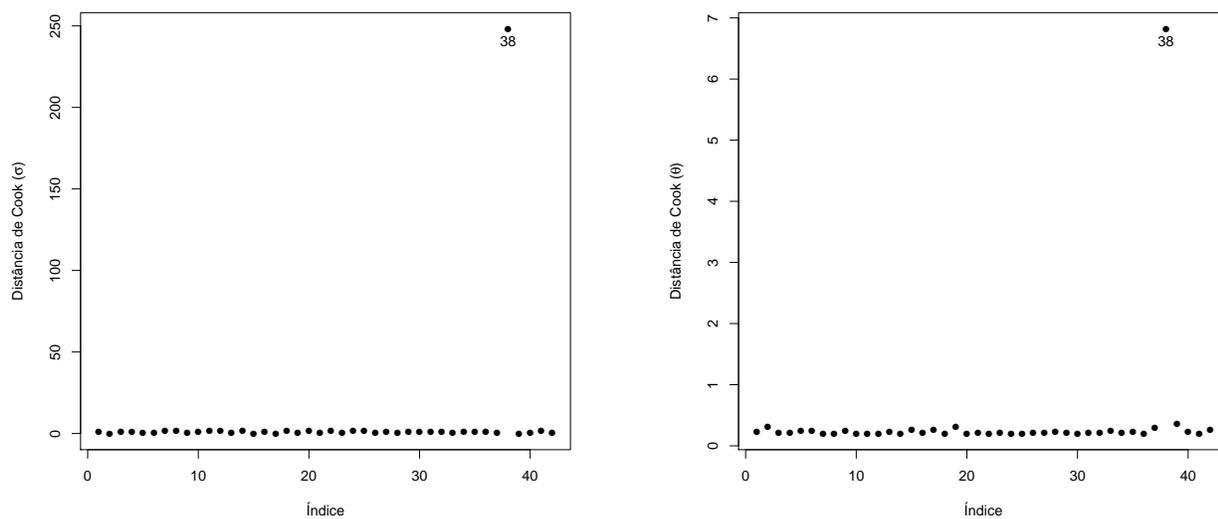


Figura 3.27: Distância Generalizada de Cook para os diferentes para o parâmetro  $\sigma$  e  $\theta$  para os dados da produção de cebola.

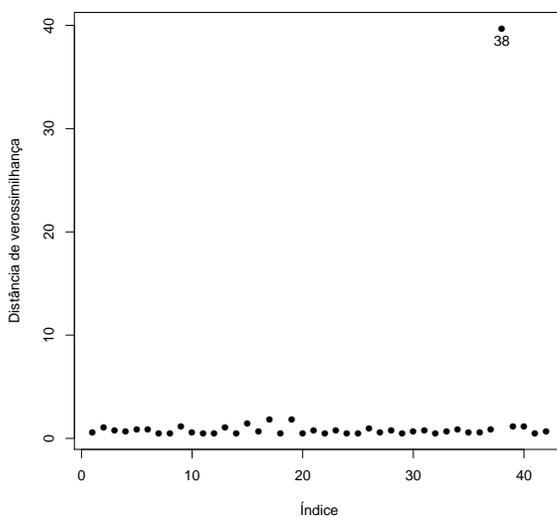


Figura 3.28: Distância de verossimilhança para os dados da produção de cebola.

### Esquemas de perturbação

Nas Figuras 3.29, 3.30 e 3.31 temos os gráficos  $M(0)$  para o parâmetro global e também para cada parâmetro dos esquemas de perturbação de ponderação de casos, perturbação na variável

explicativa e perturbação na variável resposta, respectivamente. Observe que, de acordo com a marca de nível, indicada nos gráficos por uma linha pontilhada, a observação 38 é identificada como influente nos três esquemas.

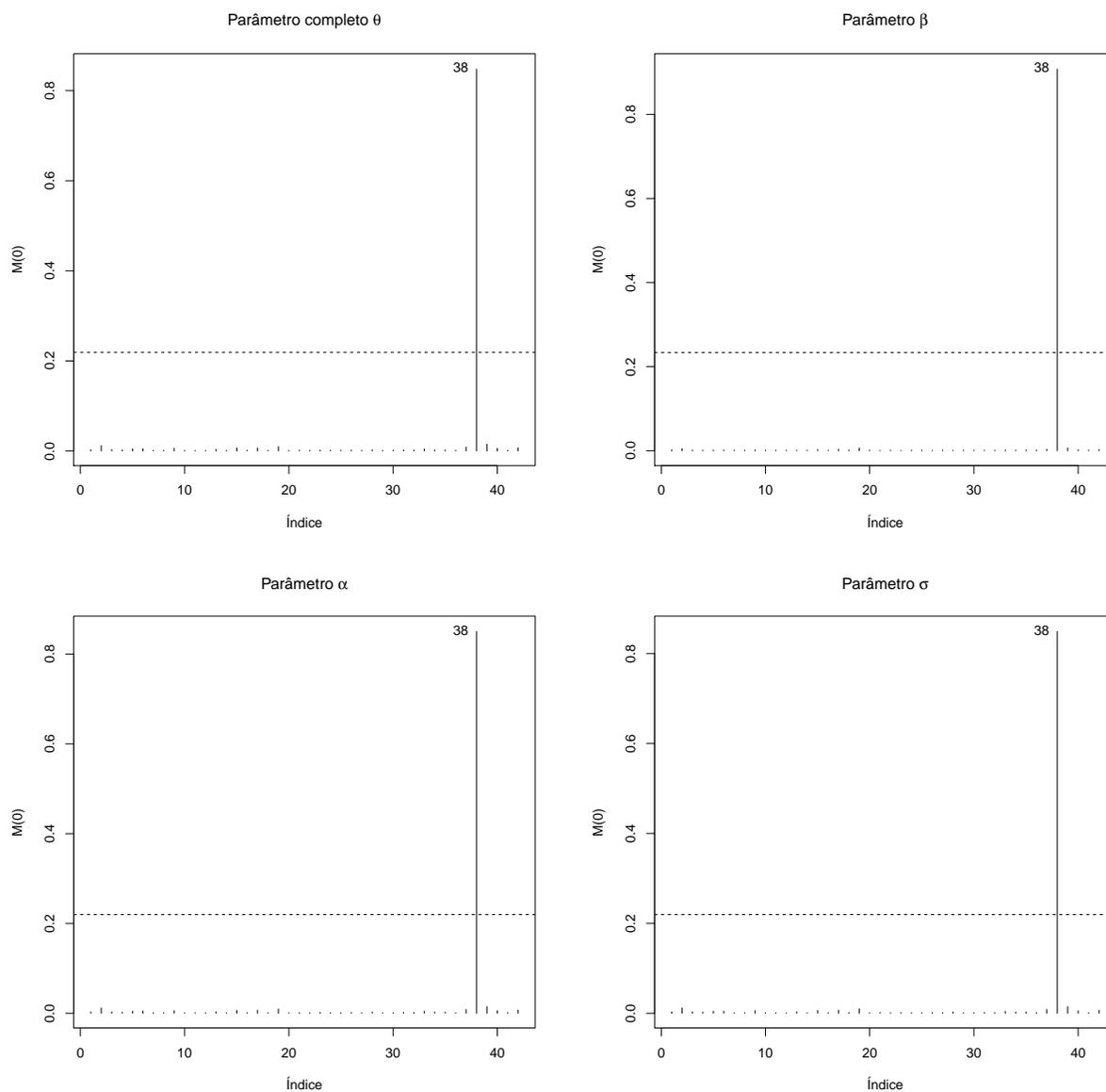


Figura 3.29: Gráfico  $M(0)$  sob perturbação de ponderação de casos para os dados da produção de cebola.

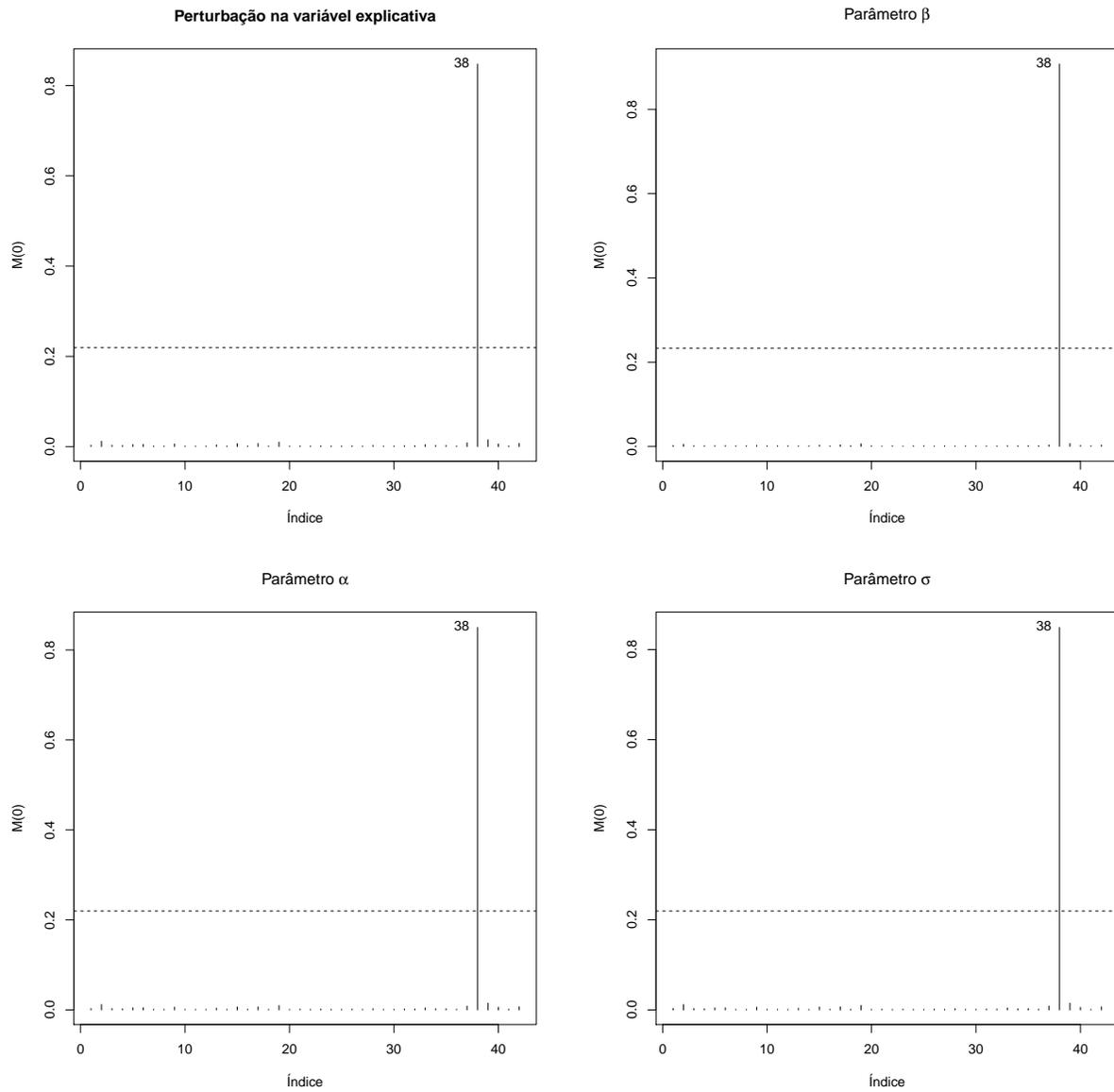


Figura 3.30: Gráfico  $M(0)$  sob perturbação na variável explicativa para os dados da produção de cebola.

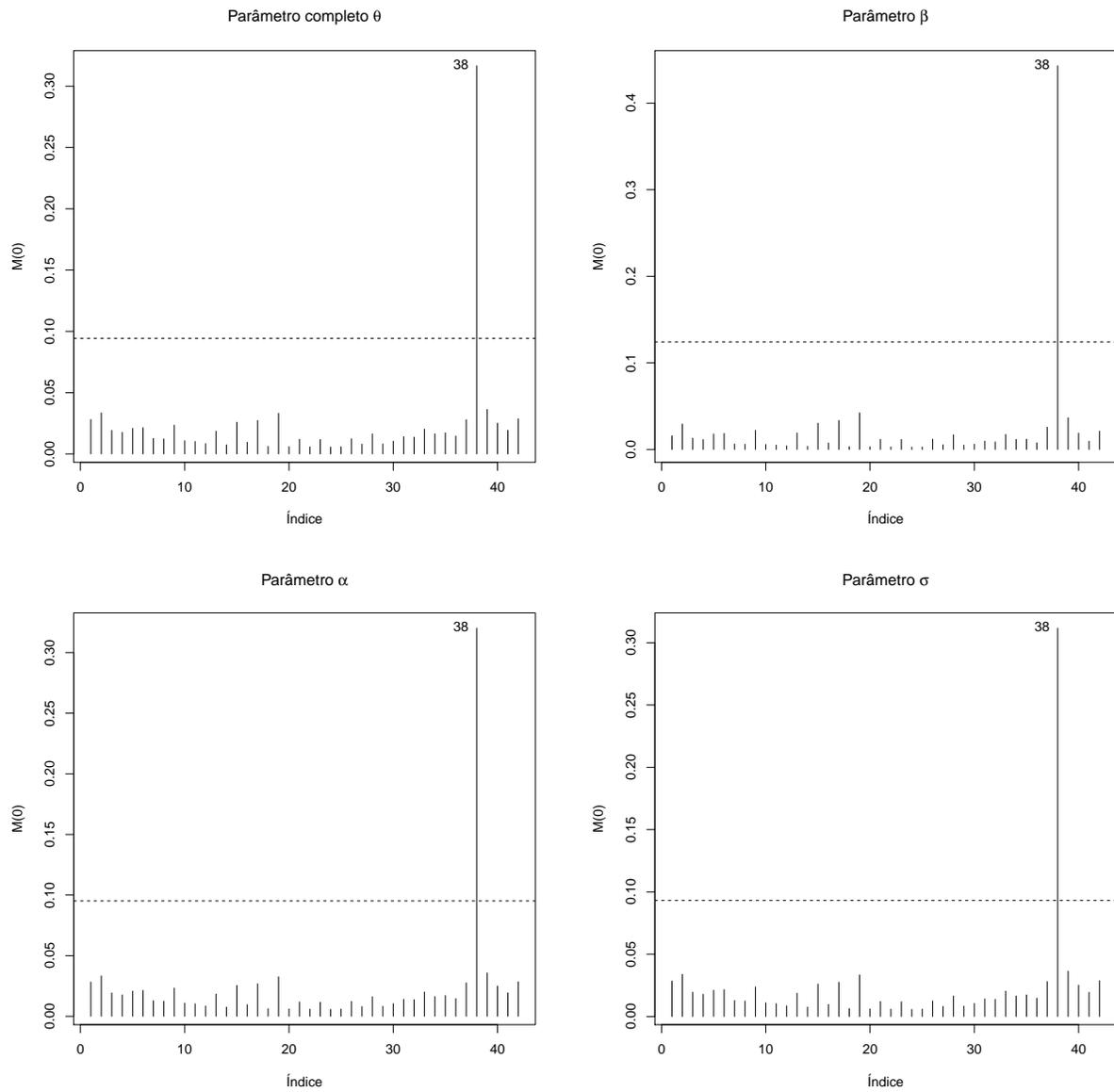


Figura 3.31: Gráfico  $M(0)$  sob perturbação na variável resposta para os dados da produção de cebola.

Para os diferentes esquemas de perturbação, nas Figuras 3.32, 3.33 e 3.34 temos os gráficos de influência total  $C_i$  vs Observações e os gráficos das medias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ . Observe que, de acordo com a marca de nível, indicada no gráfico por uma linha pontilhada, a observação 38 é identificada como influente nos três esquemas.

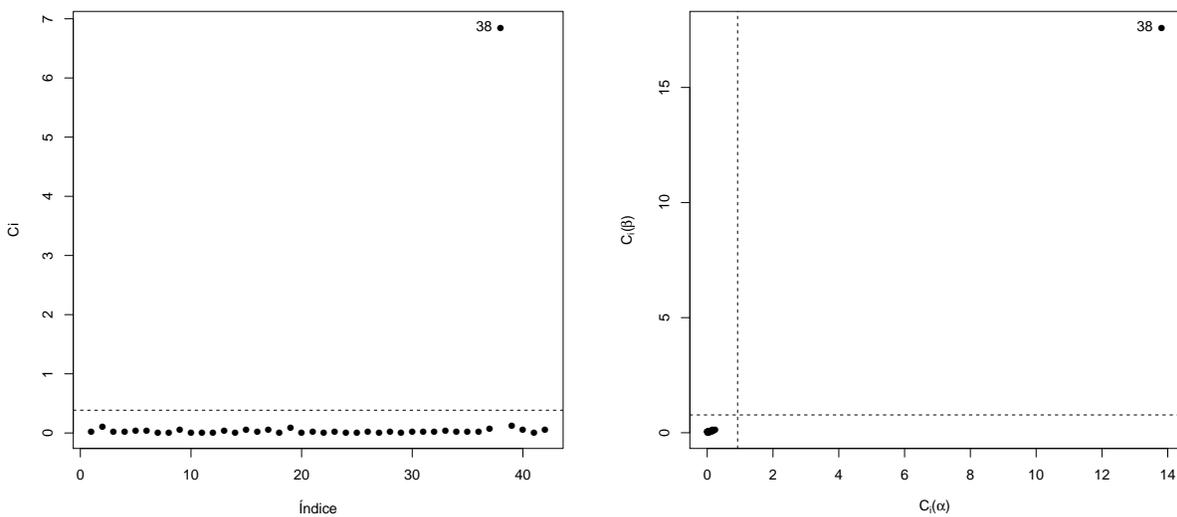


Figura 3.32: Ponderação de casos para os dados da produção de cebola. À esquerda gráfico de influência total  $C_i$  vs observações e à direita gráfico das médias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ .

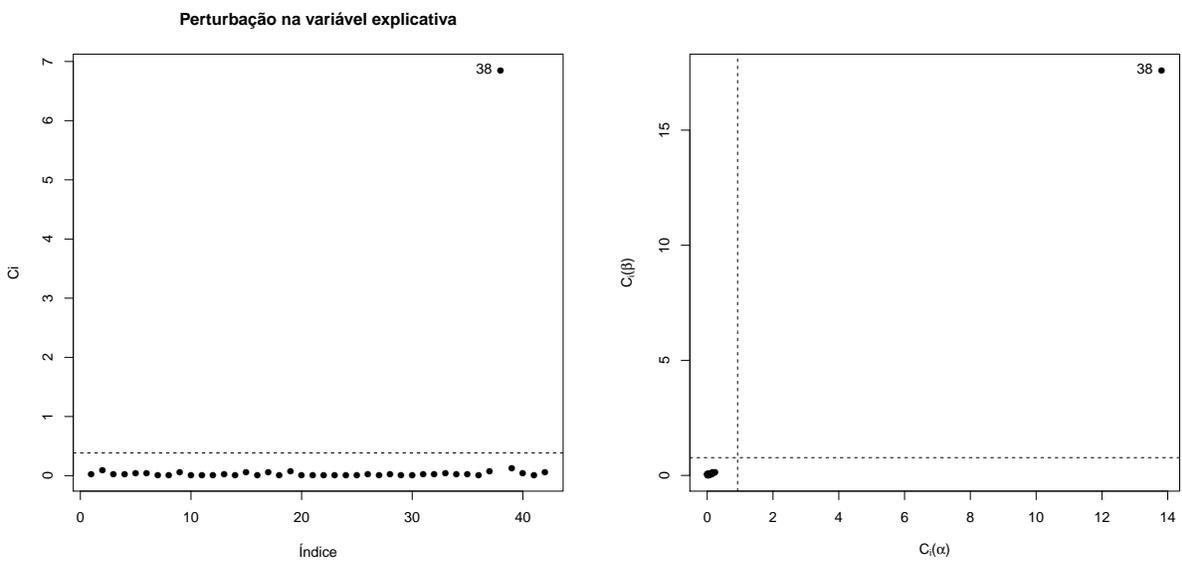


Figura 3.33: Perturbação na variável explicativa. À esquerda gráfico de influência total  $C_i$  vs observações e à direita Gráfico das médias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ .

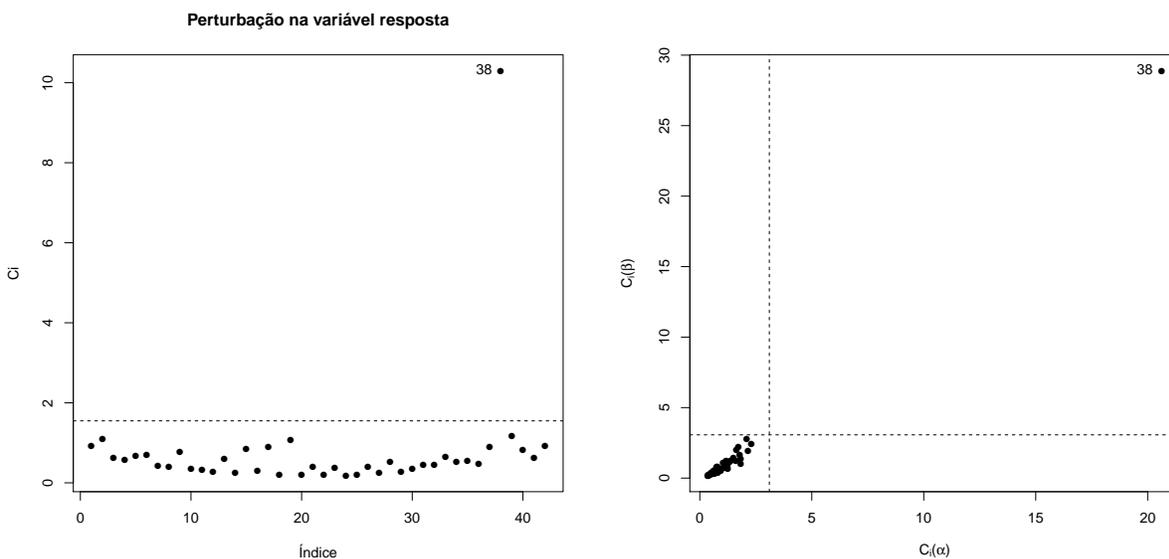


Figura 3.34: Perturbação na variável resposta para os dados da produção de cebola. À esquerda gráfico de influência total  $C_i$  vs observações e à direita gráfico das médias de influência local  $C_i(\alpha)$  vs  $C_i(\beta)$ .

Os resultados obtidos após as análises de diagnóstico são similares aos obtidos por Santos dos Reis (2013) em que a observação 38 é identificada como influente e o modelo apresenta um bom ajuste ao conjunto de dados.

# Capítulo 4

## Modelos lineares heteroscedásticos Sinh-Normal

### 4.1 Introdução

Os modelos lineares heteroscedásticos, conhecidos como aqueles que tem um parâmetro de dispersão que é atribuído para cada observação, sendo relacionado linearmente através de combinações lineares de variáveis explicativas, por meio de uma função de ligação conhecida. Estes modelos, tem sido largamente discutidos na área de econometria, veja por exemplo Park (1966) e Harvey (1976), assim como, na literatura Estatística. Por exemplo, Barroso & Cordeiro (2005) consideraram correções de Bartlett em modelos de regressão heteroscedásticos com distribuições de  $t$ -Student; Cysneiros *et al.* (2007) discutiram métodos de diagnóstico em modelos lineares heteroscedásticos com erros simétricos e Cysneiros *et al.* (2010) introduziram a classe de modelos de regressão não-lineares heteroscedásticos simétricos onde um processo iterativo conjunto para estimar a média e os parâmetros de dispersão é proposto. Neste trabalho, vamos considerar um modelo de regressão SN, quando a variância dos erros muda com as observações. Uma primeira proposta nesta linha de estudo foi discutida em Xie & Wei (2007), para o modelo de regressão log BS, em que o parâmetro de forma muda com as observações. No modelo de regressão SN homocedástico, os erros  $\varepsilon_1, \dots, \varepsilon_n$  são independentes com variância constante,  $\text{Var}(\varepsilon_i) = \sigma^2\eta(\alpha)$ , em que  $\alpha$  e  $\sigma$  são os parâmetros de forma e escala, respectivamente.

### 4.2 O modelo de regressão heteroscedástico

Propomos um modelo de regressão linear heteroscedástico definido por

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.2.1)$$

em que  $Y_i$  é a variável resposta,  $\mathbf{x}_i$  é um vetor  $p \times 1$  de variáveis explicativas conhecidas associadas com a  $i$ -ésima resposta observável  $Y_i$  e  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  é um vetor de parâmetros desconhecidos. A especificação do modelo é obtida ao considerar

$$\varepsilon_i \sim \text{SN}(\alpha_i, 0, \sigma_i), \quad i = 1, \dots, n. \quad (4.2.2)$$

Assim, das propriedades da distribuição SN dada por Rieck & Nedelman (1991),  $Y_i \sim \text{SN}(\alpha_i, \mathbf{x}_i^\top \boldsymbol{\beta}, \sigma_i)$  e sua função de densidade toma neste caso a forma

$$f(y_i) = \frac{2}{\sigma_i \alpha_i} \phi \left( \frac{2}{\alpha_i} \sinh \left( \frac{y_i - \mu_i}{\sigma_i} \right) \right) \cosh \left( \frac{y_i - \mu_i}{\sigma_i} \right) \quad (4.2.3)$$

$$= \frac{1}{\sigma_i} \phi(\xi_{2i}) \xi_{1i}, \quad (4.2.4)$$

em que  $\xi_{1i} = \xi_1(y_i; \alpha_i, \mu_i, \sigma_i)$  e  $\xi_{2i} = \xi_2(y_i; \alpha_i, \mu_i, \sigma_i)$  são como em (3.2.2), com  $\mu_i = \mathbf{x}_i^\top \boldsymbol{\beta}$ . Para  $\alpha_i = \alpha$  para todo  $i$  e  $\sigma = 2$ , obtemos o modelo de regressão linear log-BS, definido por Rieck & Nedelman (1991). Além disso, para  $\sigma = 2$ , o modelo proposto aqui reduz ao que foi discutido por Xie & Wei (2007). Neste modelo, temos de (2.1.2) que

$$d_i = \frac{4}{\alpha_i^2} \sinh^2 \left( \frac{Y_i - \mu_i}{\sigma_i} \right) \sim \chi_1^2, \quad (4.2.5)$$

e é extremamente útil para testar a qualidade do ajuste e para a detecção de outliers da mesma forma como é feito no contexto distribuição normal. De Rieck & Nedelman (1991), podemos obter a média e a variância de  $Y_i$  que são, aproximadamente, dadas por

$$\mathbb{E}[Y_i] = \mathbf{x}_i^\top \boldsymbol{\beta} \quad \text{e} \quad \text{Var}(Y) = \sigma_i^2 \eta(\alpha_i), \quad (4.2.6)$$

em que  $\eta(\alpha_i) = \frac{\alpha_i^2}{4} (1 - \frac{\alpha_i^2}{4})$  para  $0 < \alpha_i < 2$  e  $\eta(\alpha_i) = 2.197543451 - 1.963510026 \ln(\alpha_i \sqrt{2}) + [\ln(\alpha_i \sqrt{2})]^2$  para  $\alpha_i > 2$ .

A função de log-verossimilhança para  $\boldsymbol{\beta}$  dada a amostra observada  $\mathbf{y}$  é dada por

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2} \log(2\pi) - \sum_{i=1}^n \log(\sigma_i) + \sum_{i=1}^n \log(\xi_{1i}) - \frac{1}{2} \sum_{i=1}^n \xi_{2i}^2, \quad (4.2.7)$$

em que  $\xi_{1i}$  e  $\xi_{2i}$  são como em (4.2.3). Note que, a variância de  $Y_i$ , depende dos parâmetros de forma  $\alpha_i$  e de escala  $\sigma_i$ . Assim em nosso estudo vamos estudar dois casos, de maneira similar como é feita em Vilca *et al.* (2012) e Xie *et al.* (2009):

- (a)  $\alpha_i$  varia com as observações e  $\sigma_i = \sigma$  para todas  $i = 1, \dots, n$ .
- (b)  $\alpha_i = \alpha$  para todas  $i = 1, \dots, n$  e  $\sigma_i$  varia com as observações.

A seguir, vamos discutir a homogeneidade dos parâmetros de forma  $\alpha_i$  definida no caso (a) e depois o caso (b).

### 4.2.1 Teste escore para homogeneidade do parâmetro de forma

O parâmetro de forma  $\alpha$  no modelo de regressão linear SN é muito importante, porque define a forma e o modo da distribuição. Seguindo a ideia de Xie & Wei (2007), o parâmetro de forma foi assumido como sendo dependente das variáveis explicativas. Aqui, consideramos

$$\alpha_i = \alpha m_i, \quad (4.2.8)$$

em que  $m_i = m(\boldsymbol{\lambda}, \mathbf{z}_i)$  é uma função positiva conhecida continuamente diferenciável e  $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top$ ,  $i = 1, \dots, n$ , contendo valores das variáveis explicativas e  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_q)^\top$  um vector de parâmetros desconhecidos, os quais constituem, em geral, embora não seja necessário, um subconjunto de  $x_1, \dots, x_n$ .

Neste modelo, assume-se que existe um valor único  $\boldsymbol{\lambda}_0$  de  $\boldsymbol{\lambda}$  de tal modo que  $m_i(\boldsymbol{\lambda}_0, \mathbf{z}_i) = 1$  para todo  $i$ . É fácil ver que  $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ , então os  $Y_i$  têm variância constante. Se as variâncias dependem dos valores de algumas variáveis explicativas  $\mathbf{z}_i$ , duas formas específicas de  $m_i$  são usualmente tomadas, ver Cook & Weisberg (1983) e Simonoff & Tsai (1994):

- (i) uma forma específica de  $m_i$  é o modelo log-linear dado por  $m_i(\boldsymbol{\lambda}, \mathbf{z}_i) = \exp \left\{ \sum_{j=1}^q \lambda_j z_{ij} \right\}$ ;
- (ii) modelo do produto potência  $m_i(\boldsymbol{\lambda}, \mathbf{z}_i) = \prod_{j=1}^q z_{ij}^{\lambda_j} = \exp \left\{ \sum_{j=1}^q \lambda_j \log(z_{ij}) \right\}$ .

Neste caso é necessário que o  $z_{ij}$  seja estritamente positivo. Os resultados de Chen (1983) sugere que o teste heteroscedástico não é muito sensível para a forma funcional.

Tendo em conta os pressupostos acima e a ideia de Xie & Wei (2007), o teste de homogeneidade para o parâmetro de forma é equivalente ao teste de hipóteses

$$H_0 : \boldsymbol{\lambda} = \boldsymbol{\lambda}_0; \quad H_1 : \boldsymbol{\lambda} \neq \boldsymbol{\lambda}_0. \quad (4.2.9)$$

Para testar as hipóteses apresentadas acima, podemos fazer uso do teste escore. Note-se que sob esta especificação  $\boldsymbol{\lambda}$  é o parâmetro de interesse, e  $\boldsymbol{\beta}$ ,  $\alpha$  e  $\sigma$  são os parâmetros de perturbação. Assim seja  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha, \sigma, \boldsymbol{\lambda}^\top)^\top$ , e a função de log-verossimilhança para o parâmetro  $\boldsymbol{\theta}$  pode ser expresso como

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2} \log(2\pi) - n \log(\sigma) + \sum_{i=1}^n \log(\xi_{1im}) - \frac{1}{2} \sum_{i=1}^n \xi_{2im}^2, \quad (4.2.10)$$

em que, sob este modelo

$$\xi_{1im} = \frac{2}{\alpha m_i} \cosh \left( \frac{y_i - \mu_i}{\sigma} \right) = \frac{1}{m_i} \xi_{1i}, \quad \xi_{2im} = \frac{2}{\alpha m_i} \sinh \left( \frac{y_i - \mu_i}{\sigma} \right) = \frac{1}{m_i} \xi_{2i}.$$

As seguintes notações serão de utilidade para expressar a matriz de informação observada que é usada para obter a estatística escore. Seja  $\mathbf{m} = (m_1, \dots, m_n)^\top$  o vector com componentes  $m_i$ .

Por outro lado, denotamos o vetor da primeira derivada parcial de  $m_i$  em relação a  $\lambda$  e a matriz de segundas derivadas de  $m_i$ , respectivamente, por

$$\dot{\mathbf{m}} = \frac{\partial \mathbf{m}^\top}{\partial \boldsymbol{\lambda}} = \left( \frac{\partial m_1}{\partial \boldsymbol{\lambda}}, \dots, \frac{\partial m_n}{\partial \boldsymbol{\lambda}} \right) \text{ e } \mathbf{D}_{\lambda\lambda} = \left( \frac{\partial m_1}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top}, \dots, \frac{\partial m_n}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top} \right).$$

Derivando  $\ell(\boldsymbol{\theta})$  em relação a  $\boldsymbol{\beta}$ ,  $\alpha$ ,  $\sigma$  e  $\boldsymbol{\lambda}$ , temos as funções score para  $\boldsymbol{\theta}$  dada por  $U(\boldsymbol{\theta}) = (U_\beta^\top(\boldsymbol{\theta}), U_\alpha(\boldsymbol{\theta}), U_\sigma(\boldsymbol{\theta}), U_\lambda^\top(\boldsymbol{\theta}))^\top$  cujos elementos tomam, respectivamente, as formas

$$\begin{aligned} U_\beta(\boldsymbol{\theta}) &= \frac{1}{\sigma} \mathbf{X}^\top [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)] \boldsymbol{\xi}_2, \\ U_\alpha(\boldsymbol{\theta}) &= \frac{1}{\alpha} (\boldsymbol{\xi}_2^\top \boldsymbol{\xi}_2 - n), \\ U_\sigma(\boldsymbol{\theta}) &= \frac{1}{\sigma} [\boldsymbol{\epsilon}^\top [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)] \boldsymbol{\xi}_2 - n], \\ U_\lambda(\boldsymbol{\theta}) &= -\dot{\mathbf{m}}^\top D^{-1}(\mathbf{m}) [\mathbf{1}_n - D(\boldsymbol{\xi}_2) \boldsymbol{\xi}_2]. \end{aligned}$$

A matriz de informação observada  $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$  que é obtida a partir da matriz de segundas derivadas da função de log-verossimilhança  $\ell(\boldsymbol{\theta})$ ,  $\ddot{\mathbf{L}}(\boldsymbol{\theta}) = \partial^2 \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$  cujos elementos podem ser expressados como

$$\begin{aligned} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{1}{\sigma} \left[ -\frac{1}{\sigma} \mathbf{X}^\top \mathbf{A}_{12} \mathbf{X} \right], \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \alpha} &= -\frac{2}{\alpha \sigma} \mathbf{X}^\top D(\boldsymbol{\xi}_1) \boldsymbol{\xi}_2, \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma} &= -\frac{1}{\sigma^2} \mathbf{X}^\top [\mathbf{A}_{12} \boldsymbol{\epsilon} + [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)] \boldsymbol{\xi}_2], \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \alpha} &= \frac{1}{\alpha^2} [n - 3\mathbf{1}_n^\top D(\boldsymbol{\xi}_2) \boldsymbol{\xi}_2], \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma \partial \alpha} &= -\frac{2}{\alpha \sigma} \boldsymbol{\epsilon}^\top D(\boldsymbol{\xi}_1) \boldsymbol{\xi}_2, \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma \partial \sigma} &= \frac{1}{\sigma^2} [n - \boldsymbol{\epsilon}^\top \mathbf{A}_{12} \boldsymbol{\epsilon} - 2\boldsymbol{\epsilon}^\top [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_1)] \boldsymbol{\xi}_2], \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top} &= \dot{\mathbf{m}}^\top D^{-2}(\mathbf{m}) [\mathbf{I}_n - 3D^2(\boldsymbol{\xi}_2)] \dot{\mathbf{m}} + \mathbf{D}_{\lambda\lambda} D^{-1}(\mathbf{m}) [D(\boldsymbol{\xi}_2) \boldsymbol{\xi}_2 - \mathbf{1}_n] \otimes \mathbf{I}_q, \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\beta}^\top} &= -\frac{2}{\sigma} \dot{\mathbf{m}}^\top D^{-1}(\mathbf{m}) D(\boldsymbol{\xi}_1) D(\boldsymbol{\xi}_2) \mathbf{X}, \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \alpha} &= -\frac{2}{\alpha} \dot{\mathbf{m}}^\top D^{-1}(\mathbf{m}) D(\boldsymbol{\xi}_2) \boldsymbol{\xi}_2, \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \sigma} &= -\frac{2}{\sigma} \dot{\mathbf{m}}^\top D^{-1}(\mathbf{m}) D(\boldsymbol{\xi}_1) D(\boldsymbol{\xi}_2) \boldsymbol{\epsilon}, \end{aligned}$$

em que  $\mathbf{A}_{12} = [D^2(\boldsymbol{\xi}_1) + D^2(\boldsymbol{\xi}_2)] + D^2(\boldsymbol{\xi}_2)D^{-2}(\boldsymbol{\xi}_1) - \mathbf{I}_n$  e  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$ , com  $\epsilon_i = (y_i - \mu_i)/\sigma$ .

O seguinte teorema apresenta a estatística escore para testar  $H_0$ .

**Teorema 4.2.1.** *Para o modelo em (4.5.1), o teste de escore para avaliar  $H_0$  contra  $H_1$  para o parâmetro de forma é dado por*

$$SC = \tilde{\mathbf{A}}_\lambda^\top \left[ -\ddot{\mathbf{L}}(\tilde{\boldsymbol{\theta}}) \right]^{-1} \tilde{\mathbf{A}}_\lambda,$$

em que  $\tilde{\mathbf{A}}_\lambda = (0, 0, 0, 0, U_\lambda(\tilde{\boldsymbol{\theta}}))^\top$  com  $U_\lambda(\tilde{\boldsymbol{\theta}}) = -\dot{\mathbf{m}}^\top D^{-1}(\mathbf{m}) [\mathbf{1}_n - D(\boldsymbol{\xi}_2)\boldsymbol{\xi}_2]$ . Para  $n$  grande tem-se que  $SC \sim \chi_{(r)}^2$  sob  $H_0$ .

## 4.2.2 Teste escore para a homogeneidade do parâmetro de escala

Aqui, em lugar de  $\sigma$  vamos considerar  $\sigma_i$ , variando com as observações. Vamos assumir que os  $\sigma_i$  dependem de algumas variáveis explicativas  $\mathbf{z}_i$ . Seguindo Cook & Weisberg (1983), Tsai (1986), Smyth (1989) e Simonoff & Tsai (1994), vamos assumir que o  $\sigma_i$  pode ser modelada por

$$\sigma_i = \sigma m_i, \quad (4.2.11)$$

em que  $m_i = m(\boldsymbol{\lambda}, \mathbf{z}_i)$  é uma função positiva conhecida, continuamente diferenciável e  $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^\top$ ,  $i = 1, \dots, n$ , contendo valores das variáveis explicativas e  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_q)^\top$  um vector de parâmetros desconhecidos, os quais constituem, em geral, embora não seja necessário, um subconjunto de  $x_1, \dots, x_n$ .

Neste modelo, assume-se que existe um valor único  $\boldsymbol{\lambda}_0$  de  $\boldsymbol{\lambda}$  de tal modo que  $m_i(\boldsymbol{\lambda}_0, \mathbf{z}_i) = 1$  para todo  $i$ . É fácil ver que  $\boldsymbol{\lambda} = \boldsymbol{\lambda}_0$ , então os  $Y_i$  tem variância constante. Se as variâncias dependem dos valores de algumas variáveis explanatórias  $\mathbf{z}_i$ , vamos considerar as formas específicas de  $m_i$  de acordo com a Secção 4.2.1.

Para testar a hipótese apresentada em (4.2.9), podemos fazer uso do teste de escore. Note-se que sob esta especificação  $\boldsymbol{\lambda}$  é o parâmetro de interesse, e  $\boldsymbol{\beta}$ ,  $\alpha$  e  $\sigma$  são os parâmetros de perturbação. Assim, o parâmetro do modelo é determinada por  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \alpha, \sigma, \boldsymbol{\lambda}^\top)^\top$ , e a função de log-verossimilhança para o parâmetro  $\boldsymbol{\theta}$  pode ser expresso como

$$\ell(\boldsymbol{\theta}) = -\frac{n}{2} \log(2\pi) - n \log(\sigma) - \sum_{i=1}^n \log(m_i) + \sum_{i=1}^n \log(\xi_{1im}) - \frac{1}{2} \sum_{i=1}^n \xi_{2im}^2, \quad (4.2.12)$$

em que, sob este modelo

$$\xi_{1im} = \frac{2}{\alpha} \cosh\left(\frac{y_i - \mu_i}{\sigma m_i}\right), \quad \xi_{2im} = \frac{2}{\alpha} \sinh\left(\frac{y_i - \mu_i}{\sigma m_i}\right).$$

Derivando  $\ell(\boldsymbol{\theta})$  em relação à  $\boldsymbol{\beta}$ ,  $\alpha$ ,  $\sigma$  e  $\boldsymbol{\lambda}$ , temos as funções escore para  $\boldsymbol{\theta}$  dada por  $U(\boldsymbol{\theta}) = (U_\beta^\top(\boldsymbol{\theta}), U_\alpha(\boldsymbol{\theta}), U_\sigma(\boldsymbol{\theta}), U_\lambda^\top(\boldsymbol{\theta}))^\top$  cujos elementos tomam, respectivamente, as formas

$$\begin{aligned}
 U_\beta(\boldsymbol{\theta}) &= \frac{1}{\sigma} \mathbf{X}^\top D^{-1}(\mathbf{m}) [D(\boldsymbol{\xi}_{1m}) - D^{-1}(\boldsymbol{\xi}_{1m})] \boldsymbol{\xi}_{2m}, \\
 U_\alpha(\boldsymbol{\theta}) &= \frac{1}{\alpha} (\boldsymbol{\xi}_{2m}^\top \boldsymbol{\xi}_{2m} - n), \\
 U_\sigma(\boldsymbol{\theta}) &= \frac{1}{\sigma} [\boldsymbol{\epsilon}^\top D^{-1}(\mathbf{m}) [D(\boldsymbol{\xi}_{1m}) - D^{-1}(\boldsymbol{\xi}_{1m})] \boldsymbol{\xi}_{2m} - n], \\
 U_\lambda(\boldsymbol{\theta}) &= -\dot{\mathbf{m}}^\top \{ D^{-1}(\mathbf{m}) \mathbf{1}_n - D^{-2}(\mathbf{m}) D(\boldsymbol{\xi}_{2m}) [D(\boldsymbol{\xi}_{1m}) - D^{-1}(\boldsymbol{\xi}_{1m})] \boldsymbol{\epsilon} \},
 \end{aligned}$$

em que

$$\boldsymbol{\xi}_{1m} = (\xi_{11m}, \dots, \xi_{1nm})^\top \quad \text{e} \quad \boldsymbol{\xi}_{2m} = (\xi_{21m}, \dots, \xi_{2nm})^\top.$$

A matriz de informação observada  $-\ddot{\mathbf{L}}(\boldsymbol{\theta})$  que é obtida a partir da matriz de segundas derivadas da função de log-verossimilhança  $\dot{\mathbf{L}}(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$ , que neste caso os seus elementos podem ser expresso como

$$\begin{aligned}
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} &= \frac{1}{\sigma} \left[ -\frac{1}{\sigma} \mathbf{X}^\top D^{-2}(\mathbf{m}) \mathbf{A}_{12m} \mathbf{X} \right], \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \alpha} &= -\frac{2}{\alpha \sigma} \mathbf{X}^\top D^{-1}(\mathbf{m}) D(\boldsymbol{\xi}_{1m}) \boldsymbol{\xi}_{2m}, \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma} &= -\frac{1}{\sigma^2} \mathbf{X}^\top [D^{-2}(\mathbf{m}) \mathbf{A}_{12m} \boldsymbol{\epsilon} + D^{-1}(\mathbf{m}) [D(\boldsymbol{\xi}_1) - D^{-1}(\boldsymbol{\xi}_{1m})] \boldsymbol{\xi}_{2m}], \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \alpha \partial \alpha} &= \frac{1}{\alpha^2} [n - 3 \mathbf{1}_n^\top D(\boldsymbol{\xi}_{2m}) \boldsymbol{\xi}_{2m}], \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma \partial \alpha} &= -\frac{2}{\alpha \sigma} \boldsymbol{\epsilon}^\top D^{-1}(\mathbf{m}) D(\boldsymbol{\xi}_{1m}) \boldsymbol{\xi}_{2m}, \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma \partial \sigma} &= \frac{1}{\sigma^2} [n - \boldsymbol{\epsilon}^\top D^{-2}(\mathbf{m}) \mathbf{A}_{12m} \boldsymbol{\epsilon} - 2 \boldsymbol{\epsilon}^\top D^{-1}(\mathbf{m}) [D(\boldsymbol{\xi}_{1m}) - D^{-1}(\boldsymbol{\xi}_{1m})] \boldsymbol{\xi}_{2m}], \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\lambda}^\top} &= -\dot{\mathbf{m}}^\top [D^{-4}(\mathbf{m}) \mathbf{A}_{12m} D^2(\boldsymbol{\epsilon}) + 2 D^{-3}(\mathbf{m}) \mathbf{C}_{12} D(\boldsymbol{\epsilon}) - D^{-2}(\mathbf{m})] \dot{\mathbf{m}} - \mathbf{D}_{\lambda\lambda} \mathbf{D}_{12}(m), \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \boldsymbol{\beta}^\top} &= -\frac{1}{\sigma} \dot{\mathbf{m}}^\top [D^{-2}(\mathbf{m}) \mathbf{C}_{12} + D^{-3}(\mathbf{m}) D(\boldsymbol{\epsilon}) \mathbf{A}_{12m}] \mathbf{X}, \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \alpha} &= -\frac{2}{\alpha} \dot{\mathbf{m}}^\top D^{-2}(\mathbf{m}) D(\boldsymbol{\xi}_{1m}) D(\boldsymbol{\xi}_{2m}) \boldsymbol{\epsilon}, \\
 \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda} \partial \sigma} &= -\frac{1}{\sigma} \dot{\mathbf{m}}^\top [D^{-3}(\mathbf{m}) \mathbf{A}_{12m} D(\boldsymbol{\epsilon}) + D^{-2}(\mathbf{m}) \mathbf{C}_{12}] \boldsymbol{\epsilon},
 \end{aligned}$$

em que  $\mathbf{A}_{12m} = [D^2(\boldsymbol{\xi}_{1m}) + D^2(\boldsymbol{\xi}_{2m})] + D^2(\boldsymbol{\xi}_{2m}) D^{-2}(\boldsymbol{\xi}_{1m}) - \mathbf{I}_n$ ,  $\mathbf{C}_{12} = D(\boldsymbol{\xi}_{2m}) [D(\boldsymbol{\xi}_{1m}) - D^{-1}(\boldsymbol{\xi}_{1m})]$ ,  $\mathbf{D}_{12}(m) = [D^{-1}(\mathbf{m}) \mathbf{1}_n - D^{-2}(\mathbf{m}) \mathbf{C}_{12}] \boldsymbol{\epsilon} \otimes \mathbf{I}_q$ , e  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$ , com  $\epsilon_i = (y_i - \mu_i) / \sigma$ .

**Teorema 4.2.2.** Para o modelo em (4.5.1), o teste de escore para avaliar  $H_0$  contra  $H_1$  para o parâmetro de escala é dado por

$$SC = \tilde{\mathbf{A}}_\lambda^\top \left[ -\ddot{\mathbf{L}}(\tilde{\boldsymbol{\theta}}) \right]^{-1} \tilde{\mathbf{A}}_\lambda,$$

em que

$$\tilde{\mathbf{A}}_\lambda = \left( 0, 0, 0, 0, U_\lambda(\tilde{\boldsymbol{\theta}}) \right)^\top,$$

com

$$U_\lambda(\tilde{\boldsymbol{\theta}}) = -\dot{\mathbf{m}}^\top \left\{ D^{-1}(\mathbf{m}) \mathbf{1}_n - D^{-2}(\mathbf{m}) D(\boldsymbol{\xi}_{2m}) \left[ D(\boldsymbol{\xi}_{1m}) - D^{-1}(\boldsymbol{\xi}_{1m}) \right] \boldsymbol{\epsilon} \right\}.$$

Para  $n$  grande tem-se que  $SC \sim \chi_{(r)}^2$  sob  $H_0$ .

### Estimativa dos parâmetros

Seja  $\tilde{\boldsymbol{\theta}} = (\tilde{\boldsymbol{\beta}}^\top, \tilde{\alpha}, \tilde{\sigma}, \boldsymbol{\lambda}_0^\top)^\top$  a estimativa de máxima verossimilhança restrita de  $\boldsymbol{\theta}$  sob  $H_0$  e  $\tilde{m}_i = m(\mathbf{z}_i, \boldsymbol{\lambda}_0) = 1$ . Uma vez que sob modelos restritos a expressão para estimativa do parâmetro  $\tilde{\boldsymbol{\theta}}$  não tem uma forma analítica fechada, devem ser usados procedimentos iterativos como o algoritmo de Newton-Raphson. Para fazer uso deste algoritmo precisamos de valores iniciais, neste caso  $\boldsymbol{\beta}^{(0)}$  é obtido pelo método de mínimos quadrados,  $\sigma^{(0)}$  como em 3.4.2 com  $g_0(\mathbf{x}_i^\top \boldsymbol{\beta}^{(0)}) = \mathbf{x}_i^\top \boldsymbol{\beta}$ , no caso dos outros parâmetros  $\boldsymbol{\lambda}^{(0)}$  e  $\alpha^{(0)}$  são respectivamente dados por

$$\boldsymbol{\lambda}^{(0)} = 0 \quad \text{e} \quad \alpha^{(0)} = \frac{4}{n} \sum_{i=1}^n \sinh^2 \left( \frac{y_i - \mathbf{x}_i^\top \boldsymbol{\beta}^{(0)}}{\hat{\sigma}^{(0)}} \right).$$

O conjunto de equações iterativas do método de Newton Raphson é assim definido:

$$\tilde{\boldsymbol{\theta}}^{(t+1)} = \tilde{\boldsymbol{\theta}}^{(t)} + [\mathbf{I}(\tilde{\boldsymbol{\theta}})^{(t)}]^{-1} \mathbf{U}(\tilde{\boldsymbol{\theta}}^{(t)}), \quad (4.2.13)$$

em que  $\mathbf{I}(\tilde{\boldsymbol{\theta}}) = -\ddot{\mathbf{L}}$ , sendo que  $\tilde{\boldsymbol{\theta}}^{(t)}$  e  $\tilde{\boldsymbol{\theta}}^{(t+1)}$  são os estimadores de  $\boldsymbol{\theta}$  nos passos  $t$  e  $t+1$ , respectivamente e  $\mathbf{U}(\tilde{\boldsymbol{\theta}})$  é o vetor escore, que fornece os erros padrão para as estimativas dos parâmetros. As iterações são repetidas até que uma regra de convergência adequada seja satisfeita. Como critério de convergência, podemos utilizar  $\left\| \tilde{\boldsymbol{\theta}}^{(t+1)} - \tilde{\boldsymbol{\theta}}^{(t)} \right\| < \epsilon$ , onde  $\|\mathbf{a}\|$  indica a norma do vetor  $\mathbf{a}$  e  $\epsilon > 0$ .

O método de Newton-Raphson pode gerar uma sequência de valores que convergem para um ponto onde a função tem um máximo local, isto acontece quando os valores iniciais não são apropriados. Em consequência, é recomendável repetir algumas vezes o processo considerando valores iniciais distintos para determinar o ponto onde a função alcança seu valor máximo.

### 4.3 Estudo de simulação

Nesta seção, examinamos o desempenho das propriedades do teste estatístico escore através de simulações de Monte Carlo para amostras finitas. As simulações são realizadas para dois casos: (a) o teste de homogeneidade do parâmetro de forma, (b) o teste de homogeneidade do parâmetro de escala.

O modelo utilizado para o estudo de simulação foi

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, 2, \dots, n, \quad (4.3.1)$$

para os dois casos discutidos: (a)  $\epsilon_i \sim SN(\alpha_i, 0, \sigma)$ , (b)  $\epsilon_i \sim SN(\alpha, 0, \sigma_i)$ . Aqui consideramos as seguintes suposições  $m_i = m(\boldsymbol{\lambda}, \mathbf{z}_i) = m(\lambda, x_i)$ ,  $i = 1, \dots, n$ . Seja  $\sigma_i = \sigma m(\lambda, x_i)$  e  $\alpha_i = \alpha m(\lambda, x_i)$  em que  $m(\lambda, x_i) = x_i^\lambda$  ou  $m(\lambda, x_i) = \exp(x_i \lambda)$ .

Podemos definir os verdadeiros valores como

$$\beta_0 = 0, \quad \beta_1 = 3 \text{ e } \alpha = 0.5,$$

respectivamente. Geramos primeiro números pseudo aleatórios de uma distribuição uniforme no intervalo  $[1, 6]$  para os  $x_i$  ( $i = 1, \dots, n$ ). Para obter os valores de  $y_i$ , os erros  $\epsilon_i$  foram gerados a partir de  $SN(\alpha_i, 0, \sigma)$  para o caso (a) e  $SN(\alpha, 0, \sigma_i)$  no caso (b), primeiro fazemos uma amostragem  $Z$  de uma  $N(0, 1)$  e, em seguida, o colocamos em  $\epsilon_i = \sigma \operatorname{arcsinh}(0.5\alpha_i Z)$  no caso (a) e em  $\epsilon_i = \sigma_i \operatorname{arcsinh}(0.5\alpha Z)$  no caso (b), o verdadeiro valor do parâmetro  $\alpha$  e dado  $\lambda$ , de acordo com o modelo (4.3.1), obteve-se o valores de  $y_i$ .

Neste caso, nós, respectivamente, definimos  $\lambda = 0, 0.2, 0.4, 0.6, 0.8$  de acordo com Xie & Wei (2007). Fazendo o procedimento referido acima  $n$  vezes, nós temos um conjunto de dados  $y_i, x_i$ ,  $i = 1, 2, \dots, n$ , que serão usados para calcular os valores da estatística de teste. Para cada valor de  $\lambda$  dado, fizemos 5000 repetições (os valores de  $x_i$  de foram fixados para cada repetição). Em seguida, a proporção de vezes que rejeitou a hipótese nula foi apenas o valor simulado de poder. Aqui, todas as estatísticas foram comparadas com o valor crítico  $\chi_1^2$  a um nível de significância  $\alpha = 0,05$ .

A segunda coluna da Tabela 4.1 apresenta os níveis de significância estimados, do teste escore, correspondentes ao nível nominal de 5%. Podemos observar que o teste escore para o caso da função exponencial (entre parênteses) converge ligeiramente rápido ao nível nominal de 5% em comparação com o caso da função potência. Isto também pode ser notado nas Figuras 4.1 e 4.2 que representam a comparação entre a função de distribuição empírica da estatística escore e a função de distribuição  $\chi_1^2$  para função potência e a função exponencial para o caso (a) respectivamente para a hipótese definida na Subseção 4.2.1. A terceira coluna até a sexta da Tabela 4.1 apresenta os poderes estimados do teste escore para diferentes valores de  $\lambda > 0$ . O teste de escore considerando o caso (a) sobre função exponencial,  $\exp(x_i \lambda)$ , apresenta um melhor comportamento em termos do poder em comparação com função potência,  $x_i^\lambda$ . Note que para  $n$  grande o comportamento do

testes escore é otimizado consideravelmente.

Tabela 4.1: Nível de significância e poderes, para o caso (a) sob a função de potência (sem parênteses) e exponencial (com parênteses).

$n$	$\lambda = 0$	$\lambda = 0.2$	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$
10	0,0982 (0,0966)	0,0994 (0,1532)	0,1260 (0,5252)	0,2708 (0,6798)	0,3224 (0,7566)
20	0,0738 (0,0734)	0,1026 (0,2226)	0,2848 (0,7868)	0,3818 (0,9144)	0,5360 (0,9376)
40	0,0612 (0,0600)	0,1184 (0,3842)	0,4256 (0,9624)	0,4682 (0,9840)	0,5386 (0,9880)
60	0,0494 (0,0472)	0,1416 (0,5170)	0,5522 (0,9922)	0,8460 (0,9932)	0,9648 (0,9962)
80	0,0474 (0,0486)	0,1606 (0,6200)	0,6696 (0,9953)	0,9656 (0,9970)	0,9908 (0,9982)

A segunda coluna da Tabela 4.2 apresenta os níveis de significância estimados, do teste escore, correspondentes ao nível nominal de 5%. Podemos observar que o teste escore para o caso exponencial (entre parênteses) converge mais rápido ao nível nominal de 5% em comparação com o caso potência. Isto também pode ser notado nas Figuras 4.3 e 4.4 que representam a comparação entre a função de distribuição empírica da estatística escore e a função de distribuição  $\chi_1^2$  para função potência e a função exponencial para o caso (b) respectivamente para a hipótese definida na Subseção 4.2.1. A terceira coluna até a sexta da Tabela 4.2 apresenta os poderes estimados do teste escore para diferentes valores de  $\lambda > 0$ . O teste escore considerando o caso (b) sobre função exponencial,  $\exp(x_i\lambda)$ , apresenta um melhor comportamento em termos do poder em comparação com função potencia,  $x_i^\lambda$ . Note que para  $n$  grande o comportamento do testes escore é otimizado consideravelmente.

Tabela 4.2: Tamanhos simulados e poderes para o caso (b) em função de potência (sem parênteses) e a função exponencial (com parênteses).

$n$	$\lambda = 0$	$\lambda = 0.2$	$\lambda = 0.4$	$\lambda = 0.6$	$\lambda = 0.8$
10	0,0652 (0,0640)	0,0816 (0,3106)	0,1138 (0,5582)	0,3332 (0,8028)	0,4172 (0,9248)
20	0,0712 (0,0704)	0,0848 (0,4884)	0,1468 (0,8694)	0,4972 (0,9768)	0,6840 (0,9948)
40	0,0528 (0,0592)	0,1166 (0,7874)	0,2266 (0,9764)	0,7878 (0,9926)	0,9372 (0,9998)
60	0,0444 (0,0468)	0,1390 (0,9238)	0,3050 (0,9792)	0,9248 (0,9968)	0,9922 (1,0000)
80	0,0480 (0,0480)	0,1572 (0,9742)	0,3762 (0,9768)	0,9726 (0,9988)	0,9992 (1,0000)

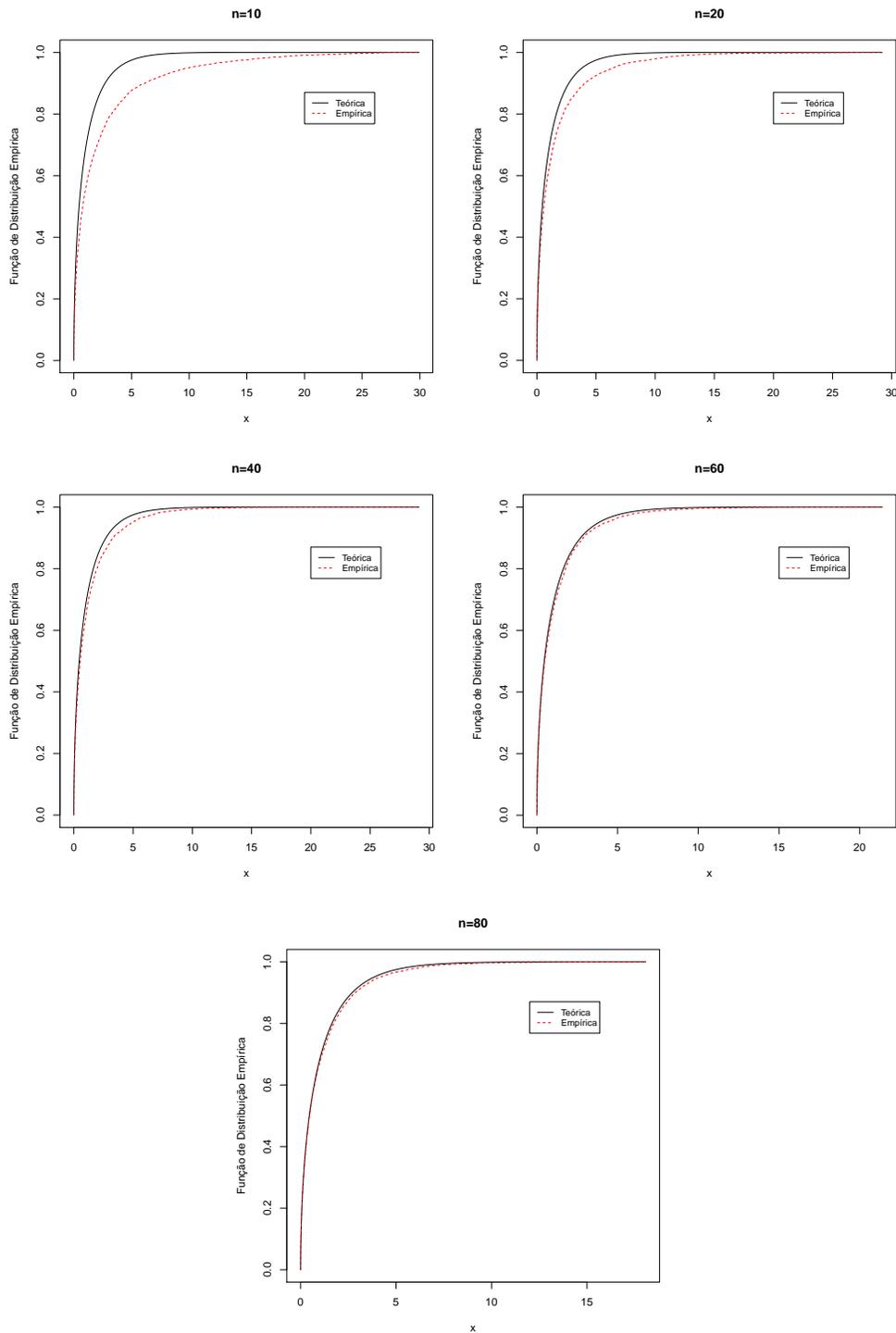


Figura 4.1: Distribuição empírica e teórica da estatística escore, para testar  $H_0$  para o caso (a) sobre a função potência, para diferente tamanhos de amostra.

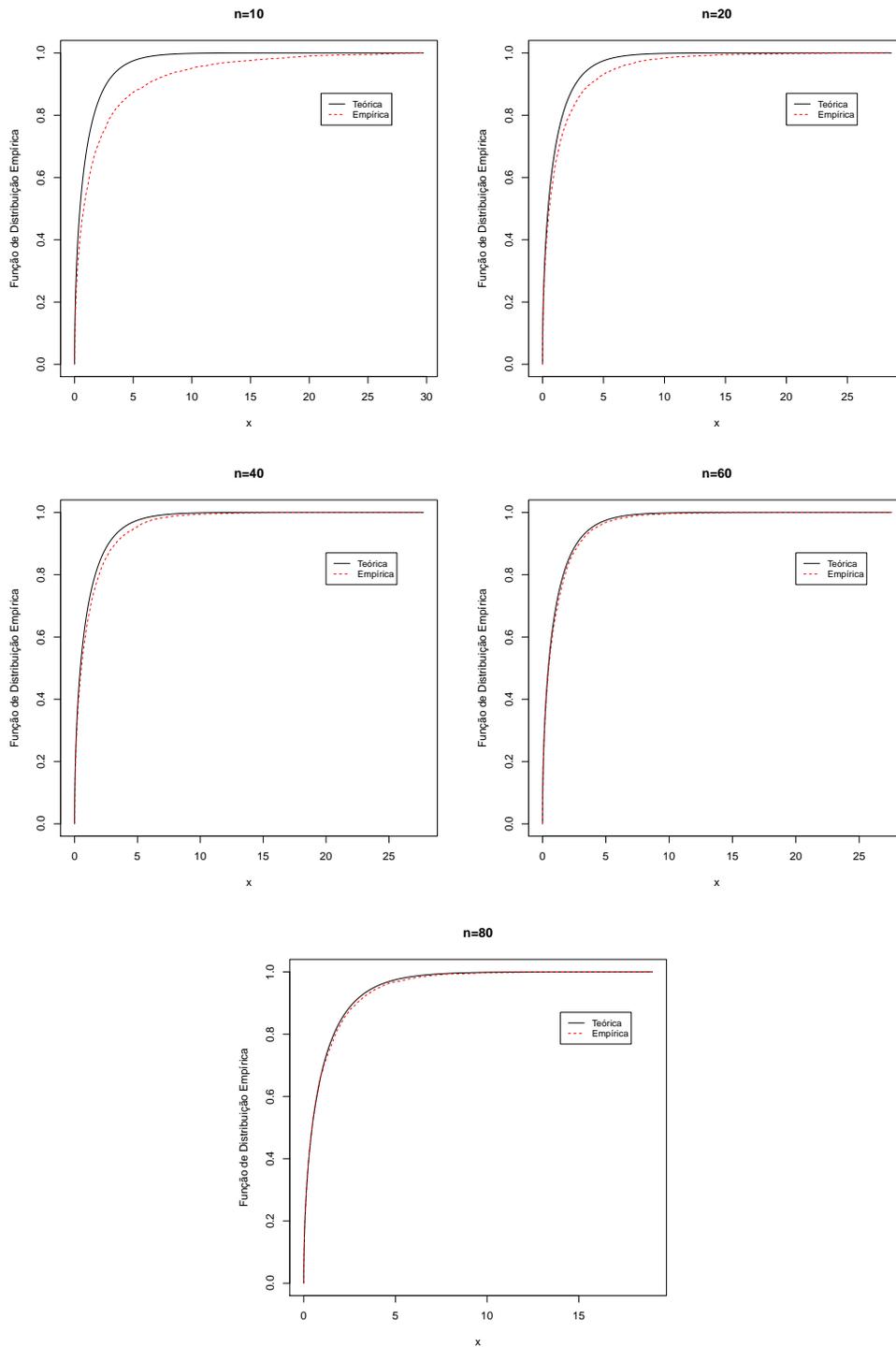


Figura 4.2: Distribuição empírica e teórica da estatística escore, para testar  $H_0$  para o caso (a) sobre a função exponencial, para diferente tamanhos de amostra.

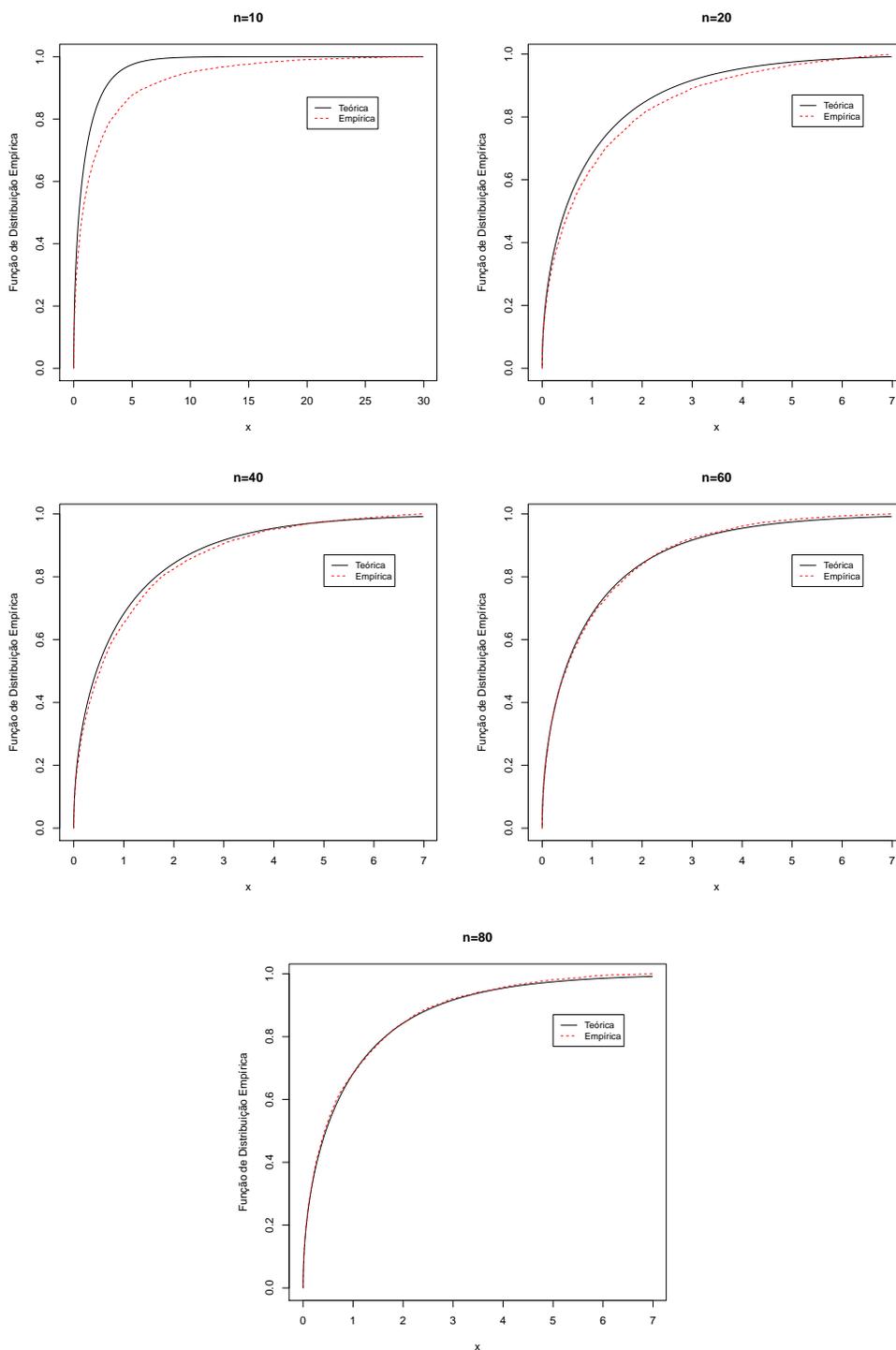


Figura 4.3: Distribuição empírica e teórica da estatística escore, para testar  $H_0$  para o caso (b) sobre a função potência, para diferente tamanhos de amostra.

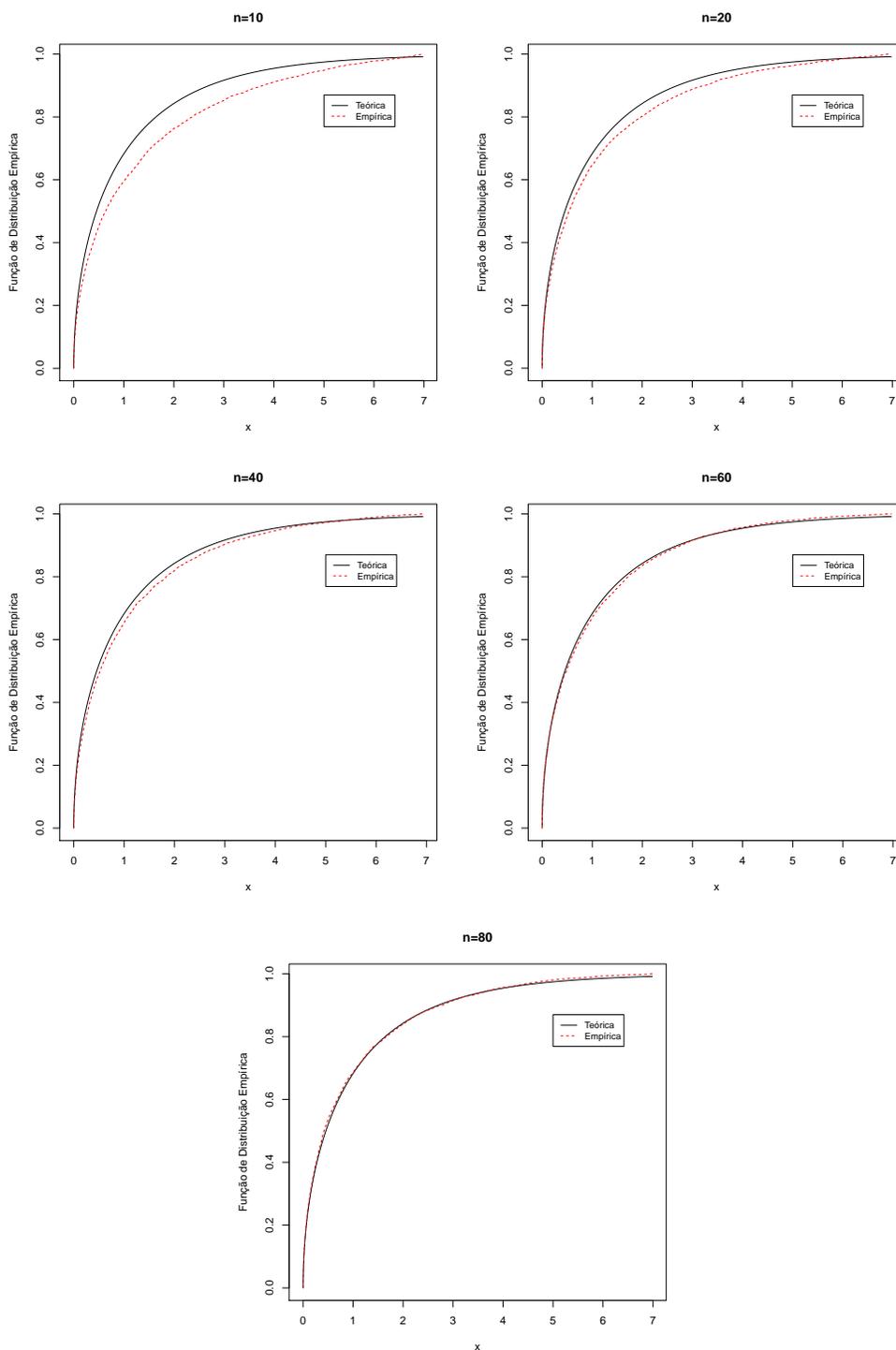


Figura 4.4: Distribuição empírica e teórica da estatística de escore, para testar  $H_0$  para o caso (b) sobre a função exponencial, para diferente tamanhos de amostra.

## 4.4 Aplicação 1: Pesquisa salarial

Nesta seção, usaremos os dados referentes aos dados da pesquisa salarial descritos e analisados por Chatterjee & Hadi (1988). Neste caso, com o intuito de ilustrar a metodologia proposta nas Subseções 4.2.1 e 4.2.2, temos que a variável dependente  $S$  é o salário mensal de uma amostra aleatória de funcionários de uma determinada empresa está relacionada com as seguintes variáveis explicativas: pontos de avaliação do trabalho ( $x_1$ ), sexo (1: masculino e 0: feminino) ( $x_2$ ), número de anos com a empresa ( $x_3$ ), número de anos no trabalho atual ( $x_4$ ) e avaliação de desempenho ( $x_5$ ). Os dados são apresentados no Anexo I, Tabela I.3. Nesta seção, vamos considerar o seguinte modelo de regressão

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, 31,$$

em que  $y_i = \log(S_i)$ ,  $x_{1i} = \log(x_{1i})$ ,  $x_{3i}$ ,  $x_{4i}$ ,  $x_{5i}$ ,  $\varepsilon_i \sim \text{SN}(\alpha_i, 0, \sigma)$  ou  $\varepsilon_i \sim \text{SN}(\alpha, 0, \sigma_i)$ .

### 4.4.1 Teste de heteroscedasticidade

Para fazer o teste de heteroscedasticidade do parâmetro de forma para o modelo de regressão SN com os dados da pesquisa salarial, aplicamos o teste de escore descrito no Teorema 4.2.1 considerando  $\alpha_i = \alpha m(\lambda, x_i)$  em que  $m(\lambda, x_i) = x_i^\lambda$  ou  $m(\lambda, x_i) = \exp(x_i \lambda)$ . Note que, quando  $\lambda = 0$ , temos que  $\alpha_i = \alpha$  para todo  $i$ .

Também vamos considerar o teste de heteroscedasticidade do parâmetro de escala para o modelo SN com os dados da pesquisa salarial, aplicamos o teste de escore descrito no Teorema 4.2.2 considerando  $\sigma_i = \sigma m(\lambda, x_i)$  em que  $m(\lambda, x_i) = x_i^\lambda$  ou  $m(\lambda, x_i) = \exp(x_i \lambda)$ . Note que, quando  $\lambda = 0$ , temos que  $\sigma_i = \sigma$  para todo  $i$ .

Na Tabela 4.3 temos a estatística escore com seus, respectivos, p-valor para as covariáveis  $x_1$ ,  $x_3$ ,  $x_4$  e  $x_5$  dos dados da pesquisa salarial. Observe que, para a covariável  $x_1$  o modelo é heteroscedástico e os modelos com as covariáveis  $x_3$ ,  $x_4$  e  $x_5$  são homocedásticos.

Sob o modelo heteroscedástico com a covariável  $x_1$ , fazemos o ajuste de cada modelo ao conjunto de dados com a finalidade de escolher o melhor modelo. Na Tabela 4.4 apresentamos as EMV com seus, respectivos, erros padrão. Observe que, o modelo com  $\alpha_i = \alpha \exp(x_i \lambda)$  tem erros padrão menores em comparação com os outros modelos.

Na Tabela 4.5 temos os AIC e BIC para os modelos com os parâmetros de forma e escala. Observe que, segundo os valores do AIC e BIC, apresentados na Tabela 4.5, o modelo com melhor ajuste e com  $\alpha_i = \alpha \exp(x_i \lambda)$ .

Tabela 4.3: Estatística de escore (SC) e (p-valor) para os modelos heteroscedásticos dos parâmetros de forma e escala para dados da pesquisa salarial.

Parâmetros	$\alpha_i = \alpha \exp(x_i \lambda)$	$\alpha_i = \alpha x_i^\lambda$	$\sigma_i = \sigma \exp(x_i \lambda)$	$\sigma_i = \sigma x_i^\lambda$
$x_1$	21,4969 (0,0000)	21,18245 (0,0000)	4,8302 (0,0280)	4,2529 (0,0392)
$x_3$	0,7117 (0,3989)	0,7001 (0,4027)	0,6957 (0,4042)	0,6885 (0,4067)
$x_4$	2,0724 (0,1500)	1,0065 (0,3157)	2,2324 (0,1351)	1,1532 (0,2829)
$x_5$	2,1856 (0,1393)	0,8304 (0,3621)	2,3306 (0,1269)	0,9022 (0,3422)

Tabela 4.4: As estimativas de máxima verossimilhança (erros padrão aproximados) para os modelos heteroscedásticos dos parâmetros de forma e escala para dados da pesquisa salarial.

Parâmetros	$\alpha_i = \alpha \exp(x_i \lambda)$	$\alpha_i = \alpha x_i^\lambda$	$\sigma_i = \sigma \exp(x_i \lambda)$	$\sigma_i = \sigma x_i^\lambda$
$\beta_1$	-1,0222 (1,0459)	-1,0962 (1,0713)	-1,0047 (1,0596)	-1,0944 (1,0776)
$\beta_2$	1,4436 (0,1682)	1,4555 (0,1726)	1,4411 (0,1707)	1,4553 (0,1737)
$\alpha$	0,4529 (0,4890)	0,2624 (0,3800)	0,4435 (0,3735)	0,2618 (0,3237)
$\sigma$	2,0331 (1,7844)	2,0344 (2,1364)	2,1507 (2,1635)	2,0498 (2,9530)
$\lambda$	-0,1228 (0,2272)	-0,1065 (0,9916)	-0,1238 (0,2140)	-0,1058 (1,0477)

Tabela 4.5: AIC e BIC para os testes heteroscedasticidade dos parâmetros de forma e escala para dados da pesquisa salarial.

Modelo	AIC	BIC
$\alpha_i = \alpha \exp(x_i \lambda)$	3,0928	10,2627
$\alpha_i = \alpha x_i^\lambda$	4,1499	11,3199
$\sigma_i = \sigma \exp(x_i \lambda)$	3,4462	10,6161
$\sigma_i = \sigma x_i^\lambda$	4,2053	11,3752

Na Figura 4.5, temos os gráficos de probabilidade normal com seus envelopes simulados considerando os resíduos de Pearson versus os índices das observações. Observe que, os gráficos confirmam os resultados baseados nos critérios AIC e BIC, apresentados na Tabela 4.5, em que o modelo da forma  $\alpha_i = \alpha \exp(x_i \lambda)$  tem melhor ajuste ao conjunto de dados.

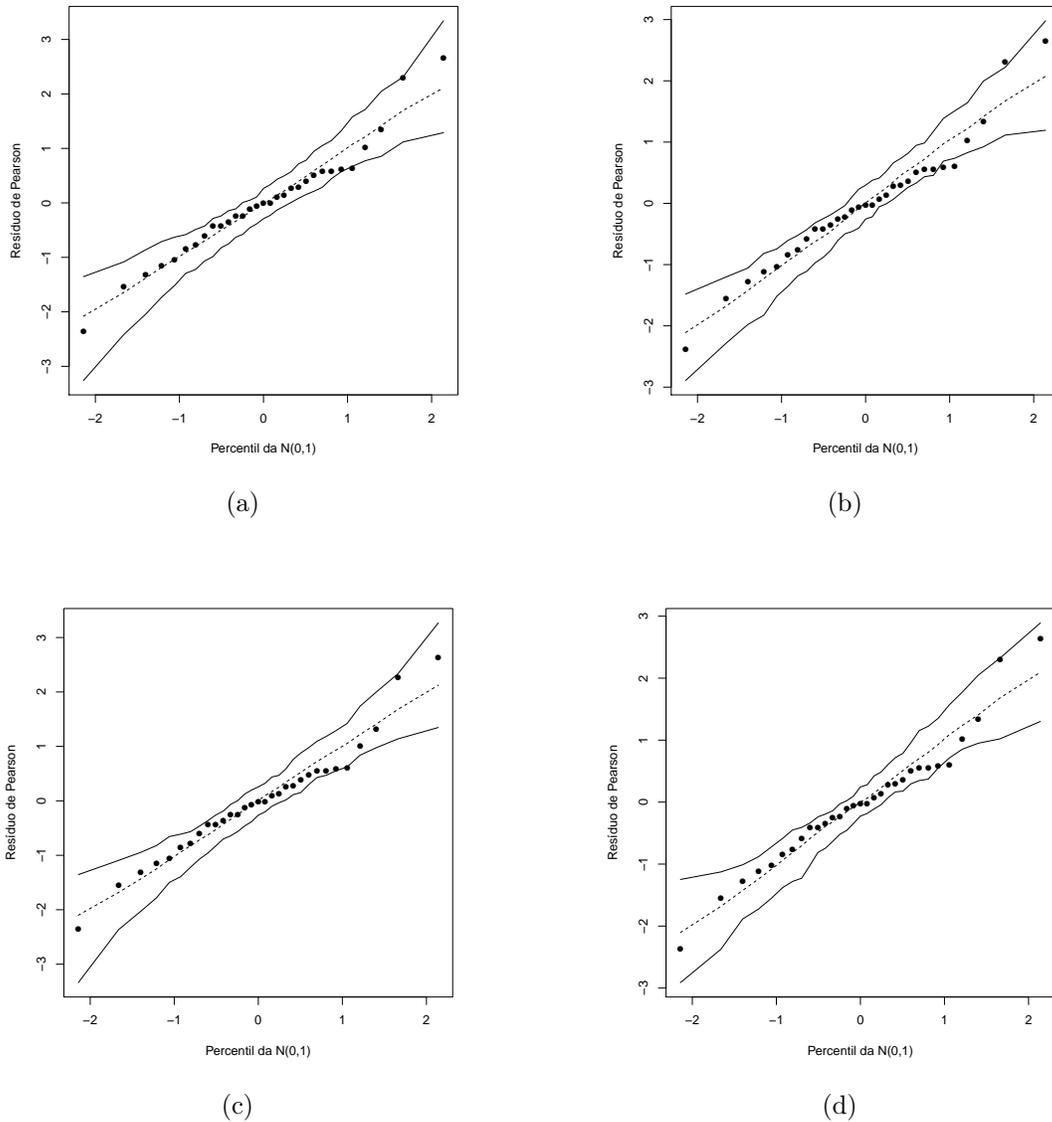


Figura 4.5: Gráfico de probabilidade normal com envelope simulado para os dados da pesquisa salarial. Usando em (a)  $\alpha_i = \alpha \exp(x_i \lambda)$ , (b)  $\alpha_i = \alpha x_i^\lambda$ , (c)  $\sigma_i = \sigma \exp(x_i \lambda)$  e (d)  $\sigma_i = \sigma x_i^\lambda$ .

Na Figura 4.6, temos os gráficos de resíduos de Pearson vs. Observações. Observe que, os resíduos apresentam um comportamento aleatório. Portanto o modelo corrigiu a heteroscedasticidade.

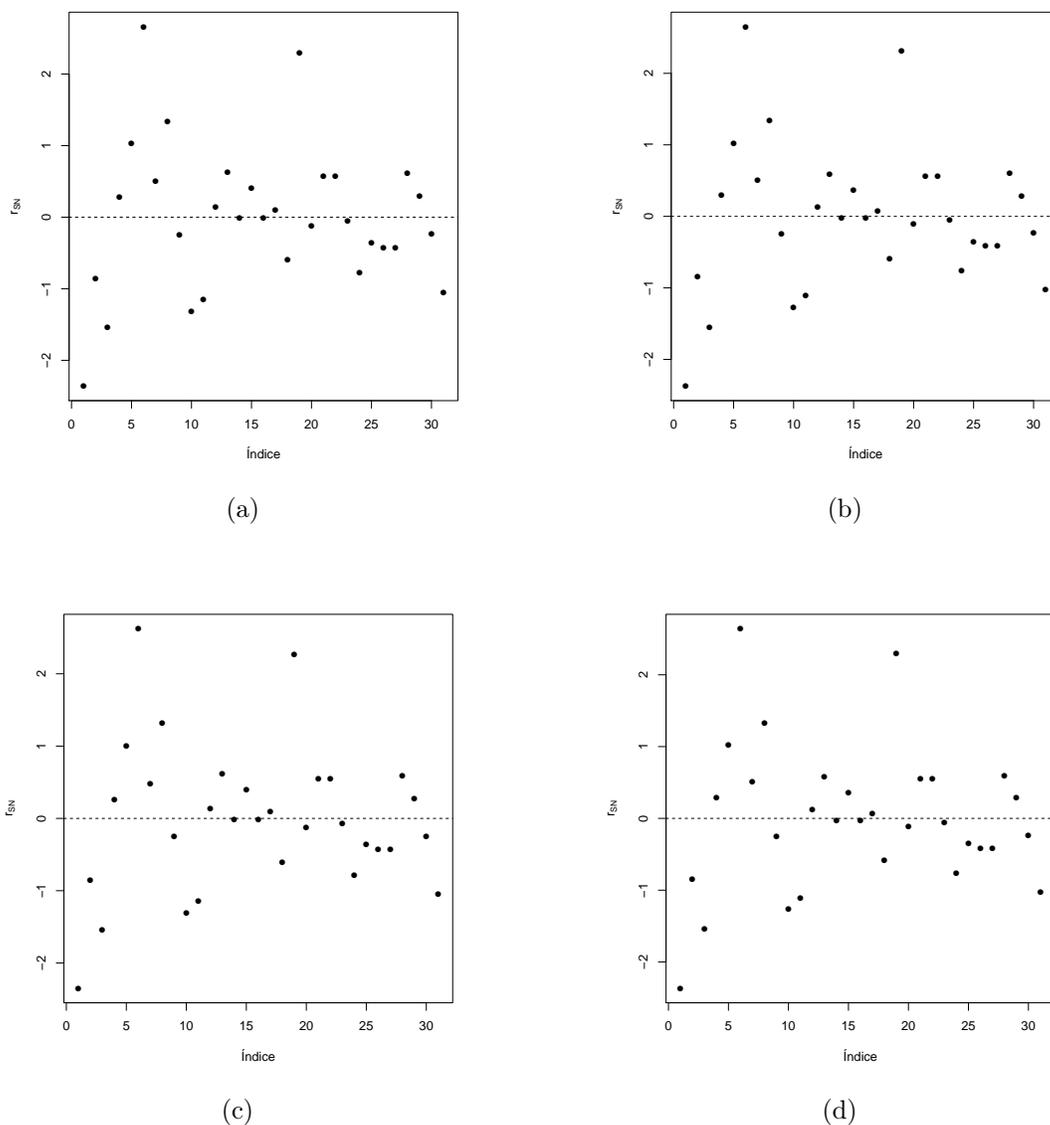


Figura 4.6: Gráfico de resíduos para os dados da pesquisa salarial usando o resíduo de Pearson. Usando em. (a)  $\alpha_i = \alpha \exp(x_i \lambda)$ , (b)  $\alpha_i = \alpha x_i^\lambda$ , (c)  $\sigma_i = \sigma \exp(x_i \lambda)$  e (d)  $\sigma_i = \sigma x_i^\lambda$ .

Na Figura 4.7, temos os envelopes simulados para os modelos heteroscedásticos. Observe que, novamente os gráficos confirmam os resultados obtidos baseado nos critérios AIC e BIC, apresentados na Tabela 4.5, em que o modelo da forma  $\alpha_i = \alpha \exp(x_i \lambda)$  tem melhor ajuste ao conjunto de dados.

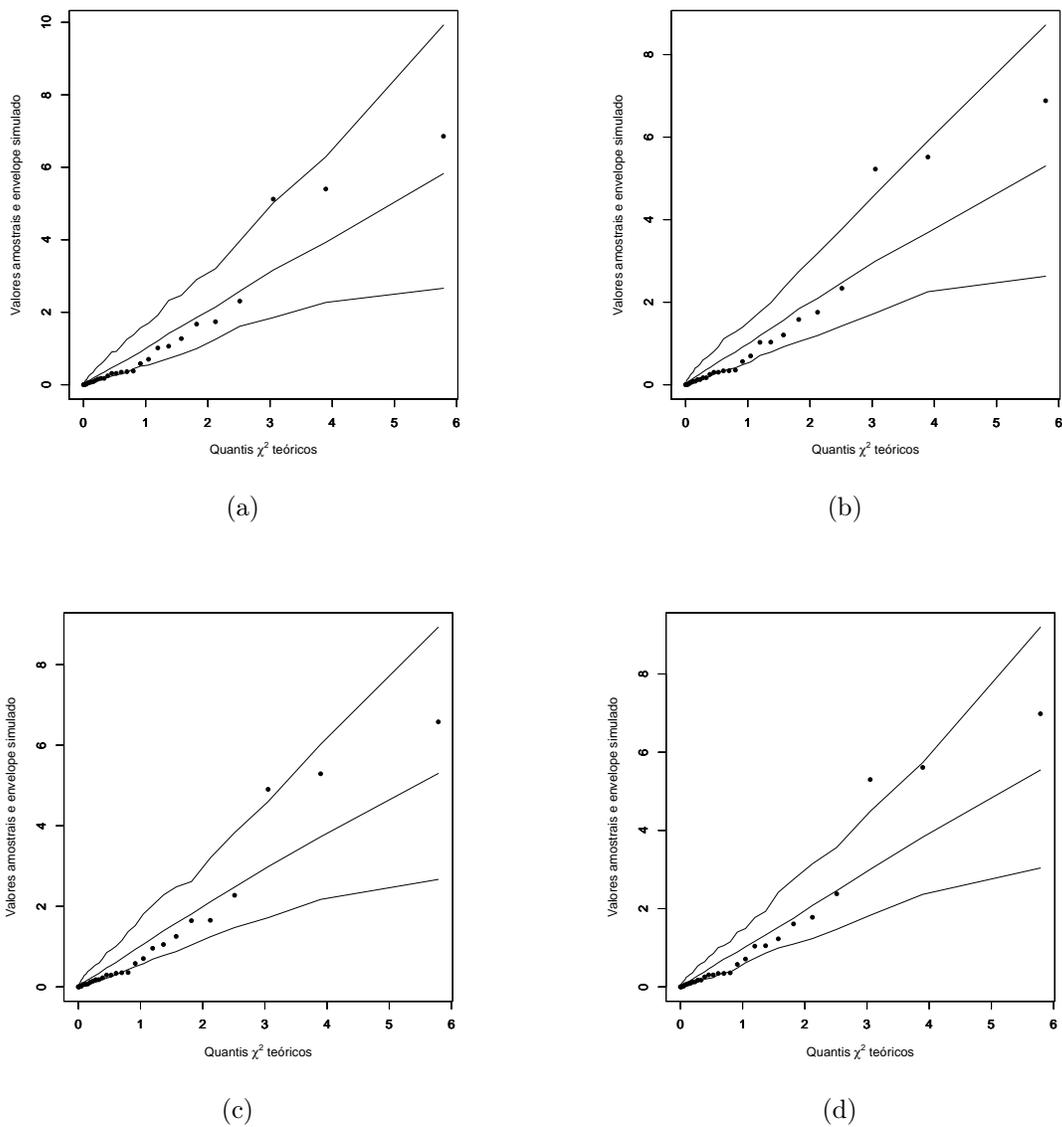


Figura 4.7: Envelope simulado para os dados da pesquisa salarial. Usando em (a)  $\alpha_i = \alpha \exp(x_i \lambda)$ , (b)  $\alpha_i = \alpha x_i^\lambda$ , (c)  $\sigma_i = \sigma \exp(x_i \lambda)$  e (d)  $\sigma_i = \sigma x_i^\lambda$ .

## 4.5 Aplicação 2: Fadiga biaxial

Com o intuito de ilustrar a metodologia proposta nas Subseções 4.2.1 e 4.2.2 utilizamos o conjunto de dados utilizado na Seção 3.7, que correspondem a 46 peças de metal que foram submetidas à fadiga em ciclos. O objetivo do estudo foi verificar o número de ciclos até que ocorresse a falha. A variável resposta  $N$  é o número de ciclos até que ocorra a falha e a variável explicativa  $z$  é o trabalho por ciclo ( $mJ/m^3$ ). Nesta seção, vamos considerar o seguinte modelo de regressão

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, 31, to$$

em que  $y_i = \log(N_i)$ ,  $x_i = \log(z_i)$  e  $\varepsilon_i \sim \text{SN}(\alpha_i, 0, \sigma_i)$ .

Na Tabela 4.6, temos as EMV dos parâmetros, com seus respectivos erros padrão (entre parêntesis). Observe que, para os quatro modelos apresentados, obtemos estimativas similares tanto para o parâmetro de forma como para o parâmetro de escala.

Tabela 4.6: As estimativas de máxima verossimilhança (erros padrão aproximados) para os modelos heteroscedásticos dos parâmetros de forma e escala para dados da fadiga biaxial.

Parâmetros	$\alpha_i = \alpha \exp(x_i \lambda)$	$\alpha_i = \alpha x_i^\lambda$	$\sigma_i = \sigma \exp(x_i \lambda)$	$\sigma_i = \sigma x_i^\lambda$
$\beta_1$	12,279 (0,390)	12,279 (0,385)	12,2791 (0,3909)	12,2791 (0,3904)
$\beta_2$	-1,671 (0,109)	-1,671 (0,107)	-1,6705 (0,1088)	-1,6705 (0,1085)
$\alpha$	0,424 (0,362)	0,419 (0,251)	0,4240 (0,3151)	0,4240 (0,3458)
$\sigma$	1,915 (1,900)	1,937 (1,339)	1,9152 (1,3032)	1,9152 (1,4007)
$\lambda$	-0,0008 (0,172)	-0,010 (0,510)	-0,0022 (0,1581)	-0,0050 (0,5352)

Na Tabela 4.7, temos as estatísticas score com seus, respectivos, p-valores e critérios AIC e BIC. Observe que, para este conjunto de dados o modelo é homocedástico, o qual é avaliado pelo teste de heterocedasticidade que não rejeita a hipótese nula descrita na Subseção 4.2.1. Nota-se, que os resultados fazendo uso dos critérios AIC e BIC apresentam resultados similares, sendo melhor para o modelo  $\sigma_i = \sigma m_i$  com função exponencial.

Tabela 4.7: Estatística de escore (E), p-valor, AIC e BIC para os testes heteroscedasticidade dos parâmetros de forma e escala para dados da fadiga biaxial.

Modelo	SC	p-valor	AIC	BIC
$\alpha_i = \alpha \exp(x_i \lambda)$	0,4990	0,4799	56,6772	65,8204
$\alpha_i = \alpha x_i^\lambda$	0,0551	0,8144	56,7003	65,8435
$\sigma_i = \sigma \exp(x_i \lambda)$	2,7505	0,0972	56,6724	65,8156
$\sigma_i = \sigma x_i^\lambda$	0,0631	0,8017	56,6743	65,8175

Na Figura 4.8, temos o gráfico de probabilidade normal com seu envelope simulado para o modelo homocedástico. Observe que, o modelo apresenta um bom ajuste ao conjunto de dados.

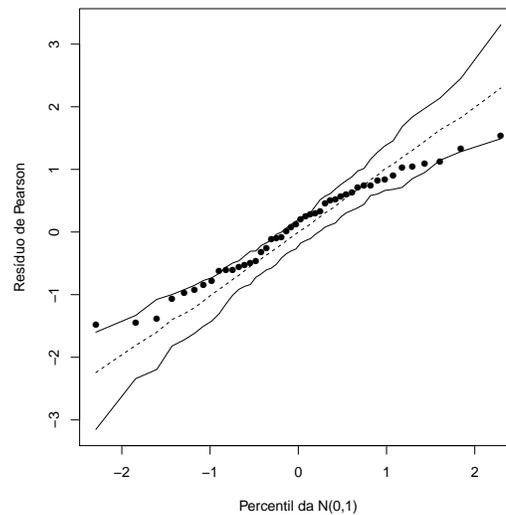


Figura 4.8: Gráfico de probabilidade normal com envelope simulado para os dados da fadiga biaxial sob modelo homocedástico.

Na Figura 4.9, temos o gráfico de resíduos de Pearson contra o índice das observações. Observe que, os resíduos apresentam um comportamento aleatório ao longo das observações. Portanto, as informações obtidas indicam que o modelo homocedástico é o mais apropriado.

Na Figura 4.10, temos o envelopes simulado para o modelos homocedástico. Observe que, um modelo homocedástico apresenta um bom ajuste para o conjunto de dados.

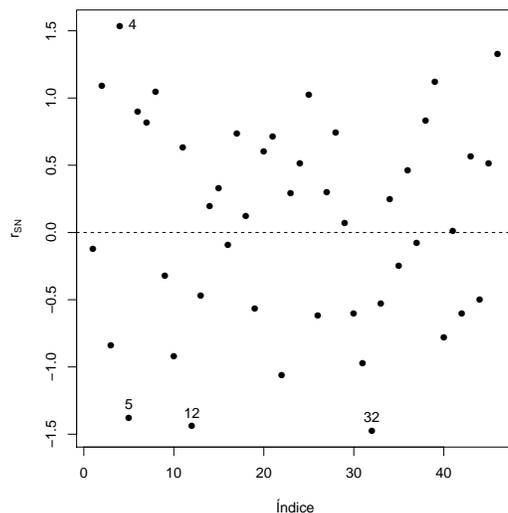


Figura 4.9: Gráfico de resíduos de Pearson para os dados da fadiga biaxial sob modelo homocedástico.

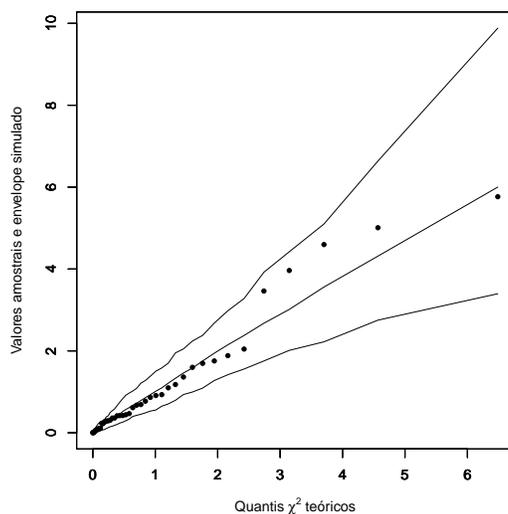


Figura 4.10: Envelope simulado para os dados da fadiga biaxial sob modelo homocedástico.

Embora o conjunto de dados tem um bom ajuste para um modelo linear, lembre-se que na Seção 3.7 concluímos que um modelo não-linear homocedástico tem um melhor ajuste de acordo com os critérios AIC e BIC. O que também foi avaliado com os diagramas de dispersão com o valor médio previsto e as bandas de confiança.

Os resultados obtidos são avaliados no artigo Xie & Wei (2007), em que obtiveram o teste de

escore para o mesmo conjunto de dados, sendo homocedástico para o modelo da forma  $\alpha_i = \alpha m_i$ .



# Capítulo 5

## Considerações Finais

Neste trabalho, consideramos modelos de regressão onde os erros seguem uma distribuição Sinh-Normal. Foi dedicada uma atenção especial a estimação, análise de diagnóstico e teste de hipóteses.

No Capítulo 2, apresentamos a distribuição Sinh-Normal com suas respectivas propriedades e distribuições relacionadas.

No Capítulo 3, descrevemos o modelo de regressão Sinh-Normal, modelando a média da variável resposta que relaciona um preditor linear, baseado em um conjunto de variáveis explicativas conhecidas e parâmetros desconhecidos, por meio de uma função de ligação, seguindo a linha de estudo de Villegas *et al.* (2013), que apresentou modelos de regressão simétricos sob as classes de distribuições elípticas. Também realizamos um estudo de diagnóstico seguindo a metodologia de Cook (1986) e as ideias de Villegas *et al.* (2013).

No Capítulo 4, apresentamos um teste de homogeneidades dos parâmetros de forma e escala no modelos de regressão Sinh-Normal usando a estatística de escore. Através de simulações de Monte Carlo examinamos o desempenho das propriedades do teste um exemplo numérico foi apresentado para ilustrar a metodologia.

A principal dificuldade, foi na obtenção das EMV num modelo heterocedástico, porque os parâmetros  $\beta$  e  $\lambda$  não tem forma fechada e os outros parâmetros dependem deles. Portanto, se o número de parâmetros é grande é melhor trabalhar com um modelo homocedástico.

Um resultado interessante, foi obtido com o conjunto de dados da fadiga biaxial, em que foi obtido o mesmo resultado que em Lemonte & Cordeiro (2010), um modelo não linear tem o melhor ajuste ao conjunto de dados. O que sugere como trabalho futuro, teste de homogeneidade dos parâmetros de forma e escala para este modelo não linear.



# Apêndice A

## Programas em R

### A.1 Distribuição Sinh-Normal

```
dsn<-function(y,alpha,mu,sigma)
{
  if(!is.numeric(y)||!is.numeric(alpha)||!is.numeric(mu)||!is.numeric(sigma))
  {stop("non-numeric argument to mathematical function")}
  if(alpha<=0){stop("alpha must be positive")}
  if(sigma<=0){stop("sigma must be positive")}
  y<-y
  p<- -2/(alpha^2)
  c<-cosh((y-mu)/sigma)
  s<-(sinh((y-mu)/sigma))^2
  pdf<-(2/(alpha*sigma*sqrt(2*pi)))*c*exp(p*s)
  return(pdf)
}
```

### A.2 Distribuição Birbaum-Saunders generalizada

```
dbs<-function(y,alpha,beta,sigma)
{
  if(!is.numeric(y)||!is.numeric(alpha)||!is.numeric(beta)||!is.numeric(sigma))
  {stop("non-numeric argument to mathematical function")}
  if(alpha<=0){stop("alpha must be positive")}
  if(sigma<=0){stop("sigma must be positive")}
  y<-y
  at=(1/alpha)*((y/beta)^(1/sigma)-(beta/y)^(1/sigma))
  z=1/sigma
  At=(y^(-z-1)/(sigma*alpha*beta^z))*(y^(2*z)+beta^(2*z))
  pdf<-dnorm(at)*At
}
```

```
return(pdf)  
}
```

# Anexo I

## Conjuntos de dados

Tabela I.1: Fadiga biaxial

Peça	X ( $mJ/m^3$ )	Ciclos de vida	Peça	X ( $mJ/m^3$ )	Ciclos de vida
1	11,5	3280	24	34,5	736
2	13	5046	25	40,1	750
3	14,3	1563	26	40,1	316
4	15,6	4707	27	43	456
5	16	977	28	44,1	552
6	17,3	2834	29	46,5	355
7	19,3	2266	30	47,3	242
8	21,1	2208	31	48,7	190
9	21,5	1040	32	52,9	127
10	22,6	700	33	56,9	185
11	22,6	1583	34	59,9	255
12	24	482	35	60,2	195
13	24	804	36	60,3	283
14	24,6	1093	37	60,5	212
15	25,2	1125	38	62,1	327
16	25,5	884	39	62,8	373
17	26,3	1300	40	66,5	125
18	27,9	852	41	67	187
19	28,3	580	42	67,1	135
20	28,4	1066	43	67,9	245
21	28,6	1114	44	68,8	137
22	30,9	386	45	75,4	200
23	31,9	745	46	100,5	190

Tabela I.2: Produção de cebola

Número	Densidade	Produção	Número	Densidade	Produção
1	22,3	148,57	22	49,78	99,85
2	25,86	125,3	23	50,43	111,65
3	29,09	150,69	24	51,72	98,09
4	29,74	147,42	25	61,42	87,85
5	31,68	117,1	26	65,29	75,45
6	31,68	116,64	27	67,23	87,01
7	32	129,66	28	71,44	90,1
8	32,32	131,54	29	73,05	81,08
9	32,32	151,5	30	86,63	65,33
10	34,91	121,8	31	96	58,49
11	35,23	125,67	32	98,91	65,67
12	38,47	117,78	33	103,44	67,19
13	39,44	101,5	34	105,05	54,01
14	41,05	113,22	35	111,19	60,92
15	41,7	136,43	36	113,78	53,48
16	44,28	117,54	37	119,92	61,62
17	45,9	87,2	38	120,89	26,32
18	46,55	107,41	39	126,71	61,21
19	48,16	129,68	40	138,99	41,67
20	48,49	104,63	41	146,75	45,26
21	48,81	114,15	42	160,97	46,45

Tabela I.3: Pesquisa salarial

Número	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$y$
1	350	1	2	2	5	1000
2	350	1	5	5	5	1400
3	350	0	4	4	4	1200
4	350	1	20	20	1	1800
5	425	0	10	2	3	2800
6	425	1	15	10	3	4000
7	425	0	1	1	4	2500
8	425	1	5	5	4	3000
9	600	1	10	5	2	3500
10	600	0	8	8	3	2800
11	600	0	4	3	4	2900
12	600	1	20	10	2	3800
13	600	1	7	7	5	4200
14	700	1	8	8	1	4600
15	700	0	25	15	5	5000
16	700	1	19	16	4	4600
17	700	0	20	14	5	4700
18	400	0	6	4	3	1800
19	400	1	20	8	3	3400
20	400	0	5	3	5	2000
21	500	1	22	12	3	3200
22	500	1	25	10	3	3200
23	500	0	8	3	4	2800
24	500	0	2	1	5	2400
25	800	1	10	10	3	5200
26	475	1	10	4	3	2400
27	475	0	3	3	4	2400
28	475	1	8	8	2	3000
29	475	1	6	6	4	2800
30	475	0	12	4	3	2500
31	475	0	4	2	5	2100



# Anexo II

## Licença

Copyright (c) 2014 de Rocío Paola Maehara Sánchez.

Exceto quando indicado o contrário, esta obra está licenciada sob a licença Creative Commons Atribuição 3.0 Não Adaptada. Para ver uma cópia desta licença, visite <http://creativecommons.org/licenses/by/3.0/>.



A marca e o logotipo da UNICAMP são propriedade da Universidade Estadual de Campinas. Maiores informações sobre encontram-se disponíveis em <http://www.unicamp.br/unicamp/a-unicamp/logotipo/normas%20oficiais-para-uso-do-logotipo>.

### II.1 Sobre a licença dessa obra

A licença Creative Commons Atribuição 3.0 Não Adaptada utilizada nessa obra diz que:

1. Você tem a liberdade de:

- Compartilhar — copiar, distribuir e transmitir a obra;
- Remixar — criar obras derivadas;
- fazer uso comercial da obra.

2. Sob as seguintes condições:

- Atribuição — Você deve creditar a obra da forma especificada pelo autor ou licenciante (mas não de maneira que sugira que estes concedem qualquer aval a você ou ao seu uso da obra).



# Referências Bibliográficas

- Abramowitz, M. & Stegun, I. A. (1972). *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables*. Dover, New York.
- Achar, J. A. (1993). Inference for birnbaum–saunders fatigue life model using bayesian method. *Computational Statistics and Data Analysis*, **15**, 367–380.
- Aitkin, M. (1987). Modelling variance heterogeneity in normal regression using glim. *Applied Statistics*, **36**, 332–339.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *International Symposium on Information Theory. Akadémiai Kiadó, Budapest, Hungary*, pages 267–281.
- Athayde, E., Azevedo, C., L. V. & Sanhueza, A. (2012). About birnbaum-saunders distributions based on the johnson system. *Communications in Statistics - Theory and Methods*, **41**, 2061–2079.
- Atkinson, A. (1981). Two graphical displays for outlying and influential observations in regression. *Biometrika*, **68**, 13.
- Atkinson, A. C. (1985). *Plots, Transformation and Regression*. Clarendon Press: Oxford.
- Barroso, L. P. & Cordeiro, G. M. (2005). Bartlett corrections in heteroskedastic  $t$  regression models. *Statistics and Probability Letters*, **75**, 86–96.
- Birnbaum, Z. W. & Saunders, S. C. (1969b). Estimation for a family of life distributions with applications to fatigue. *Journal of Applied Probability*, **6**, 328–347.
- Birnbaum, Z. W. & Saunders, S. C. (1969a). A new family of life distributions. *J. Appl. Prob.*, **6**, 637–652.
- Cancho, V. G., Ortega, E. E. M. & Paula, G. A. (2010). On estimation and influence diagnostics for log-birnbaum–saunders student- $t$  regression models: Full bayesian analysis. *Journal of Statistical Planning and Inference*, **140**, 2486–2496.
- Chatterjee, S. & Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*. Wiley, New York.

- Chen, C. F. (1983). Score test for regression models. *Journal of the American Statistical Association*, **78**, 158–161.
- Christensen, R., Pearson, L. & Johnson, W. (1992). Case-deletion diagnostics for mixed models. *Technometrics*, **34**, 38–45.
- Cook, R. (1977). Detection of influential observation in linear regression. *Technometrics*, **19**, 15–18.
- Cook, R. & Weisberg, S. (1982). *Residuals and Influence in Regression*. Chapman & Hall, Boca Raton, FL.
- Cook, R. & Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*, **70**, 1–10.
- Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, **48**, 133–169.
- Cysneiros, F., Paula, G. & Galea, M. (2007). Heteroscedastic symmetrical linear models. *Statistics and Probability Letters*, **77**, 1084–1090.
- Cysneiros, F., Cordeiro, G. & Cysneiros, A. (2010). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation*, **80**, 451–461.
- Davison, A. C. & Tsai, C. L. (1992). Regression model diagnostics. *International Statistical Review*, **60**, 337–355.
- Edgeworth, F. Y. (1898). On the representation of statistics by mathematical formulae. *Journal of the Royal Statistical Society*, **61**, 670–700.
- Galea, M., Leiva, V. & Paula, G. A. (2004). Influence diagnostics in log-birnbaum-saunders regression models. *Journal of Applied Statistics*, **31**, 1049–1064.
- Galea, M., Paula, G. A. & Cysneiros, F. J. (2005). On diagnostics in symmetrical nonlinear models. *Statistics and Probability Letters*, **73**, 459–467.
- Harvey, A. C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, **41**, 461–465.
- Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 149–176.
- Johnson, N. L. & Kotz, S. (1970). *Continuous Univariate Distributions - 1 e 2*. Boston: Houghton Mifflin Company.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Distributions in Statistics: Continuous Univariate Distributions*, volume 2. John Wiley and Sons.

- Lachos, V. H. (2002). *Inferência e Diagnóstico em Modelos de Grubbs*. Master's thesis, Instituto De Matemática, Estatística e Computação Científica, Campinas, Brasil.
- Leiva, V., Barros, M., Paula, G. A. & Galea, M. (2007). Influence diagnostics in log-birnbaum-saunders regression models with censored data. *Computational Statistics and Data Analysis*, **51**, 5694–5707.
- Leiva, V., Vilca, F., Balakrishnan, N. & Sanhueza, A. (2010). A skewed sinh-normal distribution and its properties and application to air pollution. *Communications in Statistics - Theory and Methods*, **30**, 426–443.
- Lemonte, A. J. & Cordeiro, G. M. (2009). Birnbaum-saunders nonlinear regression models. *Computational Statistics and Data Analysis*, **53**, 4441–4452.
- Lemonte, A. J. & Cordeiro, G. M. (2010). Asymptotic skewness in birnbaum-saunders nonlinear regression models. *Statistics and Probability Letters*, **80**, 892–898.
- Ng, H. K. T., Kundu, D. & Balakrishnan, N. (2003). Modified moment estimation for the two-parameter birnbaum-saunders distribution. *Computational Statistics and Data Analysis*, **43**, 283–298.
- Ortega, E. M. M., Bolfarine, H. & Paula, G. A. (2003). Influence diagnostics in generalized log-gamma regression models. *Computational Statistics and Data Analysis*, **42**, 165–186.
- Park, R. E. (1966). Estimating with heteroscedastic error terms. *Econometrica*, **34**, 888.
- Patel, J. K. & Read, C. B. (1996). *Handbook of the Normal Distribution*, volume 150. Marcel Dekker, INC.
- Poon, W. Y. & Poon, Y. S. (1999). Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society B*, **61**, 51–61.
- Ratkowsky, D. A. (1983). *Nonlinear Regression Modeling: A United Practical Approach*. Marcel Dekker, New York.
- Rieck, J. R. (1989). *Statistical Analysis for the Birnbaum Saunders fatigue life distribution*. Ph.D. thesis, Clemson University.
- Rieck, J. R. & Nedelman, J. R. (1991). A log-linear model for the birnbaum-saunders distribution. *Technometrics*, **33**, 51–60.
- Rietz, H. L. (1922). Frequency distributions obtained by certain transformations of normally distributed variables. *Ann. Matn.*, pages 23–292.
- Rolim Santana, L. (2009). *Estimação e Diagnóstico em Modelos Birnbaum-Saunders Skew-Normal*. Master's thesis, Instituto De Matemática, Estatística e Computação Científica, Campinas, Brasil.

- Sanhueza, A., Leiva, V. & Balakrishnan, N. (2008). The generalized birnbaum–saunders distribution and its theory, methodology and application. *Communications in Statistics - Theory and Methods*, **37**, 645–670.
- Santos dos Reis, S. (2013). *Diagnóstico em Modelos de Regressão Linear e Não-linear com Erros Simétricos*. Master's thesis, Instituto De Matemática, Estatística e Computação Científica, Campinas, Brasil.
- Schwarz, C. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.
- Seber, G. A. (1984). *Multivariate Observations*. John Wiley Sons, New York.
- Simonoff, J. S. & Tsai, C. L. (1994). Use of modified profile likelihood for improved tests of constancy of variance in regression. *Applied Statistics*, **43**, 357–370.
- Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society*, **51**, 47–60.
- Tang, N. S., Wei, B. C. & Wang, X. R. (2000). Influence diagnostics in nonlinear reproductive dispersion models. *Statistics and Probability Letters*, **26**, 59–68.
- Tisonas, E. G. (2001). Bayesian inference in birnbaum–saunders regression. *Communications in Statistics - Theory and Methods*, **30**, 179–193.
- Tsai, C. L. (1986). Score test for the rst-order autoregressive model with heteroscedasticity. *Biometrika*, **73**, 455–460.
- Verveke, G. & Molenberghs, G. (1997). *Linear Mixed Models in Practice*. Springer-Verlag, New York.
- Vilca, F., Sanhueza, A., Leiva, V. & Christakos, G. (2010). An extended birnbaum saunders model and its application in the study of environmental quality in santiago, chile. *Stochastic Environmental Research and Risk Assessment*, **24**, 771–782.
- Vilca, F., A., G., Lachos, V. H. & M, O. E. M. (2012). Estimation and diagnostics for heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *Journal of Statistical Planning and Inference*, **142**, 2149–2165.
- Villegas, C., Paula, G., Cysneros, F. & Galea, M. (2013). Influence diagnostics in generalized symmetric linear models. *Computational Statistics and Data Analysis*, **59**, 161–170.
- Wang, Z. H., Desmond, A. F. & Lu, X. W. (2006). Modified censored moment estimation for the two-parameter birnbaum–saunders distribution. *Computational Statistics and Data Analysis*, **4**, 1033–1051.
- Wei, B. C. (1998). *Exponential Family Nonlinear Models*. Springer, Singapore.
- Wicksell, S. D. (1917). On the genetic theory of frequency. *Ark. Mat. Astr. Fys.*, **12**, 1–56.

- Xiao, Q., Liu, Z., Balakrishnan, N. & Lu, X. (2010). Estimation of the birnbaum–saunders regression model with current status data. *Computational Statistics and Data Analysis*, **54**, 326–332.
- Xie, F. C. & Wei, B. C. (2007). Diagnostics analysis for log-birnbaum-saunders regression models. *Computational Statistics and Data Analysis*, **51**, 4692–4706.
- Xie, F. C., Wei, B. C. & Lin, J. G. (2009). Homogeneity diagnostics for skew-normal nonlinear regression models. *Statistics and Probability Letters*, **79**, 821–827.
- Zhu, H. & Lee, S. (2001). Local influence for incomplete data models. *Journal of the Royal Statistical Society B*, **63**, 111–126.