

---

Uso dos métodos clássico e bayesiano para os  
modelos não-lineares heterocedásticos simétricos

*Márcia Aparecida Centanin Macera*

---

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

# Uso dos métodos clássico e bayesiano para os modelos não-lineares heterocedásticos simétricos

**Márcia Aparecida Centanin Macera**

***Orientador:* Prof. Dr. Marinho Gomes de Andrade Filho**

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências - Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA.*

**USP – São Carlos**  
**Agosto de 2011**

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi  
e Seção Técnica de Informática, ICMC/USP,  
com os dados fornecidos pelo(a) autor(a)

M142u Macera, Márcia Aparecida Centanin  
Uso dos métodos clássico e bayesiano para os  
modelos não-lineares heterocedásticos simétricos /  
Márcia Aparecida Centanin Macera; orientador Marinho  
Gomes de Andrade Filho -- São Carlos, 2011.  
70 p.

Dissertação (Mestrado - Programa de Pós-Graduação em  
Ciências de Computação e Matemática Computacional) --  
Instituto de Ciências Matemáticas e de Computação,  
Universidade de São Paulo, 2011.

1. Modelos simétricos. 2. Modelos não-lineares  
heterocedásticos. 3. MCMC. 4. Divergência de  
Kullback-Leibler. I. Filho, Marinho Gomes de  
Andrade, orient. II. Título.

*Aos meus pais  
com gratidão,*

*Ao meu namorado  
Rafael*

*com eterna paixão,*

*dedico esse trabalho com amor e carinho.*

## Agradecimentos

Aos meus pais, que sempre me estimularam a dar esse grande passo. Estas duas pessoas com muita sabedoria, discernimento e dedicação estiveram ao meu lado encorajando-me nas horas difíceis e aplaudindo-me nos momentos de glória, portanto têm minha eterna gratidão.

Ao meu namorado Rafael, pelo incentivo, amor, compreensão e atenção. Agradeço por sempre ter uma palavra de conforto e estímulo nos momentos difíceis. Obrigada também por toda a ajuda nas correções e, acima de tudo, por estar sempre ao meu lado.

Ao Prof. Marinho Gomes de Andrade, pela orientação dessa dissertação e também pela orientação de iniciação científica durante minha graduação. Obrigada por todo seu empenho, sabedoria e compreensão com que me conduziu durante todo esse tempo. Agradeço ainda por ter acreditado em mim, pelo conhecimento que adquiri com você e por todo incentivo no decorrer do trabalho.

Aos professores de graduação e pós-graduação, em especial aos professores Mário de Castro, Vicente Garibay Cancho e Dorival Leão pela atenção e ajuda nos momentos que precisava.

Aos funcionários do ICMC pela dedicação e convívio.

À todos os meus amigos, de graduação e pós-graduação, pela amizade e pelos momentos compartilhados.

À todos aqueles que, direta ou indiretamente, tive o prazer de trocar conhecimentos para que essa dissertação atingisse seus objetivos.

Finalmente, agradeço à CNPq pelo apoio financeiro durante o desenvolvimento desse trabalho.

## Resumo

Os modelos normais de regressão têm sido utilizados durante muitos anos para a análise de dados. Mesmo nos casos em que a normalidade não podia ser suposta, tentava-se algum tipo de transformação com o intuito de alcançar a normalidade procurada. No entanto, na prática, essas suposições sobre normalidade e linearidade nem sempre são satisfeitas. Como alternativas à técnica clássica, foram desenvolvidas novas classes de modelos de regressão. Nesse contexto, focamos a classe de modelos em que a distribuição assumida para a variável resposta pertence à classe de distribuições simétricas. O objetivo geral desse trabalho é a modelagem desta classe no contexto bayesiano, em particular a modelagem da classe de modelos não-lineares heterocedásticos simétricos. Vale ressaltar que esse trabalho tem ligação com duas linhas de pesquisa, a saber: a inferência estatística abordando aspectos da teoria assintótica e a inferência bayesiana considerando aspectos de modelagem e critérios de seleção de modelos baseados em métodos de simulação de Monte Carlo em Cadeia de Markov (MCMC). Uma primeira etapa consiste em apresentar a classe dos modelos não-lineares heterocedásticos simétricos bem como a inferência clássica dos parâmetros desses modelos. Posteriormente, propomos uma abordagem bayesiana para esses modelos, cujo objetivo é mostrar sua viabilidade e comparar a inferência bayesiana dos parâmetros estimados via métodos MCMC com a inferência clássica das estimativas obtidas por meio da ferramenta GAMLSS. Além disso, utilizamos o método bayesiano de análise de influência caso a caso baseado na divergência de Kullback-Leibler para detectar observações influentes nos dados. A implementação computacional foi desenvolvida no software R e para detalhes dos programas pode ser consultado aos autores do trabalho.

## Abstract

The normal regression models have been used for many years for data analysis. Even in cases where normality could not be assumed, was trying to be some kind of transformation in order to achieve the normality sought. However, in practice, these assumptions about normality and linearity are not always satisfied. As alternatives to classical technique new classes of regression models were developed. In this context, we focus on the class of models in which the distribution assumed for the response variable belongs to the symmetric distributions class. The aim of this work is the modeling of this class in the bayesian context, in particular the modeling of the nonlinear models heteroscedastic symmetric class. Note that this work is connected with two research lines, the statistical inference addressing aspects of asymptotic theory and the bayesian inference considering aspects of modeling and criteria for models selection based on simulation methods Monte Carlo Markov Chain (MCMC). A first step is to present the nonlinear models heteroscedastic symmetric class as well as the classic inference of parameters of these models. Subsequently, we propose a bayesian approach to these models, whose objective is to show their feasibility and compare the estimated parameters bayesian inference by MCMC methods with the classical inference of the estimates obtained by GAMLSS tool. In addition, we use the bayesian method of influence analysis on a case based on the Kullback-Leibler divergence for detecting influential observations in the data. The computational implementation was developed in the software R and programs details can be found at the study's authors.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação e definição dos objetivos . . . . .	1
1.2	Apresentação dos capítulos . . . . .	2
<b>2</b>	<b>Modelos Simétricos</b>	<b>3</b>
2.1	Classe dos modelos simétricos . . . . .	5
2.2	Distribuição Normal . . . . .	6
2.3	Distribuição Normal Generalizada . . . . .	7
2.4	Distribuição $t$ de Student . . . . .	10
2.5	Distribuição Logística-II . . . . .	11
<b>3</b>	<b>Modelo de regressão simétrico: Abordagem clássica</b>	<b>14</b>
3.1	Modelos não-lineares simétricos . . . . .	14
3.1.1	Comentários a respeito dos parâmetros $s$ e $\nu$ . . . . .	18
3.2	Modelos não-lineares heterocedásticos simétricos . . . . .	19
3.2.1	Testes de heterocedasticidade . . . . .	22



3.3	Métodos de diagnóstico . . . . .	23
3.3.1	Resíduos . . . . .	24
3.3.2	Critérios de Seleção . . . . .	26
<b>4</b>	<b>Resultados da abordagem clássica</b>	<b>27</b>
4.1	Aplicação . . . . .	27
4.2	Seleção do modelo . . . . .	30
<b>5</b>	<b>Abordagem Bayesiana</b>	<b>39</b>
5.1	A função de verossimilhança . . . . .	40
5.2	Distribuição <i>a priori</i> . . . . .	41
5.3	Distribuição <i>a posteriori</i> . . . . .	42
5.4	Critérios de seleção dos modelos . . . . .	43
5.4.1	Critério de Informação Bayesiano (ou Schwarz) Esperado (EBIC)	43
5.4.2	Critério de Informação de Akaike Esperado (EAIC) . . . . .	44
5.4.3	Critério de Informação de Desvio (DIC) . . . . .	44
5.4.4	Critério da Densidade Preditiva Ordenada (CPO) . . . . .	45
5.5	Análise de influência caso a caso . . . . .	46
<b>6</b>	<b>Resultados da abordagem bayesiana</b>	<b>48</b>
6.1	Aplicação . . . . .	48
6.2	Seleção do modelo . . . . .	49
6.3	Influência de observações extremas . . . . .	54

6.3.1	Dados simulados . . . . .	54
6.3.2	Dados de diâmetro-altura de <i>Eucalyptus sp.</i> . . . . .	60
<b>7</b>	<b>Conclusão</b>	<b>63</b>
	<b>Referências Bibliográficas</b>	<b>64</b>
<b>A</b>	<b>Características de algumas distribuições simétricas</b>	<b>71</b>
<b>B</b>	<b>Demais resultados da abordagem clássica</b>	<b>74</b>
B.1	Resultados dos testes de heterocedasticidade . . . . .	74
B.2	Estimativas de máxima verossimilhança e gráficos dos resíduos via GAMLSS	76
<b>C</b>	<b>Demais resultados da abordagem bayesiana</b>	<b>91</b>
C.1	Estimativas bayesianas . . . . .	91

# Lista de Figuras

2.1	Funções densidade da distribuição normal para $\mu = 0$ . . . . .	8
2.2	Funções densidade da distribuição normal generalizada para $\mu = 0$ e $\sigma^2 = 1$ . . . . .	9
2.3	Funções densidade da distribuição $t$ de Student para $\mu = 0$ e $\sigma = 1$ . . .	12
2.4	Funções densidade da distribuição logística-II para $\mu = 0$ . . . . .	13
4.1	Gráfico de dispersão das alturas contra os diâmetros de <i>Eucalyptus sp.</i> .	30
4.2	Valores fixados para o parâmetro $s$ da distribuição normal generalizada contra o critério SBC calculado para os modelos hipsométricos de (a) Stofel, (b) Curtis, (c) Petterson e (d) Prodan. . . . .	31
4.3	Gráfico das estimativas dos parâmetros da distribuição normal. (a) valores estimados para $\mu_i$ (b) valores estimados para $\sigma_i^2$ . . . . .	32
4.4	Gráfico das estimativas dos parâmetros da distribuição normal generalizada com $s = 3$ . (a) valores estimados para $\mu_i$ (b) valores estimados para $\sigma_i^2$ . . . . .	32
4.5	Gráfico das estimativas dos parâmetros da distribuição $t$ de Student com 30 g.l. (a) valores estimados para $\mu_i$ (b) valores estimados para $\sigma_i^2$ . . .	33

4.6	Gráfico das estimativas dos parâmetros da distribuição logística-II. (a) valores estimados para $\mu_i$ (b) valores estimados para $\sigma_i^2$ . . . . .	33
4.7	Percentis da distribuição normal generalizada com $s = 3$ quando ajustado o modelo de Prodan. . . . .	36
4.8	Quantis residuais do Modelo A. (a) quantis residuais versus os valores ajustados de $\mu_i$ (b) quantis residuais versus os índices $i$ (c) gráfico da densidade estimada (d) gráfico normal QQ plot. . . . .	37
4.9	Gráfico <i>worm-plot</i> para o Modelo A. . . . .	38
6.1	Valores fixados para o parâmetro $s$ da distribuição normal generalizada contra o critério EBIC calculado para os modelos hipsométricos de (a) Stofel, (b) Curtis, (c) Petterson e (d) Prodan. . . . .	50
6.2	Gráfico de índices de $CPO_i$ para as distribuições (a) normal (b) normal generalizada com $s = 3$ (c) $t$ de Student com $\nu = 30$ e (d) logística-II. . . . .	52
6.3	Histograma dos parâmetros para o Modelo B. . . . .	53
6.4	Simulação dos parâmetros para o Modelo B. . . . .	54
6.5	Dados simulados da NG. Gráfico de índices de $K(\pi, \pi_{(-i)})$ a partir do ajuste da NG. . . . .	59
6.6	Dados simulados da NG. Gráfico de índices de $K(\pi, \pi_{(-i)})$ a partir do ajuste da tS. . . . .	60
6.7	Dados de diâmetro-altura de <i>Eucalyptus sp.</i> . Gráfico de índices de $K(\pi, \pi_{(-i)})$ para o Modelo B. . . . .	61
B.1	Quantis residuais do modelo de Stofel ajustado com a distribuição normal.	76
B.2	Quantis residuais do modelo de Curtis ajustado com a distribuição normal.	77

B.3	Quantis residuais do modelo de Petterson ajustado com a distribuição normal. . . . .	78
B.4	Quantis residuais do modelo de Prodan ajustado com a distribuição normal. . . . .	79
B.5	Quantis residuais do modelo de Stofel ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	80
B.6	Quantis residuais do modelo de Curtis ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	81
B.7	Quantis residuais do modelo de Petterson ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	82
B.8	Quantis residuais do modelo de Stofel ajustado com a distribuição $t$ de Student (30 g.l.). . . . .	83
B.9	Quantis residuais do modelo de Curtis ajustado com a distribuição $t$ de Student (30 g.l.). . . . .	84
B.10	Quantis residuais do modelo de Petterson ajustado com a distribuição $t$ de Student (30 g.l.). . . . .	85
B.11	Quantis residuais do modelo de Prodan ajustado com a distribuição $t$ de Student (30 g.l.). . . . .	86
B.12	Quantis residuais do modelo de Stofel ajustado com a distribuição logística-II. . . . .	87
B.13	Quantis residuais do modelo de Curtis ajustado com a distribuição logística-II. . . . .	88
B.14	Quantis residuais do modelo de Petterson ajustado com a distribuição logística-II. . . . .	89

B.15	Quantis residuais do modelo de Prodan ajustado com a distribuição logística-II. . . . .	90
C.1	Histograma e Simulação dos parâmetros para o modelo de Stofel ajustado com a distribuição normal. . . . .	92
C.2	Histograma e Simulação dos parâmetros para o modelo de Curtis ajustado com a distribuição normal. . . . .	93
C.3	Histograma e Simulação dos parâmetros para o modelo de Petterson ajustado com a distribuição normal. . . . .	94
C.4	Histograma e Simulação dos parâmetros para o modelo de Prodan ajustado com a distribuição normal. . . . .	95
C.5	Histograma e Simulação dos parâmetros para o modelo de Stofel ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	96
C.6	Histograma e Simulação dos parâmetros para o modelo de Curtis ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	97
C.7	Histograma e Simulação dos parâmetros para o modelo de Petterson ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	98
C.8	Histograma e Simulação dos parâmetros para o modelo de Stofel ajustado com a distribuição $t$ de Student ( $\nu = 30$ ). . . . .	99
C.9	Histograma e Simulação dos parâmetros para o modelo de Curtis ajustado com a distribuição $t$ de Student ( $\nu = 30$ ). . . . .	100
C.10	Histograma e Simulação dos parâmetros para o modelo de Petterson ajustado com a distribuição $t$ de Student ( $\nu = 30$ ). . . . .	101
C.11	Histograma e Simulação dos parâmetros para o modelo de Prodan ajustado com a distribuição $t$ de Student ( $\nu = 30$ ). . . . .	102

C.12 Histograma e Simulação dos parâmetros para o modelo de Stofel ajustado com a distribuição logística-II. . . . .	103
C.13 Histograma e Simulação dos parâmetros para o modelo de Curtis ajustado com a distribuição logística-II. . . . .	104
C.14 Histograma e Simulação dos parâmetros para o modelo de Petterson ajustado com a distribuição logística-II. . . . .	105
C.15 Histograma e Simulação dos parâmetros para o modelo de Prodan ajustado com a distribuição logística-II. . . . .	106

# Lista de Tabelas

3.1	Expressões de $W_g(u)$ e $W'_g(u)$ para algumas distribuições simétricas. . .	17
3.2	Valores de $d_g$ , $f_g$ e $\xi$ para algumas distribuições simétricas. . . . .	18
4.1	Relação dos modelos de hipsometria. . . . .	29
4.2	Comparação dos erros para os modelos ajustados usando o DMAP(%). . . . .	34
4.3	Estatísticas para seleção dos modelos. . . . .	35
4.4	Estimativas de máxima verossimilhança para o Modelo A via GAMLSS. . . . .	35
4.5	Sumário dos quantis residuais do Modelo A via GAMLSS. . . . .	37
6.1	Comparação dos modelos ajustados utilizando diferentes critérios bayesianos. . . . .	51
6.2	Sumário <i>a posteriori</i> para o Modelo B. . . . .	53
6.3	Dados simulados da NG. Médias e desvios padrão <i>a posteriori</i> para os parâmetros do modelo ajustado com a distribuição NG com $s = 3$ . . . . .	56
6.4	Dados simulados da NG. Médias e desvios padrão <i>a posteriori</i> para os parâmetros do modelo ajustado com a distribuição tS. . . . .	56
6.5	Dados simulados. Comparação entre os ajustes da NG e da tS usando diferentes critérios bayesianos. . . . .	57



6.6	Medidas de diagnóstico para os dados simulados. . . . .	58
6.7	Identificação de casos influentes para os dados de diâmetro-altura de <i>Eucalyptus sp.</i> . . . . .	61
6.8	Estimativas para os dados de diâmetro-altura de <i>Eucalyptus sp.</i> após retirado o caso influente. . . . .	62
A.1	Algumas distribuições pertencentes à classe simétrica. . . . .	72
A.2	Continuação da Tabela A.1. . . . .	73
B.1	Valor das estatísticas dos testes e p-valor (entre parênteses) para os modelos ajustados com a distribuição normal. . . . .	74
B.2	Valor das estatísticas dos testes e p-valor (entre parênteses) para os modelos ajustados com a distribuição normal generalizada com $s = 3$ . . . . .	75
B.3	Valor das estatísticas dos testes e p-valor (entre parênteses) para os modelos ajustados com a distribuição $t$ de Student com 30 g.l. . . . .	75
B.4	Valor das estatísticas dos testes e p-valor (entre parênteses) para os modelos hipsométricos ajustados com a distribuição logística-II. . . . .	75
B.5	Estimativas de máxima verossimilhança para o modelo de Stofel ajustado com a distribuição normal. . . . .	76
B.6	Estimativas de máxima verossimilhança para o modelo de Curtis ajustado com a distribuição normal. . . . .	77
B.7	Estimativas de máxima verossimilhança para o modelo de Petterson ajustado com a distribuição normal. . . . .	78
B.8	Estimativas de máxima verossimilhança para o modelo de Prodan ajustado com a distribuição normal. . . . .	79

B.9	Estimativas de máxima verossimilhança para o modelo de Stofel ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	80
B.10	Estimativas de máxima verossimilhança para o modelo de Curtis ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	81
B.11	Estimativas de máxima verossimilhança para o modelo de Petterson ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	82
B.12	Estimativas de máxima verossimilhança para o modelo de Stofel ajustado com a distribuição $t$ de Student (30 g.l.). . . . .	83
B.13	Estimativas de máxima verossimilhança para o modelo de Curtis ajustado com a distribuição $t$ de Student (30 g.l.). . . . .	84
B.14	Estimativas de máxima verossimilhança para o modelo de Petterson ajustado com a distribuição $t$ de Student (30 g.l.). . . . .	85
B.15	Estimativas de máxima verossimilhança para o modelo de Prodan ajustado com a distribuição $t$ de Student (30 g.l.). . . . .	86
B.16	Estimativas de máxima verossimilhança para o modelo de Stofel ajustado com a distribuição logística-II. . . . .	87
B.17	Estimativas de máxima verossimilhança para o modelo de Curtis ajustado com a distribuição logística-II. . . . .	88
B.18	Estimativas de máxima verossimilhança para o modelo de Petterson ajustado com a distribuição logística-II. . . . .	89
B.19	Estimativas de máxima verossimilhança para o modelo de Prodan ajustado com a distribuição logística-II. . . . .	90
C.1	Estimativas <i>a posteriori</i> para o modelo de Stofel ajustado com a distribuição normal. . . . .	91

C.2	Estimativas <i>a posteriori</i> para o modelo de Curtis ajustado com a distribuição normal. . . . .	92
C.3	Estimativas <i>a posteriori</i> para o modelo de Petterson ajustado com a distribuição normal. . . . .	93
C.4	Estimativas <i>a posteriori</i> para o modelo de Prodan ajustado com a distribuição normal. . . . .	94
C.5	Estimativas <i>a posteriori</i> para o modelo de Stofel ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	95
C.6	Estimativas <i>a posteriori</i> para o modelo de Curtis ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	96
C.7	Estimativas <i>a posteriori</i> para o modelo de Petterson ajustado com a distribuição normal generalizada ( $s = 3$ ). . . . .	97
C.8	Estimativas <i>a posteriori</i> para o modelo de Stofel ajustado com a distribuição $t$ de Student ( $\nu = 30$ ). . . . .	98
C.9	Estimativas <i>a posteriori</i> para o modelo de Curtis ajustado com a distribuição $t$ de Student ( $\nu = 30$ ). . . . .	99
C.10	Estimativas <i>a posteriori</i> para o modelo de Petterson ajustado com a distribuição $t$ de Student ( $\nu = 30$ ). . . . .	100
C.11	Estimativas <i>a posteriori</i> para o modelo de Prodan ajustado com a distribuição $t$ de Student ( $\nu = 30$ ). . . . .	101
C.12	Estimativas <i>a posteriori</i> para o modelo de Stofel ajustado com a distribuição logística-II. . . . .	102
C.13	Estimativas <i>a posteriori</i> para o modelo de Curtis ajustado com a distribuição logística-II. . . . .	103

C.14 Estimativas <i>a posteriori</i> para o modelo de Petterson ajustado com a distribuição logística-II. . . . .	104
C.15 Estimativas <i>a posteriori</i> para o modelo de Prodan ajustado com a distribuição logística-II. . . . .	105

# Capítulo 1

## Introdução

### 1.1 Motivação e definição dos objetivos

Os modelos de regressão simétricos, cuja distribuição assumida para os erros tem caudas mais leves ou mais pesadas do que a normal, têm mostrado ajustes tão bons ou melhores do que os modelos normais. Além do mais, é bem conhecido que a modelagem com a suposição de respostas normalmente distribuídas pode ser altamente influenciada por observações extremas, também chamadas de observações aberrantes, o que incentiva o uso de metodologias mais robustas. Dentre essas metodologias podemos destacar os modelos que podem produzir estimativas robustas contra as observações aberrantes, podendo assim reduzir e controlar a influência dessas observações. Dessa forma, os modelos simétricos proporcionam uma grande flexibilidade na especificação da distribuição da variável resposta. Essa classe de modelos é discutida nesse trabalho.

Podemos então relacionar os seguintes objetivos específicos que pretendemos alcançar no decorrer desse trabalho:

- (i) revisão dos modelos de regressão simétricos, com foco nos modelos não-lineares heterocedásticos simétricos. Análise inferencial, testes de heterocedasticidade e

resíduos para esses modelos;

(ii) propor uma abordagem bayesiana para os modelos citados no item (i) considerando distribuições *a priori* informativas, uma modelagem baseada em métodos de simulação MCMC e uma análise de influência caso a caso com base na divergência de Kullback-Leibler;

(iii) aplicação dos métodos clássico e bayesiano a um conjunto de dados reais.

## 1.2 Apresentação dos capítulos

Neste primeiro capítulo apresentamos a contextualização e motivação desta dissertação, bem como seus principais objetivos. No Capítulo 2 apresentamos a classe dos modelos simétricos e algumas distribuições utilizadas nesse trabalho. No Capítulo 3 tratamos dos modelos não-lineares heterocedásticos simétricos, abordando a análise clássica e incluindo testes de heterocedasticidade, análise de resíduos e critérios de seleção. No Capítulo 4 apresentamos uma aplicação do método clássico em um conjunto de dados reais bem como os seus resultados. No Capítulo 5 propomos uma abordagem bayesiana baseada em métodos MCMC para fazer inferência sobre os modelos não-lineares heterocedásticos simétricos. Propomos também uma análise de influência caso a caso com base na divergência de Kullback-Leibler. No Capítulo 6 ilustramos a aplicação do método bayesiano utilizando um conjunto de dados reais bem como os seus resultados. Por fim, no Capítulo 7 apresentamos as conclusões com base nos resultados obtidos.

# Capítulo 2

## Modelos Simétricos

O estudo abordando distribuições simétricas tem crescido consideravelmente na literatura nos últimos anos, pelo fato de as mesmas serem uma alternativa robusta ao modelo normal. A área de pesquisa em distribuições simétricas começou, talvez, com as aplicações em engenharia consideradas por Blake & Thomas (1968) e McGraw & Wagner (1968). Em Chu (1973), Cambanis *et al.* (1981) e Chmielewski (1981) foi elaborado um levantamento detalhado dessas distribuições e suas aplicações em diversas áreas e, em Muirhead (1980, 1982), Berkane & Bentler (1986), Rao (1990) e (Fang *et al.*, 1990) também foram exploradas suas propriedades. Ainda, Cambanis *et al.* (1981) observou que a classe de distribuições simétricas coincide com a família de distribuições elípticas univariadas. Na linha de modelos que produzem estimativas robustas, podemos citar Lange *et al.* (1989), que propõe modelos supondo erros com distribuição  $t$  de Student com  $\nu$  graus de liberdade enquanto Little (1988) e Yamaguchi (1990) propõem modelos baseados em erros com distribuição  $t$  multivariada e normal contaminada. Outro estudo propõe um modelo de regressão linear em que os erros têm distribuição exponencial potência com um parâmetro extra de forma (Taylor, 1992). Em Arellano-Valle (1994) foram apresentados resultados sobre as propriedades, inferência e aplicações à modelos de regressão na classe de distribuições elípticas e

posteriormente estendidos para os modelos de regressão linear simétricos (Ferrari & Uribe-Opazo, 2001). Em um outro trabalho (Cordeiro *et al.*, 2000) foi apresentada a correção do viés de segunda ordem das estimativas de máxima verossimilhança na classe dos modelos de regressão não-lineares simétricos. Encontramos ainda correções de Bartlett para esses modelos de regressão não-lineares simétricos (Cordeiro, 2004), permitindo a generalização de resultados obtidos por Ferrari & Uribe-Opazo (2001). Outro trabalho desenvolveu uma teoria assintótica de segunda ordem para testes de escore na classe dos modelos de regressão lineares simétricos (Uribe-Opazo *et al.*, 2008). Em Cysneiros (2004) foram desenvolvidos métodos restritos e de validação de modelos de regressão simétricos e modelos de regressão lineares heterocedásticos simétricos, em que o trabalho propõe metodologias de influência local bem como medidas de alavancagem. Posteriormente, Cysneiros *et al.* (2005) apresentou os principais resultados inferenciais e técnicas de diagnóstico em modelos não-lineares, lineares e lineares heterocedásticos com erros simétricos e por fim, discutiu algumas extensões de modelos elípticos para dados correlacionados. Cysneiros *et al.* (2007) discutiu os modelos lineares heterocedásticos simétricos, bem como aspectos de robustez dos estimadores de máxima verossimilhança contra *outliers* e observações com variância grande, teste de escore e diagnósticos para heterocedasticidade. Ainda, podemos citar Cysneiros *et al.* (2010) que discutiu a classe dos modelos não-lineares heterocedásticos simétricos, apresentando uma fórmula geral para a matriz do viés de segunda ordem dos estimadores de máxima verossimilhança que generaliza alguns resultados anteriores. Considerando ainda a classe dos modelos não-lineares heterocedásticos podemos citar Lachos *et al.* (2011), que abordou os modelos baseados em distribuições de mistura de escala normal estendendo o trabalho de Cysneiros *et al.* (2010). Um outro trabalho (Russo *et al.*, 2009) discutiu o uso de modelos elípticos com efeitos mistos como uma forma de ajustar dados não-lineares correlacionados, também apresentou testes para componentes de variância e uma análise de diagnóstico e sensibilidade. Por fim, podemos ainda encon-



trar na literatura alternativas à classe simétrica, como a classe dos modelos simétricos transformados que estende os modelos de Box-Cox (Cordeiro & Andrade, 2009).

## 2.1 Classe dos modelos simétricos

Nesta seção apresentamos a classe simétrica univariada e alguns resultados teóricos importantes para o desenvolvimento desse trabalho.

**Definição 1** Uma variável aleatória  $Y$  tem distribuição simétrica com parâmetro de locação  $\mu \in \Re$  e parâmetro de escala  $\sigma^2$ ,  $\sigma > 0$ , se sua função densidade de probabilidade é dada por

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma} g \left\{ \left( \frac{y - \mu}{\sigma} \right)^2 \right\}, \quad y \in \Re \quad (2.1)$$

para alguma função  $g(\cdot)$  (denominada função geradora de densidades), com  $g(u) > 0$  para  $u > 0$  e  $\int_0^\infty u^{-1/2} g(u) du = 1$ , em que  $u = [(y - \mu)/\sigma]^2$ . Essa condição é necessária para que  $f(y; \mu, \sigma^2)$  seja uma função densidade de probabilidade. Dessa forma, denotamos por  $Y \sim S(\mu, \sigma^2)$  a variável aleatória simétrica definida acima.

Podemos citar como distribuições pertencentes a essa classe as distribuições normal, normal generalizada, Cauchy,  $t$  de Student,  $t$  de Student generalizada, logística dos tipos I e II, logística generalizada, Kotz, Kotz generalizada, normal contaminada, exponencial dupla, exponencial potência e família potência estendida.

Algumas propriedades da distribuição normal podem ser estendidas para a classe simétrica.

1. Fang *et al.* (1990) observa que se  $Y \sim S(\mu, \sigma^2)$  então a função característica de  $Y$ ,  $\zeta_Y(t) = E(e^{itY})$  é dada por  $e^{it\mu} \varphi(t^2 \sigma^2)$ ,  $t \in \Re$  para alguma função  $\varphi$ , com  $\varphi(u) \in \Re$  para  $u > 0$ . Quando existem, temos que  $E(Y_i) = \mu_i$  e  $Var(Y_i) = \xi \sigma^2$ , em que  $\xi > 0$  é uma constante dada por  $\xi = -2\varphi'(0)$ , com  $\varphi'(0) = d\varphi(u)/du|_{u=0}$

e que não depende dos parâmetros  $\mu$  e  $\sigma^2$ . Ainda, se  $u^{-\frac{1}{2}(k+1)}g(u)$  for integrável então o  $k$ -ésimo momento de  $Y$  existe (Kelker, 1970).

2. Podemos ver que, se  $Y \sim S(\mu, \sigma^2)$  então  $a+bY \sim S(a+b\mu, b^2\sigma^2)$ , em que  $a, b \in \Re$  com  $b \neq 0$ , isto é, a distribuição de qualquer combinação linear de uma variável aleatória com distribuição simétrica também é simétrica. Para ilustrar, temos que se  $Y \sim S(\mu, \sigma^2)$  então  $Z = (Y - \mu)/\sigma \sim S(0, 1)$ , com função densidade  $f(z) = f(z; 0, 1) = g(z^2)$  para  $z \in \Re$ , de modo que  $Z$  é chamada de variável simétrica padrão. Para mais detalhes ver (Fang *et al.*, 1990; Cysneiros *et al.*, 2005).

A seguir apresentamos algumas distribuições pertencentes à classe simétrica utilizadas nesse trabalho, em que  $u = [(y - \mu)/\sigma]^2$  sendo  $Y \sim S(\mu, \sigma^2)$  na reta real. Um resumo das demais distribuições simétricas citadas nesse texto pode ser encontrado no Apêndice A (Tabelas A.1 e A.2) em que apresentamos algumas características especiais de cada distribuição.

## 2.2 Distribuição Normal

A distribuição normal (ou Gaussiana), pertencente à classe simétrica, é considerada a distribuição mais proeminente na estatística devido ao fato de ser uma distribuição bastante tratável analiticamente e uma escolha conveniente para a modelagem de uma grande variedade de variáveis aleatórias encontradas na prática.

Uma variável aleatória  $Y \sim S(\mu, \sigma)$  tem distribuição normal se sua função geradora de densidade  $g(\cdot)$  é da forma

$$g(u) = \frac{1}{\sqrt{2\pi}} \exp\{-u/2\}, \quad u > 0. \quad (2.2)$$

Logo, a função densidade de probabilidade (f.d.p.) de  $Y$ , denotada por  $Y \sim N(\mu, \sigma^2)$

e obtida aplicando (2.2) em (2.1), é definida como

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(y - \mu)^2}{2\sigma^2}\right\}, \quad y \in \mathfrak{R}. \quad (2.3)$$

A função característica da distribuição normal é dada por

$$\zeta_Y(t) = e^{it\mu} \exp\left\{-\frac{1}{2}t^2\sigma^2\right\}, \quad t \in \mathfrak{R}.$$

Sendo  $Y \sim N(\mu, \sigma^2)$  temos que  $E(Y) = \mu$ ,  $Var(Y) = \sigma^2$  e os momentos centrais de ordem  $r$  são

$$\mu_r = E\{(Y - \mu)^r\} = \begin{cases} 0, & r \text{ ímpar;} \\ \sigma^r r! / \{2^{r/2}(r/2)!\}, & r \text{ par.} \end{cases}$$

Em particular, o coeficiente de curtose é  $\gamma_2 = 3$ .

A Figura 2.1 apresenta os gráficos da função densidade da distribuição normal com  $\mu = 0$  e diferentes valores de  $\sigma^2$ .

## 2.3 Distribuição Normal Generalizada

A distribuição normal generalizada (Nadarajah, 2005) é uma generalização natural da distribuição normal, em que substituímos a potência de ordem 2 com a potência  $s > 0$  de qualquer ordem.

Dizemos que uma variável aleatória  $Y \sim S(\mu, \sigma^2)$  tem distribuição normal generalizada se sua função geradora de densidade é dada por

$$g(u) = \frac{s}{2\Gamma(1/s)} \exp\{-\sqrt{u^s}\}, \quad u > 0, \quad s > 0.$$

Contudo, a função densidade de probabilidade de  $Y$ , denotada por  $NG(\mu, \sigma^2, s)$ , é

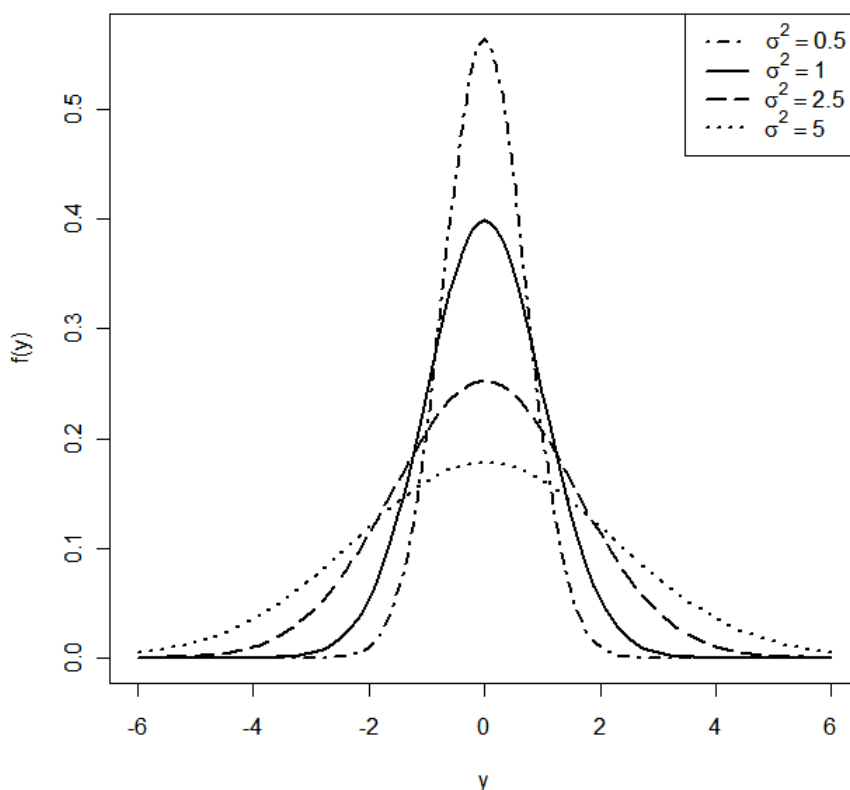


Figura 2.1: Funções densidade da distribuição normal para  $\mu = 0$ .

dada por

$$f(y; \mu, \sigma^2, s) = \frac{s}{2\sigma\Gamma(1/s)} \exp\left\{-\left|\frac{y-\mu}{\sigma}\right|^s\right\}, \quad y \in \mathfrak{R}, \quad (2.4)$$

em que  $\Gamma(\cdot)$  denota a função Gama.

A distribuição normal generalizada tem como caso particular a distribuição normal quando  $s = 2$  (com média  $\mu$  e variância  $\sigma^2/2$ ), e quando  $s = 1$  tem como caso particular a distribuição de Laplace. Quando  $s \rightarrow \infty$ , a f.d.p. (2.4) converge pontualmente para uma distribuição uniforme em  $(\mu - \sigma, \mu + \sigma)$ . Assim como a distribuição normal, a f.d.p. dada por (2.4) é unimodal com moda em  $y = \mu$ .

Se  $Y \sim \text{NG}(\mu, \sigma^2, s)$  temos que  $E(Y) = \mu$  e os momentos centrais de ordem  $r$  são

$$E\{(Y - \mu)^r\} = \begin{cases} 0, & r \text{ ímpar;} \\ \sigma^r \Gamma[(r+1)/s] / \Gamma(1/s), & r \text{ par.} \end{cases} \quad (2.5)$$

Em particular, a variância e o coeficiente de curtose de  $Y$  são dados, respectivamente, por

$$Var(Y) = \frac{\sigma^2 \Gamma(3/s)}{\Gamma(1/s)} \quad (2.6)$$

$$\gamma_2 = \frac{\Gamma(1/s)\Gamma(5/s)}{\Gamma^2(3/s)} \quad (2.7)$$

O desvio médio em torno da média e o desvio médio em torno da mediana são dados por

$$E(|Y - \mu|) = \frac{\sigma \Gamma(2/s)}{\Gamma(1/s)} \quad (2.8)$$

visto que para a distribuição normal generalizada  $E(Y)$  e  $Mediana(Y)$  são iguais a  $\mu$ .

A Figura 2.2 apresenta os gráficos da função densidade da distribuição normal generalizada com  $\mu = 0$ ,  $\sigma^2 = 1$  e diferentes valores de  $s$ .

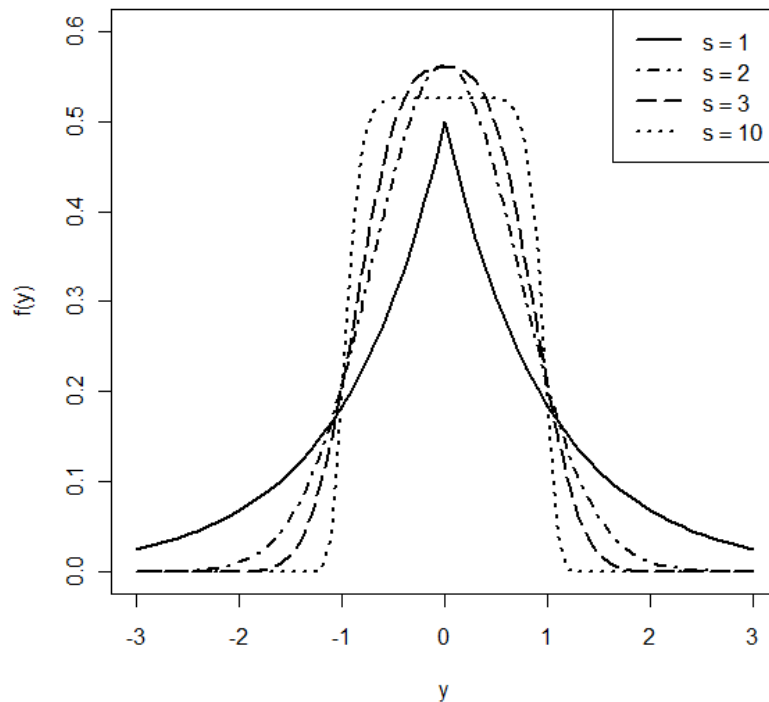


Figura 2.2: Funções densidade da distribuição normal generalizada para  $\mu = 0$  e  $\sigma^2 = 1$ .

## 2.4 Distribuição $t$ de Student

A distribuição  $t$  de Student é uma das distribuições mais utilizadas para a modelagem de dados leptocúrticos, ou seja, dados que provêm de uma distribuição com caudas mais pesadas do que a distribuição normal, permitindo reduzir a influência de observações extremas. Essa distribuição nos permite ajustar a curtose dos dados através do número de graus de liberdade  $\nu$  (Stasinopoulos *et al.*, 2008).

Uma variável aleatória  $Y \sim S(\mu, \sigma^2)$  tem distribuição  $t$  de Student com  $\nu$  graus de liberdade (g.l.) se sua função geradora de densidade  $g(\cdot)$  é dada por

$$g(u) = \frac{1}{B\left(\frac{1}{2}, \frac{\nu}{2}\right) \nu^{\frac{1}{2}}} \left[1 + \frac{u}{\nu}\right]^{-\frac{(\nu+1)}{2}}, \quad \nu > 0, \quad u > 0, \quad (2.9)$$

em que  $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$  é a função Beta, sendo  $\Gamma(\cdot)$  a função Gama.

A função densidade de probabilidade da distribuição  $t$  de Student, denotada nesse trabalho por  $tS(\mu, \sigma^2, \nu)$ , é definida como

$$f(y; \mu, \sigma^2, \nu) = \frac{1}{\sigma B\left(\frac{1}{2}, \frac{\nu}{2}\right) \nu^{\frac{1}{2}}} \left[1 + \frac{(y - \mu)^2}{\sigma^2 \nu}\right]^{-\frac{(\nu+1)}{2}} \quad y \in \mathfrak{R}. \quad (2.10)$$

Sabemos que, a distribuição  $t$  de Student com parâmetros  $(\mu, \sigma^2, \nu)$  tende a uma distribuição normal com média  $\mu$  e variância  $\sigma^2$  quando  $\nu \rightarrow \infty$ . De forma similar, quando  $\nu = 1$  temos a distribuição de Cauchy com parâmetros  $(\mu, \sigma^2)$ .

A função característica da distribuição  $t$  de Student pode ser encontrada em Fang *et al.* (1990). Se  $Y \sim tS(\mu, \sigma^2, \nu)$  então  $E(Y) = \mu$  para  $\nu > 1$  e os momentos de ordem

$r$  para  $\nu > r$  são dados por

$$E\{(Y - \mu)^r\} = \begin{cases} 0, & r \text{ ímpar;} \\ \sigma^r \frac{\nu^{r/2} \Gamma(\frac{r+1}{2}) \Gamma(\frac{\nu-r}{2})}{\Gamma(\frac{1}{2}) \Gamma(\frac{\nu}{2})}, & r \text{ par.} \end{cases} \quad (2.11)$$

Em particular, se  $r \geq \nu$  e  $r$  é par, então o momento de ordem  $r$  é infinito. Ainda, a variância e coeficiente de curtose de  $Y$  são dados, respectivamente, por

$$Var(Y) = \frac{\sigma^2 \nu}{(\nu - 2)}, \quad \text{para } \nu > 2. \quad (2.12)$$

$$\gamma_2 = 3 + \frac{6}{(\nu - 4)}, \quad \text{para } \nu > 4. \quad (2.13)$$

O desvio médio, dado pela média dos desvios absolutos da amostra com relação à média, é definido como

$$E\{|Y - \mu|\} = \frac{\sigma \nu^{1/2} \Gamma(\frac{\nu-1}{2})}{\Gamma(1/2) \Gamma(\nu/2)}. \quad (2.14)$$

A Figura 2.3 apresenta os gráficos da função densidade da distribuição  $t$  de Student com  $\mu = 0$ ,  $\sigma^2 = 1$  e diferentes valores de  $\nu$ .

## 2.5 Distribuição Logística-II

A distribuição logística-II tem sido bastante usada para representar curvas de crescimento em demografia e economia. Em Johnson *et al.* (1995) podemos encontrar uma série de aplicações da distribuição logística-II em diversas áreas.

Dizemos que a variável aleatória  $Y \sim S(\mu, \sigma^2)$  tem distribuição logística-II se sua

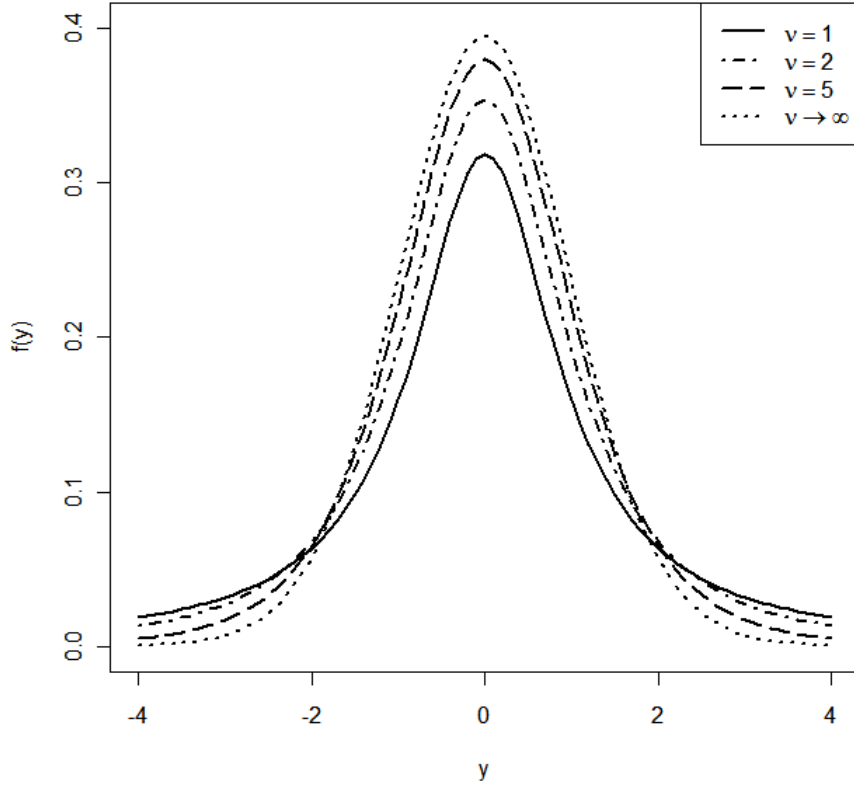


Figura 2.3: Funções densidade da distribuição  $t$  de Student para  $\mu = 0$  e  $\sigma = 1$ .

função geradora de densidade  $g(\cdot)$  é da forma

$$g(u) = \frac{e^{-u^{1/2}}}{(1 + e^{-u^{1/2}})^2}, \quad u > 0. \quad (2.15)$$

A função densidade de probabilidade da distribuição logística-II, denotada nesse trabalho por  $\text{LII}(\mu, \sigma^2)$ , é dada por

$$f(y; \mu, \sigma^2) = \frac{1}{\sigma} \left\{ \exp \left[ - \left( \frac{y - \mu}{\sigma} \right) \right] \right\} \left\{ 1 + \exp \left[ - \left( \frac{y - \mu}{\sigma} \right) \right] \right\}^{-2}, \quad y \in \mathfrak{R}. \quad (2.16)$$

A função característica da distribuição logística-II é definida como

$$\zeta_Y(t) = \frac{2(e^{it\mu}\pi\sigma t)}{(e^{\pi\sigma t} - e^{-\pi\sigma t})}, \quad t \in \mathfrak{R}. \quad (2.17)$$



Para  $Y \sim \text{LII}(\mu, \sigma^2)$  temos que  $E(Y) = \mu$ ,  $\text{Var}(Y) = \pi^2\sigma^2/3$  e coeficiente de curtose  $\gamma_2 = 4,2$ . Ainda, a mediana e a moda para essa distribuição são iguais a média. Podemos encontrar a expressão dos momentos de ordem  $r$  de  $Y \sim \text{LII}(\mu, \sigma^2)$  em Johnson *et al.* (1995). A Figura 2.4 apresenta os gráficos da função densidade da distribuição logística-II com  $\mu = 0$  e diferentes valores de  $\sigma^2$ .

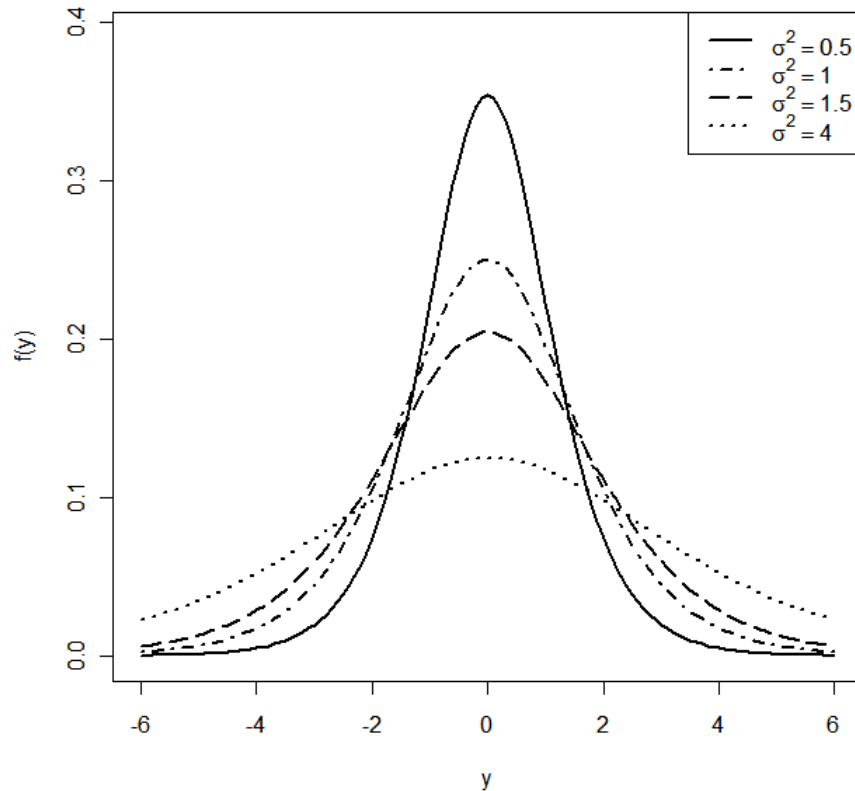


Figura 2.4: Funções densidade da distribuição logística-II para  $\mu = 0$ .

# Capítulo 3

## Modelo de regressão simétrico:

## Abordagem clássica

### 3.1 Modelos não-lineares simétricos

Para definir a classe dos modelos de regressão simétricos, supomos primeiramente que  $Y_1, \dots, Y_n$  são variáveis aleatórias independentes, cada uma com distribuição pertencente à classe simétrica definida em (2.1). Com isso, o modelo não-linear simétrico é definido por

$$y_i = \mu_i(\boldsymbol{\beta}, \mathbf{x}_i) + \epsilon_i, \quad i = 1, \dots, n \quad (3.1)$$

em que  $\mu_i = \mu_i(\boldsymbol{\beta}, \mathbf{x}_i)$  - chamada componente sistemática - é uma função não linear contínua e diferenciável em relação ao vetor de parâmetros desconhecidos  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  de dimensão  $p \times 1$ . Além disso, assumimos que a matriz de derivadas  $\mathbf{D}_{\boldsymbol{\beta}} = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta}$  tem posto  $p$  ( $p < n$ ) para todo  $\boldsymbol{\beta}$  com  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)^T$ . Os erros aleatórios,  $\epsilon_i$ , são independentes e seguem distribuição  $S(0, \sigma^2)$ ,  $\mathbf{y} = (y_1, \dots, y_n)^T$  é o vetor de respostas observadas e  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$  é o vetor que representa os valores de  $p$  variáveis explicativas, cuja dimensão é  $p \times 1$ . No caso linear temos  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$  com  $\mathbf{X} = (\mathbf{x}_1^T, \dots, \mathbf{x}_n^T)^T$

matriz  $n \times p$  de posto completo, conseqüentemente  $D_{\boldsymbol{\beta}} = \mathbf{X}$  (Cysneiros *et al.*, 2005). Portanto, os modelos de regressão não-lineares simétricos são definidos por (2.1) e (3.1). Ainda, quando a existência dos momentos for satisfeita para as distribuições da classe simétrica temos que  $E(Y_i) = \mu_i$  e  $Var(Y_i) = \xi\sigma^2$ . Alguns valores de  $\xi$  são apresentados na Tabela 3.2.

A função log-verossimilhança de  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2)^T$  é dada por

$$l(\boldsymbol{\theta}) = -\frac{n}{2} \log \sigma^2 + \sum_{i=1}^n \log\{g(u_i)\} \quad (3.2)$$

em que  $u_i = [(y_i - \mu_i)/\sigma]^2$  e  $y_i \sim S(\mu, \sigma^2)$ .

Para obter a função escore e as matrizes de informação de Fisher precisamos derivar a função  $l(\boldsymbol{\theta})$  em relação aos parâmetros desconhecidos. Para isso, supomos aqui que tais derivadas existem e que as distribuições simétricas utilizadas nesse trabalho satisfazem certas condições de regularidade necessárias para a validade dos resultados. As condições de regularidade são basicamente: (i) a existência dos momentos; (ii) a diferenciação da função log-verossimilhança em todo espaço paramétrico.

A função escore para  $\boldsymbol{\beta}$  é obtida por meio de

$$\frac{\partial}{\partial \beta_j} l(\boldsymbol{\theta}) = \frac{d}{d\beta_j} \left\{ \sum_{i=1}^n \log\{g(u_i)\} \right\}. \quad (3.3)$$

Em geral, no contexto elíptico é usual definir  $W_g(u)$  e  $W'_g(u)$  de tal forma que  $W_g(u) = [d \log\{g(u)\}]/du$  e  $W'_g(u) = [dW_g(u)]/du$ . Então,

$$\frac{\partial}{\partial \beta_j} l(\boldsymbol{\theta}) = \sum_{i=1}^n W_g(u_i) \frac{\partial u_i}{\partial \beta_j} = \sum_{i=1}^n \left\{ \frac{v_i}{\sigma^2} (y_i - \mu_i) \frac{\partial \mu_i}{\partial \beta_j} \right\}, \quad (3.4)$$

para  $j = 1, \dots, p$ . As expressões  $W_g(u)$  e  $W'_g(u)$  para as distribuições simétricas utilizadas nesse trabalho são apresentadas na Tabela 3.1.

Portanto, a função escore para  $\boldsymbol{\beta}$  na forma matricial é dada por

$$\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \mathbf{D}_{\boldsymbol{\beta}}^T \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu}), \quad (3.5)$$

em que  $\mathbf{D}(\mathbf{v}) = \text{diag}\{v_1, \dots, v_n\}$  com  $v_i = -2W_g(u_i)$ .

Analogamente, a função escore para  $\sigma^2$  é obtida por meio de

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} l(\boldsymbol{\theta}) &= \frac{d}{d\sigma^2} \left\{ -\frac{n}{2} \log(\sigma^2) + \sum_{i=1}^n \log\{g(u_i)\} \right\} \\ &= -\frac{n}{2\sigma^2} + \sum_{i=1}^n \left\{ -W_g(u_i)(\sigma^2)^{-2}(y_i - \mu_i)^2 \right\} \\ &= -\frac{n}{2\sigma^2} + \sum_{i=1}^n \left\{ \frac{v_i}{2(\sigma^2)^2}(y_i - \mu_i)^2 \right\}. \end{aligned} \quad (3.6)$$

Logo, a função escore para  $\sigma^2$  na forma matricial é dada pela expressão

$$\mathbf{U}_{\sigma^2}(\boldsymbol{\theta}) = \frac{1}{2\sigma^2} \left\{ \frac{1}{\sigma^2} Q_{\mathbf{v}}(\boldsymbol{\beta}, \sigma^2) - n \right\}, \quad (3.7)$$

em que  $Q_{\mathbf{v}}(\boldsymbol{\beta}, \sigma^2) = (\mathbf{y} - \boldsymbol{\mu})^T \mathbf{D}(\mathbf{v})(\mathbf{y} - \boldsymbol{\mu})$ .

Para a estimação dos parâmetros nos modelos de regressão simétricos podemos utilizar o método de máxima verossimilhança como apresentado em Cysneiros *et al.* (2005), o qual consiste em encontrar as estimativas do vetor de parâmetros  $\boldsymbol{\beta}$  e do escalar  $\sigma^2$  através da solução das equações dadas por  $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta}) = 0$  e  $\mathbf{U}_{\sigma^2}(\boldsymbol{\theta}) = 0$ . A solução dessas equações não possui forma fechada, tornando-se necessário o uso de métodos iterativos para obter as estimativas de máxima verossimilhança. Sendo assim, essas estimativas podem ser obtidas por meio de um algoritmo utilizando o método de escore de Fisher, discutido em Lange (2004), cujo processo iterativo conjunto é dado

por

$$\boldsymbol{\beta}^{(m+1)} = \boldsymbol{\beta}^{(m)} + (4d_g)^{-1} \{ \mathbf{D}_{\boldsymbol{\beta}}^{T(m)} \mathbf{D}_{\boldsymbol{\beta}}^{(m)} \}^{-1} \mathbf{D}_{\boldsymbol{\beta}}^{T(m)} \mathbf{D}(\mathbf{v}^{(m)}) [\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(m)})] \quad (3.8)$$

$$\sigma^{2(m+1)} = \frac{1}{n} Q_{\mathbf{v}}(\boldsymbol{\beta}^{(m+1)}, \sigma^{2(m)}), \quad m = 0, 1, 2, \dots, \quad (3.9)$$

em que  $d_g = E\{W_g^2(U^2)U^2\}$  com  $U \sim S(0, 1)$ . Os valores de  $d_g$  para as distribuições simétricas utilizadas nesse trabalho são apresentados na Tabela 3.2.

Tabela 3.1: Expressões de  $W_g(u)$  e  $W'_g(u)$  para algumas distribuições simétricas.

Distribuições	$W_g(u)$	$W'_g(u)$
Normal	$-\frac{1}{2}$	0
Normal Generalizada	$-\frac{s}{2} u^{(s/2)-1}$	$-\frac{(s^2-2s)}{4} u^{(s/2)-2}$
$t$ de Student	$-\frac{\nu+1}{2(\nu+u)}$	$\frac{(\nu+1)}{2(\nu+u)^2}$
Logística-II	$-\frac{\exp(-\sqrt{u})-1}{(-2\sqrt{u})[1+\exp(-\sqrt{u})]}$	$\frac{2\exp(-\sqrt{u})\sqrt{u}+\exp(-2\sqrt{u})-1}{-4u^{3/2}[1+\exp(-\sqrt{u})]^2}$

Dado que a função  $l(\boldsymbol{\theta})$  é assumida regular com respeito ao vetor de parâmetros  $\boldsymbol{\theta}$  segundo as condições apresentadas em Cox & Hinkley (1974), podemos assumir que o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$ ,  $\hat{\boldsymbol{\theta}}$ , é um estimador consistente e segue distribuição assintoticamente normal, cuja matriz de variâncias e covariâncias depende da matriz de informação esperada  $\mathbf{K}_{\boldsymbol{\theta}}$ , que pode ser expressa na forma

$$\mathbf{K}_{\boldsymbol{\theta}} = \begin{bmatrix} \mathbf{K}_{\boldsymbol{\beta}} & 0 \\ 0 & K_{\sigma^2} \end{bmatrix} \quad (3.10)$$

em que

$$\mathbf{K}_\beta = E \left\{ -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \beta \partial \beta^T} \right\} = \frac{4d_g}{\sigma^2} \mathbf{D}_\beta^T \mathbf{D}_\beta \quad (3.11)$$

$$K_{\sigma^2} = E \left\{ -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial (\sigma^2)^2} \right\} = \frac{n}{4(\sigma^2)^2} (4f_g - 1) \quad (3.12)$$

sendo  $f_g = E\{W_g^2(U^2)U^4\}$  com  $U \sim S(0, 1)$  (ver Tabela 3.2).

Tabela 3.2: Valores de  $d_g$ ,  $f_g$  e  $\xi$  para algumas distribuições simétricas.

Distribuições	$d_g$	$f_g$	$\xi$
Normal	$\frac{1}{4}$	$\frac{3}{4}$	1
Normal Generalizada	$\frac{(1 + (-1)^{2s})s^2(1 + s)}{8}$	$\frac{(1 + (-1)^{2s})s^{3+(2/s)}\Gamma(2 + \frac{3}{s})}{8 \Gamma(1 + \frac{1}{s})}$	$\frac{\Gamma(3/s)}{\Gamma(1/s)}$
$t$ de Student	$\frac{(\nu + 1)}{4(\nu + 3)}$	$\frac{3(\nu + 1)}{4(\nu + 3)}$	$\frac{\nu}{\nu - 2}, \nu > 2$
Logística-II	$\frac{1}{12}$	0,60749	$\frac{\pi^2}{3}$

Podemos observar por meio de (3.10) que  $\boldsymbol{\beta}$  e  $\sigma^2$  são ortogonais, ou seja,  $E\{\partial^2 l(\boldsymbol{\theta})/\partial \beta \partial \sigma^2\} = 0$ . Logo, os estimadores  $\hat{\boldsymbol{\beta}}$  e  $\hat{\sigma}^2$  são assintoticamente independentes. Contudo, temos que  $\hat{\mathbf{K}}_{\hat{\boldsymbol{\beta}}}^{-1} = \frac{\hat{\sigma}^2}{4d_g} (\mathbf{D}_{\hat{\boldsymbol{\beta}}}^T \mathbf{D}_{\hat{\boldsymbol{\beta}}})^{-1}$  é um estimador consistente da matriz de covariâncias assintótica de  $\hat{\boldsymbol{\beta}}$ . De forma similar,  $\hat{K}_{\hat{\sigma}^2}^{-1} = \frac{4(\hat{\sigma}^2)^2}{n(4f_g - 1)}$  é um estimador consistente da variância assintótica de  $\hat{\sigma}^2$ . Os estimadores  $\hat{\mathbf{K}}_{\hat{\boldsymbol{\beta}}}^{-1}$  e  $\hat{K}_{\hat{\sigma}^2}^{-1}$  podem ser utilizados para encontrar intervalos de confiança assintóticos para  $\hat{\boldsymbol{\beta}}$  e  $\hat{\sigma}^2$ .

### 3.1.1 Comentários a respeito dos parâmetros $s$ e $\nu$

Podemos encontrar na literatura basicamente três métodos para a estimação do parâmetro de forma  $s$  da distribuição normal generalizada. São eles, o método de máxima verossimilhança (Nadarajah, 2005), o método dos momentos (Dominguez-Molina *et al.*, 2003) e ainda, uma combinação de ambos (Varanasi & Aazhang, 1989). Da

mesma forma, os graus de liberdade  $\nu$  da distribuição  $t$  de Student podem ser estimados por meio do método de máxima verossimilhança e método dos momentos (Arellano-Valle, 1994). No entanto, para obtermos estimativas mais estáveis para os parâmetros dos modelos, optamos por trabalhar com os parâmetros  $s$  e  $\nu$  fixados em um conjunto de valores inteiros, de tal forma que  $s$  e  $\nu$  são determinados por meio de um procedimento de seleção que consiste em minimizar o critério de informação bayesiano.

## 3.2 Modelos não-lineares heterocedásticos simétricos

No contexto dos modelos não-lineares simétricos abordamos a situação em que o parâmetro de dispersão não é constante para todas as observações, havendo assim uma estrutura heterocedástica. De maneira análoga à estrutura estabelecida para a variância nos modelos lineares, admitimos uma forma funcional que relaciona os parâmetros da dispersão, por meio de uma função de ligação, com um conjunto de variáveis explicativas.

Consideremos o modelo de regressão não-linear heterocedástico simétrico dado por

$$y_i = \mu_i + \sigma_i \epsilon_i, \quad i = 1, \dots, n. \quad (3.13)$$

A função densidade de probabilidade de  $Y_i$  é dada por

$$f(y_i) = \frac{1}{\sigma_i} g\{u_i\}, \quad y_i \in \mathfrak{R}, \quad (3.14)$$

em que  $u_i = [(y_i - \mu_i)/\sigma]^2$ , sendo  $y_1, \dots, y_n$  as variáveis repostas observadas,  $\mu_i = \mu_i(\boldsymbol{\beta}, \mathbf{x}_i)$  uma função não-linear contínua e diferenciável em relação ao vetor de parâmetros desconhecidos  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ ,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$  os valores de  $p$  covariáveis,  $\mathbf{D}_{\boldsymbol{\beta}} = \partial \boldsymbol{\mu} / \partial \boldsymbol{\beta}$  a matriz de derivadas de posto completo ( $p < n$ ) para todo  $\boldsymbol{\beta}$  e  $\epsilon_i$  os erros aleatórios com distribuição  $S(0, 1)$ . Assumimos que o parâmetro de es-

cala  $\sigma_i^2$  é parametrizado tal que  $\sigma_i^2 = h_i = h(\tau_i)$ , em que  $h(\cdot)$  é uma função positiva, um-a-um contínua e diferenciável e  $\tau_i = \mathbf{z}_i^T \boldsymbol{\delta}$ , sendo  $\mathbf{z}_i = (z_{i1}, \dots, z_{iq})^T$  um vetor de dimensão  $q \times 1$  de variáveis explicativas e  $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^T$  um vetor de dimensão  $q \times 1$  de parâmetros desconhecidos. A função  $h(\cdot)$  é conhecida como função de ligação de dispersão. As covariáveis de dispersão  $\mathbf{z}_i$  não são necessariamente as mesmas covariáveis de locação  $\mathbf{x}_i$ . Para as distribuições que satisfazem a existência dos momentos temos que  $E(Y_i) = \mu_i$  e  $Var(Y_i) = \xi \sigma_i^2$ . Contudo, os modelos não-lineares heterocedásticos simétricos são definidos por (3.14) e pelas componentes sistemáticas

$$\mu_i = \mu_i(\boldsymbol{\beta}, \mathbf{x}_i) \quad \text{e} \quad \sigma_i^2 = h(\tau_i), \quad i = 1, \dots, n. \quad (3.15)$$

A função log-verossimilhança de  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T)^T$  é dada por

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \sum_{i=1}^n \log\{\sigma_i^2\} + \sum_{i=1}^n \log\{g(u_i)\}, \quad (3.16)$$

sendo  $u_i = [(y_i - \mu_i)/\sigma_i]^2$ . Análogo ao desenvolvimento apresentado na Seção 3.1, podemos obter as funções score para  $\boldsymbol{\beta}$  e  $\boldsymbol{\delta}$  as quais assumem, respectivamente, as formas

$$\mathbf{U}_\beta(\boldsymbol{\theta}) = \mathbf{D}_\beta^T \mathbf{D}(\mathbf{g})(\mathbf{y} - \boldsymbol{\mu}) \quad \text{e} \quad \mathbf{U}_\delta(\boldsymbol{\theta}) = \mathbf{Z}^T \mathbf{m}, \quad (3.17)$$

em que  $\mathbf{y} = (y_1, \dots, y_n)^T$ ,  $\mathbf{D}(\mathbf{g}) = \text{diag}\{g_1, \dots, g_n\}$  com  $g_i = \frac{v_i}{\sigma_i^2}$ ,  $v_i = -2W_g(u_i)$ ,  $W_g(u) = [d \log\{g(u)\}]/du$ ,  $\mathbf{m} = (m_1, \dots, m_n)^T$  e  $m_i = \frac{h'_i}{2\sigma_i^2}(v_i u_i - 1)$  sendo  $h'_i = \frac{\partial h(\tau_i)}{\partial \tau_i}$  e  $\mathbf{Z}$  é uma matriz  $n \times q$  com linhas  $\mathbf{z}_i^T$  (Cysneiros *et al.*, 2005).

A matriz de informação (esperada) de Fisher para  $\boldsymbol{\theta}$  é uma matriz bloco diagonal que pode ser expressa na forma  $\mathbf{K}_\theta = \text{diag}\{\mathbf{K}_\beta, \mathbf{K}_\delta\}$ . Logo,  $\boldsymbol{\beta}$  e  $\boldsymbol{\delta}$  são globalmente ortogonais. As matrizes de informação esperada  $\mathbf{K}_\beta$  e  $\mathbf{K}_\delta$  são dadas, respectivamente, por

$$\mathbf{K}_\beta = \mathbf{D}_\beta^T \mathbf{W}_1 \mathbf{D}_\beta \quad \text{e} \quad \mathbf{K}_\delta = \mathbf{Z}^T \mathbf{W}_2 \mathbf{Z}, \quad (3.18)$$



em que  $\mathbf{W}_1 = \text{diag}\{4d_g/\sigma_i^2\}$  e  $\mathbf{W}_2 = \text{diag}\{\frac{(4f_g-1)h_i^2}{4(\sigma_i^2)^2}\}$ , para  $i = 1, \dots, n$ . Os valores  $d_g$  e  $f_g$  são apresentados na Tabela 3.2.

Um processo iterativo para fornecer as estimativas de máxima verossimilhança de  $\boldsymbol{\beta}$  e  $\boldsymbol{\delta}$  pode ser desenvolvido usando o método de escore de Fisher. O processo iterativo conjunto, como apresentado por Cysneiros *et al.* (2007), assume a forma

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \{\mathbf{D}_{\boldsymbol{\beta}}^{T(k)} \mathbf{W}_1^{(k)} \mathbf{D}_{\boldsymbol{\beta}}^{(k)}\}^{-1} \mathbf{D}_{\boldsymbol{\beta}}^{T(k)} \mathbf{D}(\mathbf{g}^{(k)})(\mathbf{y} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(k)})) \quad (3.19)$$

$$\boldsymbol{\delta}^{(k+1)} = \boldsymbol{\delta}^{(k)} + \{\mathbf{Z}^T \mathbf{W}_2 \mathbf{Z}\}^{-1} \mathbf{Z}^T \mathbf{m}, \quad (3.20)$$

para  $k = 0, 1, 2, \dots$ . Contudo, para  $n$  suficientemente grande e sob certas condições de regularidade, os estimadores de máxima verossimilhança  $\hat{\boldsymbol{\beta}}$  e  $\hat{\boldsymbol{\delta}}$  têm distribuição aproximadamente normal com médias  $\boldsymbol{\beta}$  e  $\boldsymbol{\delta}$  e matrizes de covariâncias  $\mathbf{K}_{\boldsymbol{\beta}}^{-1}$  e  $\mathbf{K}_{\boldsymbol{\delta}}^{-1}$ , respectivamente. Portanto,  $\hat{\mathbf{K}}_{\boldsymbol{\beta}}^{-1} = (\mathbf{D}_{\boldsymbol{\beta}}^T \hat{\mathbf{W}}_1 \mathbf{D}_{\boldsymbol{\beta}})^{-1}$  é um estimador consistente da matriz de covariâncias assintótica de  $\hat{\boldsymbol{\beta}}$  e  $\hat{\mathbf{K}}_{\boldsymbol{\delta}}^{-1} = (\mathbf{Z}^T \hat{\mathbf{W}}_2 \mathbf{Z})^{-1}$  é um estimador consistente da matriz de covariâncias assintótica de  $\hat{\boldsymbol{\delta}}$ .

Cysneiros *et al.* (2007) destaca que observações com valores grandes para  $\sigma_i^2$  e valores pequenos para  $v_i$  têm pesos menores no processo de estimação de  $\boldsymbol{\beta}$ . Isso pode ser visto por meio da função  $\mathbf{U}_{\boldsymbol{\beta}}(\boldsymbol{\theta})$  e do processo iterativo (3.19). Para as distribuições normal,  $t$  de Student com  $\nu$  graus de liberdade, logística-II e normal generalizada temos, respectivamente,  $v_i = 1$ ,  $v_i = (\nu + 1)/(\nu + u_i)$ ,  $v_i = [2 \exp(-\sqrt{u_i}) - 1]\{-2\sqrt{u_i}[1 + \exp(-\sqrt{u_i})]\}^{-1}$  e  $v_i = su_i^{(s/2)-1}$ , para todo  $i$ . Com isso, é possível notar que  $v_i$  é inversamente proporcional a  $u_i$  para as distribuições  $t$  de Student, logística-II e normal generalizada quando  $0 < s < 2$ . Dessa forma, para essas distribuições que têm caudas mais pesadas temos que  $\hat{\boldsymbol{\beta}}$  é robusto diante de observações extremas e diante de observações com variância grande. Para o caso normal, a robustez de  $\hat{\boldsymbol{\beta}}$  aparece somente diante dos casos de variância grande. Por outro lado, os aspectos

tos de robustez de  $\hat{\boldsymbol{\delta}}$  depende somente da função  $h(\boldsymbol{\tau})$ . Quando a função  $h(\boldsymbol{\tau})$  assume a forma usual, ou seja,  $h(\boldsymbol{\tau}) = \exp(\boldsymbol{\tau})$ , obtemos as expressões simplificadas para  $m_i = \frac{1}{2}(v_i u_i - 1)$  e  $\mathbf{W}_2 = \frac{(4f_g - 1)}{4} \mathbf{I}_n$ , em que  $\mathbf{I}_n$  é a matriz identidade de ordem  $n$ . Assim, todas as observações assumirão o mesmo peso no processo de estimação (3.20). Consequentemente, a matriz de informação de Fisher para  $\boldsymbol{\delta}$  se reduz a  $\mathbf{K}_\delta = \frac{(4f_g - 1)}{4} \mathbf{Z}^T \mathbf{Z}$ .

### 3.2.1 Testes de heterocedasticidade

No contexto dos modelos simétricos heterocedásticos, é de grande interesse testar a presença de heterocedasticidade. Considerando o modelo (3.13), podemos representar a heterocedasticidade pelas hipóteses  $H_0 : \boldsymbol{\delta}_* = 0$  contra  $H_1 : \delta_j \neq 0$  para pelo menos um  $j = 2, \dots, q$ , em que  $\boldsymbol{\delta}_* = (\delta_2, \dots, \delta_q)^T$ . A variância assintótica de  $\hat{\boldsymbol{\delta}}$  é dada por

$$Var(\hat{\boldsymbol{\delta}}) = \frac{4}{(4f_g - 1)} (\mathbf{Z}^T \mathbf{W}_3 \mathbf{Z})^{-1},$$

em que  $\mathbf{W}_3 = \text{diag}\{w_{3i}\}$  sendo  $w_{3i} = \frac{h_i'^2}{(\sigma_i^2)^2}$  com  $h_i' = \frac{\partial h(\tau_i)}{\partial \tau_i}$  para  $i = 1, \dots, n$ . Com isso, após algumas manipulações algébricas (Cysneiros, 2004), obtemos a variância assintótica de  $\hat{\boldsymbol{\delta}}_*$  dada por

$$Var(\hat{\boldsymbol{\delta}}_*) = \frac{4}{(4f_g - 1)} (\mathbf{R}^T \mathbf{W}_3 \mathbf{R})^{-1},$$

sendo  $\mathbf{R} = \mathbf{Z}_1 - \mathbf{1}_n \mathbf{C}$  em que  $\mathbf{Z}_1$  é dada pela partição  $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{1}_n)$  sendo  $\mathbf{Z}_1 = (\mathbf{z}_2^T, \dots, \mathbf{z}_q^T)^T$  uma matriz  $n \times (q-1)$ ,  $\mathbf{1}_n$  um vetor de uns e  $\mathbf{C} = (\mathbf{1}_n^T \mathbf{W}_3 \mathbf{1}_n)^{-1} \mathbf{1}_n^T \mathbf{W}_3 \mathbf{Z}_1$ . Assim, as estatísticas para os testes da razão de verossimilhanças (RV) e score (SR) para avaliar  $H_0$  contra  $H_1$  são dadas, respectivamente, pelas expressões (3.21) e (3.22).

$$\xi_{RV} = 2\{l(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\delta}}_*, \hat{\delta}_1) - l(\hat{\boldsymbol{\beta}}^0, \hat{\delta}_1^0)\}, \quad (3.21)$$

$$\xi_{SR} = \mathbf{U}_\delta(\hat{\boldsymbol{\theta}}^0)^T \widehat{Var}_0(\hat{\boldsymbol{\delta}}_*) \mathbf{U}_\delta(\hat{\boldsymbol{\theta}}^0), \quad (3.22)$$

em que  $\mathbf{U}_\delta(\hat{\boldsymbol{\theta}}^0)$  e  $\widehat{Var}_0(\hat{\boldsymbol{\delta}}_*)$  são avaliados sob  $H_0$ . Desenvolvendo algebricamente (3.22) obtemos a expressão

$$\xi_{SR} = \frac{1}{(4f_g - 1)} (\hat{\mathbf{r}}_e^0)^T \mathbf{Z}_1 \left\{ \mathbf{Z}_1^T \left( \mathbf{I}_n - \frac{1}{n} \mathbf{J}_n \right) \mathbf{Z}_1 \right\}^{-1} \mathbf{Z}_1^T \hat{\mathbf{r}}_e^0 \quad (3.23)$$

em que  $\hat{\mathbf{r}}_e^0$  é um vetor avaliado sob  $H_0$ , com  $\mathbf{r}_e = (v_1 u_1 - 1, \dots, v_n u_n - 1)^T$  e  $\mathbf{J}_n = \mathbf{1}_n \mathbf{1}_n^T$ . Assintoticamente e sob a hipótese nula, temos que  $\xi_{RV}$  e  $\xi_{SR}$  têm distribuição qui-quadrado com  $(q - 1)$  graus de liberdade (Cysneiros, 2004).

No caso em que a função de ligação de dispersão é dada por  $h(\tau_i) = \exp\{\delta_1 + \sum_{j=2}^q \delta_j z_{ji}\}$  obtemos uma expressão simplificada para  $\mathbf{W}_3$ , ou seja,  $\mathbf{W}_3 = \mathbf{I}_n$ . Dessa forma, a variância assintótica de  $\hat{\boldsymbol{\delta}}_*$  se reduz à

$$Var(\hat{\boldsymbol{\delta}}_*) = \frac{4}{(4f_g - 1)} (\mathbf{R}^T \mathbf{R})^{-1}, \quad (3.24)$$

em que  $\mathbf{R} = \mathbf{Z}_1 - \mathbf{1}_n \bar{\mathbf{Z}}_1$  sendo  $\bar{\mathbf{Z}}_1 = (\bar{z}_2, \dots, \bar{z}_q)$  e  $\bar{z}_j$  a média da coluna  $j$  de  $\mathbf{Z}_1$ , com  $j = 2, \dots, q$ . Com isso, o teste de escore fica dado por

$$\xi_{SR} = \frac{4}{(4f_g - 1)} \hat{\mathbf{m}}_0^T \mathbf{Z}_1 (\mathbf{R}^T \mathbf{R})^{-1} \mathbf{Z}_1^T \hat{\mathbf{m}}_0, \quad (3.25)$$

em que  $\hat{\mathbf{m}}_0$  é um vetor avaliado sob a hipótese nula sendo  $\mathbf{m} = (m_1, \dots, m_n)^T$  com  $m_i = \frac{h'_i}{2\sigma_i^2} (v_i u_i - 1)$ .

### 3.3 Métodos de diagnóstico

Muitos fenômenos podem ser representados por meio de modelos estatísticos. Para validar tais modelos é necessário verificar se as suposições envolvidas estão satisfeitas e avaliar a qualidade e performance dos ajustes dos modelos em questão. A seguir apresentamos e discutimos alguns métodos de diagnóstico.

### 3.3.1 Resíduos

A análise de resíduos é uma etapa importante após o ajuste de um modelo, pois por meio dessa técnica é possível verificar se há afastamentos das suposições feitas para o modelo, ou seja, é possível detectar a presença de observações extremas bem como avaliar a adequação da distribuição proposta para a variável resposta. Dessa forma, os resíduos proporcionam evidências a respeito de observações que não são completamente explicadas pelo modelo. Cysneiros *et al.* (2005) utilizam a metodologia apresentada em Cox & Snell (1968) para determinar o resíduo para os modelos de regressão simétricos.

A metodologia descrita por Cox & Snell (1968) permite definir os resíduos para modelos em que as variáveis respostas são assumidas independentes. Com isso, podemos assumir que os valores da variável de interesse podem ser expressos como

$$y_i = b_i(\boldsymbol{\theta}, \epsilon_i), \quad i = 1, \dots, n, \quad (3.26)$$

em que  $b_1, \dots, b_n$  são funções conhecidas e  $\epsilon_1, \dots, \epsilon_n$  são variáveis aleatórias não observadas, independentes e identicamente distribuídas. Além disso, a distribuição de  $\epsilon_i$  é conhecida. Sendo  $\hat{\boldsymbol{\theta}}$  o estimador de máxima verossimilhança de  $\boldsymbol{\theta}$  podemos escrever (3.26) como

$$y_i = b_i(\hat{\boldsymbol{\theta}}, r_i). \quad (3.27)$$

Supondo que (3.27) tem uma única solução para  $r_i$ , temos que o resíduo de Cox & Snell (1968) pode ser expresso na forma

$$r_i = c_i(y_i, \hat{\boldsymbol{\theta}}). \quad (3.28)$$

Sendo assim, para os modelos simétricos de regressão o resíduo de Cox e Snell

corresponde ao resíduo ordinário dado por

$$r_i(y_i, \hat{\mu}_i, \sigma_i^2) = y_i - \hat{\mu}_i, \quad i = 1, \dots, n, \quad (3.29)$$

com  $\sigma_i^2$  conhecido ou fixo,  $\hat{\mu}_i = \mu_i(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$ ,  $y_i = \mu_i + \epsilon_i$  e  $\epsilon_i \sim S(0, \sigma_i^2)$ .

No geral, mesmo assintoticamente, os resíduos ordinários são viciados e não têm distribuição normal, o que dificulta a verificação do ajuste e a adequação do modelo pelos métodos tradicionais. No entanto, Cysneiros (2004) propõe corrigir, até a ordem  $n^{-1}$ , os dois primeiros momentos do resíduo ordinário  $r_i$  e apresenta uma versão padronizada desse resíduo, denotada  $t_{r_i}$ , que pode ser expressa na forma

$$t_{r_i} = \frac{y_i - \hat{y}_i}{\hat{\sigma}_i \xi^{1/2} \{1 - (4d_g \xi)^{-1} \hat{h}_{ii}\}^{1/2}}, \quad i = 1, \dots, n \quad (3.30)$$

em que  $h_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz  $\mathbf{H} = \mathbf{D}_\beta(\mathbf{D}_\beta^T \mathbf{D}_\beta)^{-1} \mathbf{D}_\beta^T$ .

Além disso, foi verificado através de estudos de simulação que o resíduo proposto tem, respectivamente, média e variância aproximadamente zero e um, uma assimetria desprezível e uma curtose acompanhando a curtose da distribuição do erro. Em um estudo posterior, Cysneiros & Vanegas (2008) propõe resíduos mais abrangentes para os modelos não-lineares simétricos de regressão, sendo eles o resíduo componente do desvio e o resíduo quantal, os quais apresentam média zero, variância um e assimetria desprezível, tornando o resíduo  $t_{r_i}$ , proposto anteriormente, um caso particular dos novos resíduos. Feito isso, foi realizado um estudo analítico e empírico utilizando a definição de Cox & Snell (1968) e simulação de Monte Carlo para descrever o comportamento desses resíduos.

A metodologia descrita para os resíduos, considerando  $\sigma_i^2$  conhecido ou fixo, é desenvolvida de maneira análoga para o caso em que  $\sigma_i^2 = \sigma^2$  para todo  $i = 1, \dots, n$ .

### 3.3.2 Critérios de Seleção

Os critérios de seleção são utilizados para comparar diversos modelos com quantidades diferentes de parâmetros a fim de escolher o modelo que melhor se ajusta aos dados, ou seja, selecionar um modelo com um número reduzido ou suficiente de parâmetros. Para a seleção de modelos pode-se utilizar critérios de informação, como os critérios de Akaike (Akaike, 1972) e de Schwarz (Schwarz, 1978). Além desses métodos, para avaliar a qualidade e performance dos ajustes dos modelos também é comum utilizar o desvio médio absoluto percentual dos resíduos, conhecido como erro de predição. Valores menores desses critérios indicam melhor adequação do modelo.

Os critérios de informação de Akaike (AIC - *Akaike Information Criterion*) e de Schwarz (SBC - *Schwarz Bayesian Criterion*) são dados pelas expressões (3.31) e (3.32), respectivamente.

$$AIC = -2\hat{l} + 2r, \quad (3.31)$$

$$SBC = -2\hat{l} + r \log(n), \quad (3.32)$$

em que  $\hat{l}$  é a função log-verossimilhança maximizada sob o modelo considerado,  $r$  é o número de parâmetros a serem estimados e  $n$  é o tamanho amostral.

De acordo com Kuha (2004), o critério SBC possui um melhor desempenho em relação ao AIC para amostras de tamanho grande, sendo o SBC o mais indicado em caso de divergência entre os critérios.

O desvio médio absoluto percentual dos resíduos (DMAP) é dado pela expressão (3.33). Por meio desse critério podemos medir a capacidade preditiva do modelo em questão.

$$DMAP = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{\mu}_i}{y_i} \right|, \quad (3.33)$$

em que  $\hat{\mu}_i = \mu_i(\mathbf{x}_i, \hat{\boldsymbol{\beta}})$ .

# Capítulo 4

## Resultados da abordagem clássica

### 4.1 Aplicação

Com a finalidade de ilustrar a aplicação da metodologia apresentada anteriormente, consideramos um conjunto de dados com 1309 pares de diâmetro-altura coletados em florestas plantadas de clones de *Eucalyptus sp.*, na região de Itapetininga no sul do estado de São Paulo.

Medir a altura de uma árvore grande é um processo caro, exigindo muito tempo e geralmente sendo feita de forma indireta, o que leva a erros. Desse modo, na prática procura-se medir em uma certa área, chamada parcela, a altura de parte das árvores e o diâmetro de todas. Através dessas alturas e dos diâmetros medidos, é estabelecida uma relação matemática que possibilite estimar as alturas das demais árvores contidas nas parcelas em função dos diâmetros já medidos (Do Couto & Bastos, 1987). Esses modelos matemáticos que relacionam a altura da árvore ( $h$ ) com o diâmetro do tronco ( $d$ ), medidos a  $1,3m$  do solo, são conhecidos como relações hipsométricas.

A dificuldade na escolha do melhor modelo para representar as relações hipsométricas deve-se à não linearidade da relação entre as variáveis envolvidas e às restrições

impostas, por razões biológicas, aos parâmetros do modelo. Esses dois fatores tornam o ajuste desses modelos um problema de regressão não-linear com restrições nos parâmetros. Além do mais, um grande número de variáveis não observáveis tais como: características do solo, luz solar, proximidade entre as árvores, variabilidade genética, entre outras afetam a variação da relação hipsométrica desse tipo de árvore. Esse fato nos conduz aos modelos heterocedásticos, visando assim uma melhor qualidade dos ajustes.

Para representar a relação entre o diâmetro (d) do tronco da árvore e a altura total da árvore (h) para o conjunto de dados em questão, propomos o seguinte modelo heterocedástico

$$y_i = \mu_i(x_i, \boldsymbol{\beta}) + \sigma_i \epsilon_i, \quad (4.1)$$

em que  $y_i$  e  $x_i$  são funções monotônicas, respectivamente, da altura  $h_i$  e do diâmetro  $d_i$  e  $\epsilon_i \sim S(0, 1)$  são erros mutuamente independentes. A estrutura heterocedástica pode ser abordada de uma forma bem geral, no entanto para a nossa aplicação consideramos apenas a estrutura apresentada na equação 4.2

$$\sigma_i^2 = \exp\{\delta_1 + \delta_2 z_i\} = \alpha \exp\{\delta_2 z_i\}, \quad (4.2)$$

sendo  $z_i = (d_i - \bar{d})^2$  com  $\bar{d} = \sum_{i=1}^n d_i/n$  a média dos diâmetros  $d_i$ .

Existe uma grande quantidade de modelos que descrevem a relação entre o diâmetro (d) do tronco da árvore e a sua altura total (h). No entanto, nesse trabalho consideramos apenas os modelos de Stofel e Petterson (Franco *et al.*, 2009), Curtis e Prodan (Curtis, 1967; Moreira *et al.*, 2007) os quais são apresentados na Tabela 4.1.



Tabela 4.1: Relação dos modelos de hipsometria.

Modelos	$y_i$	$x_i$	$\mu_i(x_i, \boldsymbol{\beta})$
Stofel	$h_i$	$\ln(d_i)$	$\exp\{\beta_1 + \beta_2 x_i\}$
Curtis	$h_i$	$1/d_i$	$\exp\{\beta_1 - \beta_2 x_i\}$
Petterson	$h_i$	$1/d_i$	$(\beta_1 + \beta_2 x_i)^{-3}$
Prodan	$h_i$	$1/d_i$	$(\beta_1 x_i^2 + \beta_2 x_i + \beta_3)^{-1}$

Considerando as observações pareadas  $\{(h_1, d_1), \dots, (h_n, d_n)\}$  e usando a relação apropriada para cada modelo dado na Tabela 4.1, temos as 1309 observações  $\{(y_1, x_1), \dots, (y_n, x_n)\}$ , em que  $y_i = h_i$  e  $x_i = k_1(d_i)$ , com  $k_1(\cdot)$  dado na Tabela 4.1.

O principal objetivo dessa aplicação é ajustar os modelos apresentados na Tabela 4.1 ao conjunto de dados, supondo diferentes distribuições simétricas para os erros (normal, normal generalizada,  $t$  de Student e logística-II) e comparar a eficiência dos diferentes modelos. Para determinar o parâmetro de forma  $s$  da distribuição normal generalizada fixamos um conjunto de valores e aplicamos um procedimento de seleção que consiste em minimizar a função SBC, dada pela expressão (3.32), para os valores  $s$  fixados. Encontramos o menor valor para SBC quando  $s = 3$ . Para a distribuição  $t$  de Student o mesmo procedimento é considerado. No entanto, esse procedimento não é capaz de selecionar o melhor grau de liberdade  $\nu$ , pois o valor de SBC diminui quando  $\nu \rightarrow \infty$ , mostrando que nesse caso a distribuição  $t$  de Student tende a uma normal. Contudo, afim de comparar o ajuste obtido pela  $t$  de Student com os demais fixamos  $\nu = 30$ . A implementação computacional foi desenvolvida no software R (R Development Core Team, 2010) utilizando o pacote GAMLSS (Stasinopoulos *et al.*, 2008; Stasinopoulos & Rigby, 2007). Por fim, os ajustes dos modelos são comparados por meio do critério DMAP e o modelo mais versátil aos dados é escolhido baseado nos critérios de informação AIC e SBC e na análise gráfica dos resíduos.

## 4.2 Seleção do modelo

A Figura 4.1 mostra a dispersão entre as variáveis altura e diâmetro de *Eucalyptus sp.* O gráfico indica uma possível presença de heterocedasticidade, pois parece haver uma variabilidade maior das alturas a medida que os diâmetros aumentam. Esse fato justifica a parametrização que utilizamos para  $\sigma_i^2$  dada em (4.2).

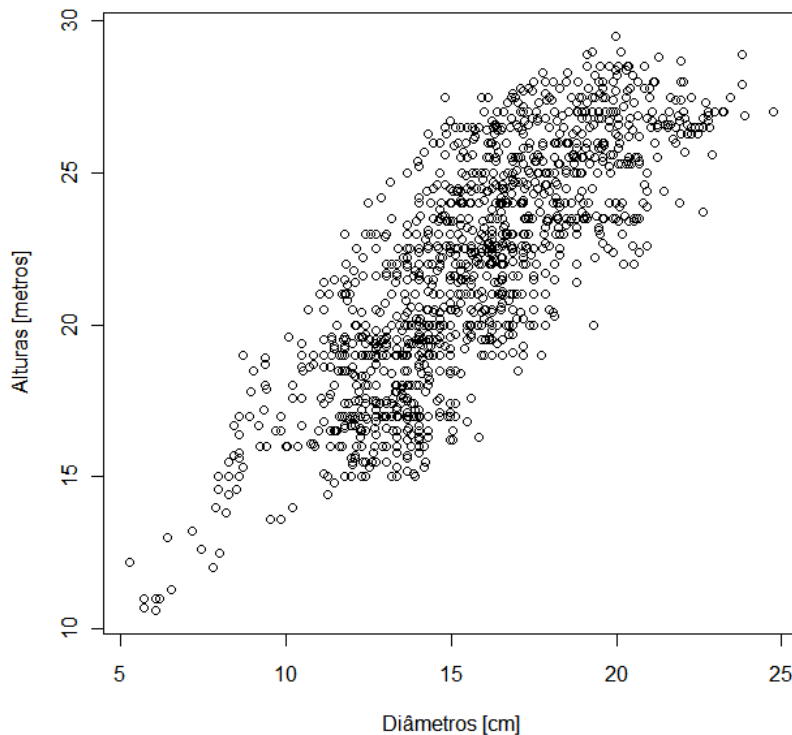


Figura 4.1: Gráfico de dispersão das alturas contra os diâmetros de *Eucalyptus sp.*

A Figura 4.2 apresenta o gráfico dos valores de SBC contra os valores fixados para o parâmetro de forma  $s$  da distribuição normal generalizada. Para os quatro modelos hipsométricos considerados obtemos o menor valor para o critério SBC quando  $s = 3$ .

As estatísticas descritas na Seção 3.2.1 para testar  $H_0 : \delta_2 = 0$  contra  $H_1 : \delta_2 \neq 0$  são significativas ao nível 5%, menos para o teste RV para o modelo de Curtis com a distribuição  $t$  de Student com 30 g.l. (veja tabelas da Seção B.1 do Apêndice B). As Figuras 4.3a, 4.4a, 4.5a e 4.6a mostram o gráfico das alturas reais e das alturas

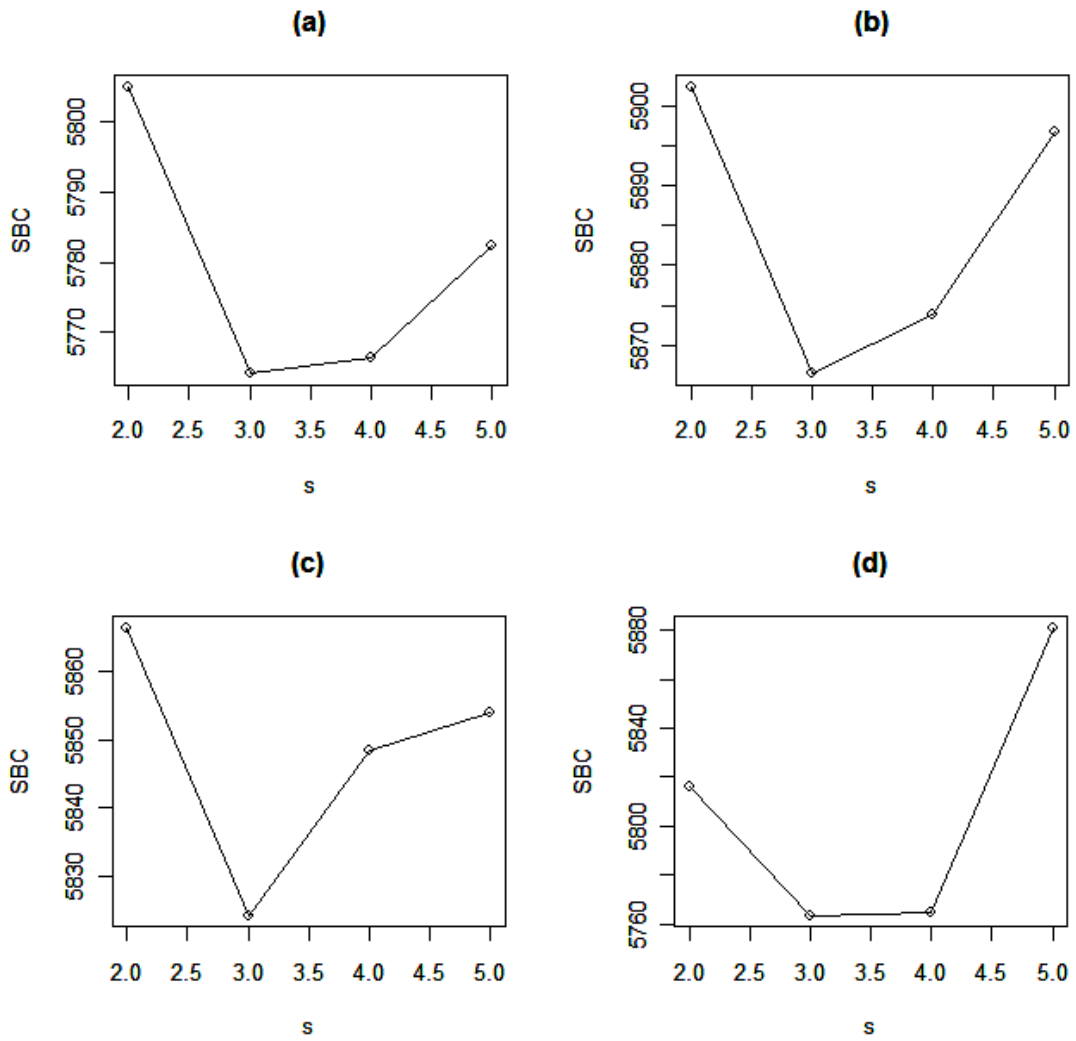


Figura 4.2: Valores fixados para o parâmetro  $s$  da distribuição normal generalizada contra o critério SBC calculado para os modelos hipsométricos de (a) Stofel, (b) Curtis, (c) Petterson e (d) Prodan.

estimadas e, ainda, as Figuras 4.3b, 4.4b, 4.5b e 4.6b mostram o comportamento do parâmetro de escala ( $\sigma_i^2$ ) para as distribuições normal, normal generalizada com  $s = 3$ ,  $t$  de Student com 30 g.l. e logística-II. Podemos observar que, em média, as alturas estimadas estão bem próximas das reais, mostrando assim que os modelos considerados apresentam um bom ajuste. Na Tabela 4.2 apresentamos os resultados da estatística DMAP(%). A análise da mesma comprova a boa performance dos modelos.

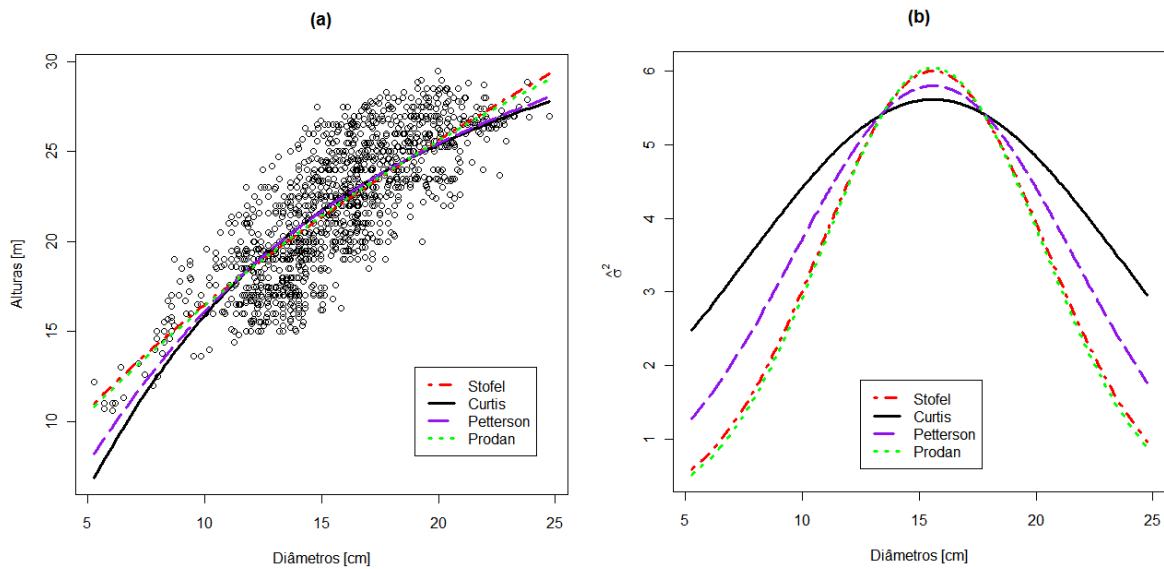


Figura 4.3: Gráfico das estimativas dos parâmetros da distribuição normal. (a) valores estimados para  $\mu_i$  (b) valores estimados para  $\sigma_i^2$ .

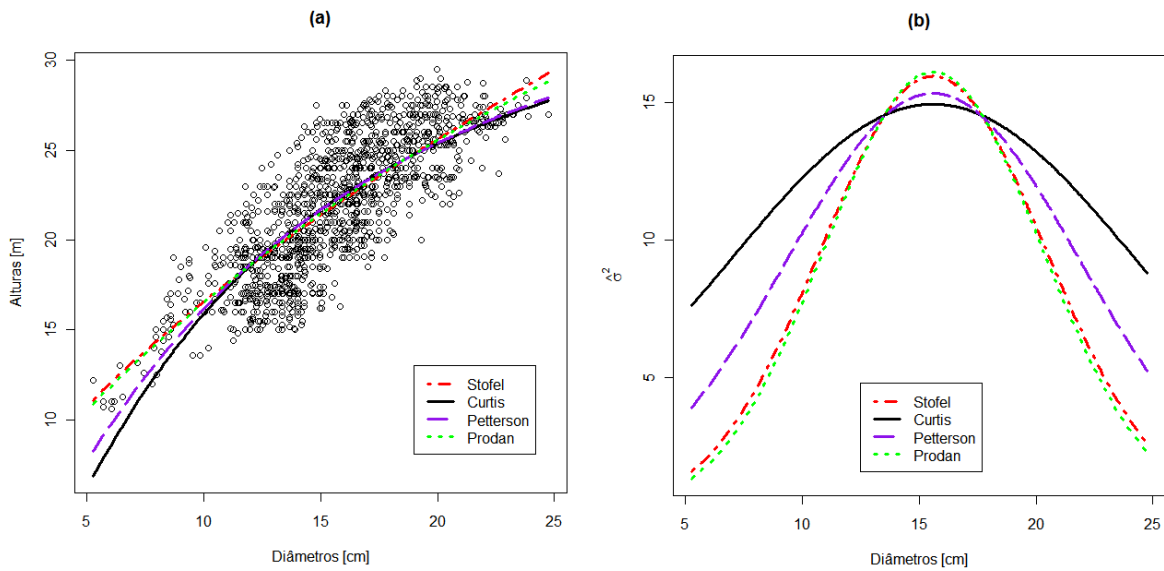


Figura 4.4: Gráfico das estimativas dos parâmetros da distribuição normal generalizada com  $s = 3$ . (a) valores estimados para  $\mu_i$  (b) valores estimados para  $\sigma_i^2$ .

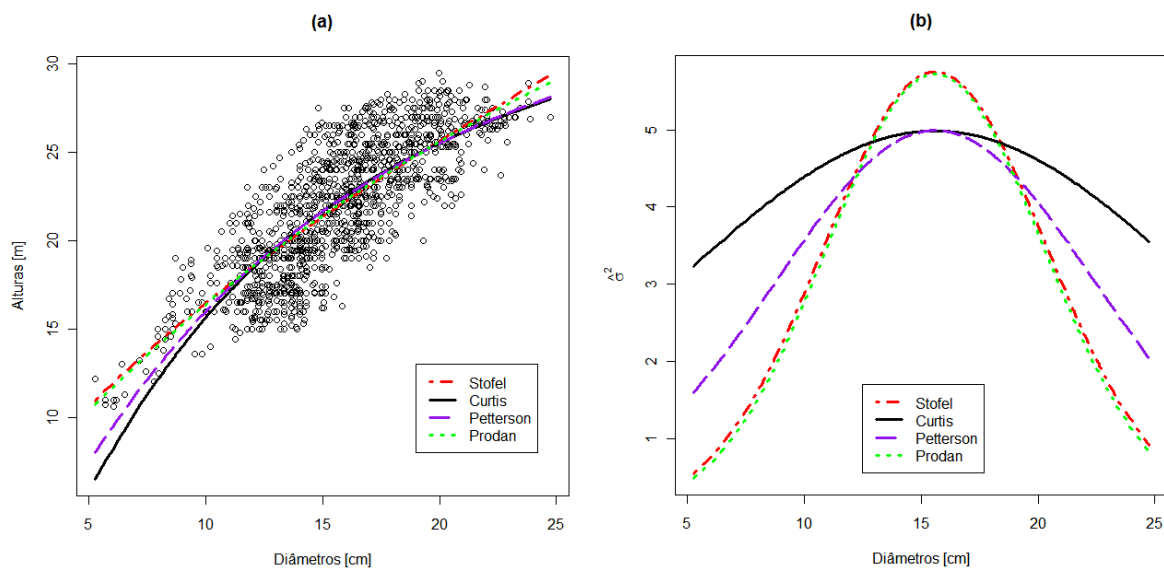


Figura 4.5: Gráfico das estimativas dos parâmetros da distribuição  $t$  de Student com 30 g.l. (a) valores estimados para  $\mu_i$  (b) valores estimados para  $\sigma_i^2$ .

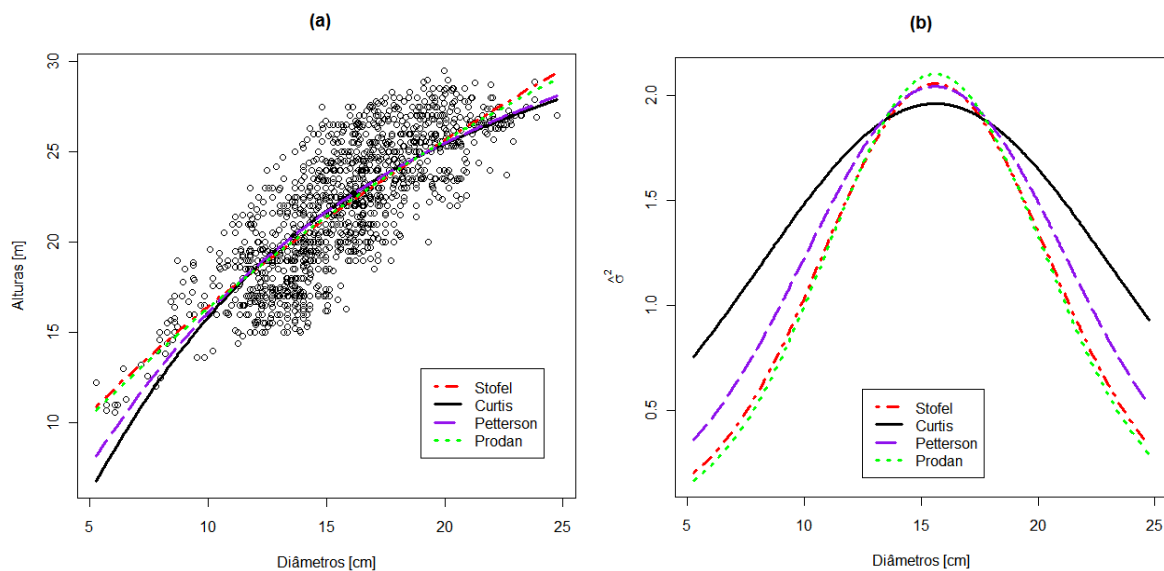


Figura 4.6: Gráfico das estimativas dos parâmetros da distribuição logística-II. (a) valores estimados para  $\mu_i$  (b) valores estimados para  $\sigma_i^2$ .

Tabela 4.2: Comparação dos erros para os modelos ajustados usando o DMAP(%).

Distribuições	DMAP(%)			
	Stofel	Curtis	Petterson	Prodan
Normal	8,732	9,147	9,010	8,739
Normal Generalizada ( $s = 3$ )	8,760	9,138	9,013	8,780
$t$ de Student ( $\nu = 30$ )	8,728	9,135	8,990	8,742
Logística-II	8,716	9,148	9,011	8,724

Para escolher o modelo que melhor representa os dados, nos baseamos na análise da log-verossimilhança maximizada  $\hat{l}$  e dos critérios de seleção AIC e SBC calculados para cada modelo, os quais são apresentados na Tabela 4.3. O modelo dado pela relação hipsométrica de Prodan com a distribuição normal generalizada com  $s = 3$  (Modelo A) é o que apresenta a maior log-verossimilhança ( $-2863, 81$ ) e os menores critérios SBC ( $5763, 51$ ) e AIC ( $5737, 62$ ) dentre os modelos considerados, sendo de acordo com esses critérios o modelo mais indicado para representar os dados.

As estimativas de máxima verossimilhança dos parâmetros do modelo escolhido (Modelo A) são apresentadas na Tabela 4.4. Todos os parâmetros são significativos ao nível 5%.

Na Figura 4.7 apresentamos os percentis da distribuição do Modelo A. A partir disso, é possível verificar que a distribuição normal generalizada com  $s = 3$  representa de maneira satisfatória os dados.

Tabela 4.3: Estatísticas para seleção dos modelos.

Distribuições		Modelos			
		Stofel	Curtis	Petterson	Prodan
Normal	$\hat{l}$	-2888,18	-2936,84	-2915,89	-2884,50
	AIC	5784,35	5881,68	5839,79	5778,99
	SBC	5805,06	5902,39	5860,49	5804,88
Normal Generalizada ( $s = 3$ )	$\hat{l}$	-2867,74	-2918,89	-2897,82	<b>-2863,81</b>
	AIC	5743,47	5845,78	5803,64	<b>5737,62</b>
	SBC	5764,18	5866,49	5824,35	<b>5763,51</b>
$t$ de Student ( $\nu = 30$ )	$\hat{l}$	-2895,49	-2945,39	-2925,16	-2891,78
	AIC	5798,99	5898,79	5858,32	5793,57
	SBC	5819,70	5919,50	5879,03	5819,45
Logística-II	$\hat{l}$	-2922,92	-2970,35	-2949,34	-2919,11
	AIC	5853,84	5948,71	5906,69	5848,22
	SBC	5874,54	5969,41	5927,40	5874,11

Tabela 4.4: Estimativas de máxima verossimilhança para o Modelo A via GAMLSS.

Parâmetros	Estimativa	Erro-padrão	p-valor
$\beta_1$	-0,575	0,227	0,012
$\beta_2$	0,516	0,039	0,000
$\beta_3$	0,015	0,002	0,000
$\alpha$	16,098	0,890	0,000
$\delta_2$	-0,023	0,002	0,000

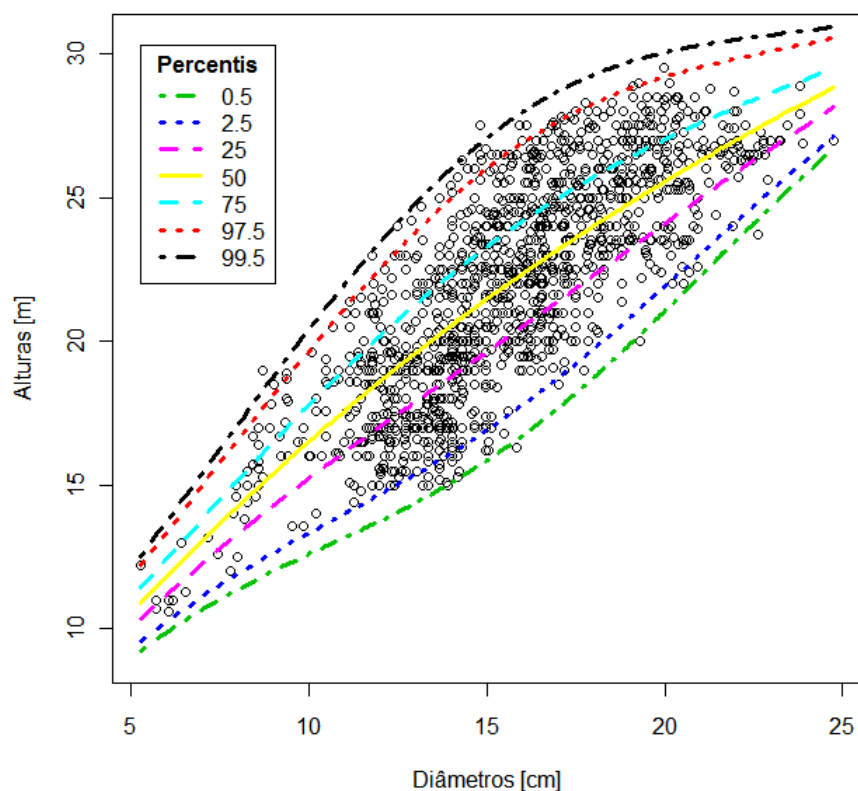


Figura 4.7: Percentis da distribuição normal generalizada com  $s = 3$  quando ajustado o modelo de Prodan.

Com base nesses resultados, avaliamos a qualidade do ajuste do Modelo A por meio dos gráficos dos resíduos apresentados nas Figuras 4.8 e 4.9. Pelas Figuras 4.8a e 4.8b observamos a presença da estrutura heterocedástica e a tendência quadrática para a altura em relação ao diâmetro. O gráfico normal QQ plot (Figura 4.8d) mostra que o conjunto de dados está bem representado pelo modelo escolhido e comprova que a distribuição normal generalizada com  $s = 3$  acomoda bem esses dados, em detrimento às outras distribuições que apresentam muitos pontos distantes de suas caudas superior e/ou inferior (para comparação veja figuras da Seção B.2 do Apêndice B).



Na Tabela 4.5 apresentamos uma análise descritiva dos resíduos do Modelo A. Podemos notar que esses resíduos seguem aproximadamente uma distribuição normal padrão.

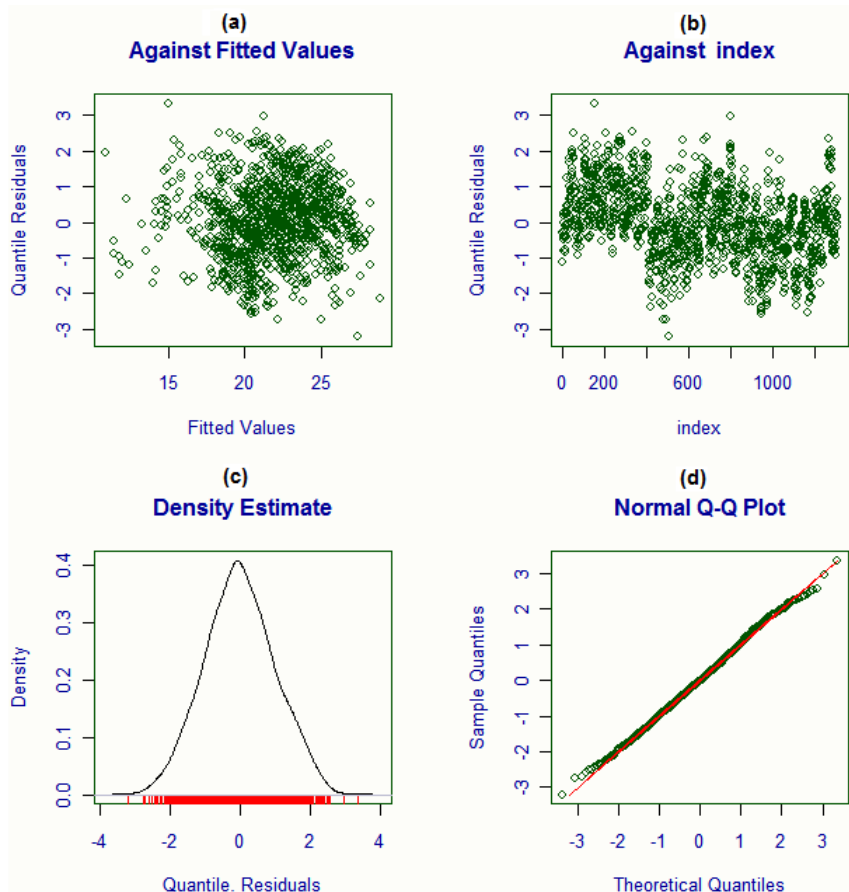


Figura 4.8: Quantis residuais do Modelo A. (a) quantis residuais versus os valores ajustados de  $\mu_i$  (b) quantis residuais versus os índices  $i$  (c) gráfico da densidade estimada (d) gráfico normal QQ plot.

Tabela 4.5: Sumário dos quantis residuais do Modelo A via GAMLSS.

Estatística	Valores
média	-0,005
variância	1,004
coef. de assimetria	0,042
coef. de curtose	2,793

Uma outra forma de analisar a adequação do modelo é por meio do gráfico *worm plot* (van Buuren & Fredriks, 2001). Esses gráficos consistem em ferramentas de diagnóstico para a análise dos resíduos em diferentes regiões (intervalos) da variável explicativa. Se nenhuma covariável é especificada, o gráfico *worm plot* funciona como o gráfico dos quantis normais dos resíduos. Se os pontos estão situados no interior da região de aceitação (entre as duas curvas elípticas), o modelo fornece um bom ajuste. Dessa forma, observando a Figura 4.9 verificamos que para o modelo escolhido todos os pontos estão situados no interior da região de aceitação, indicando um bom ajuste.

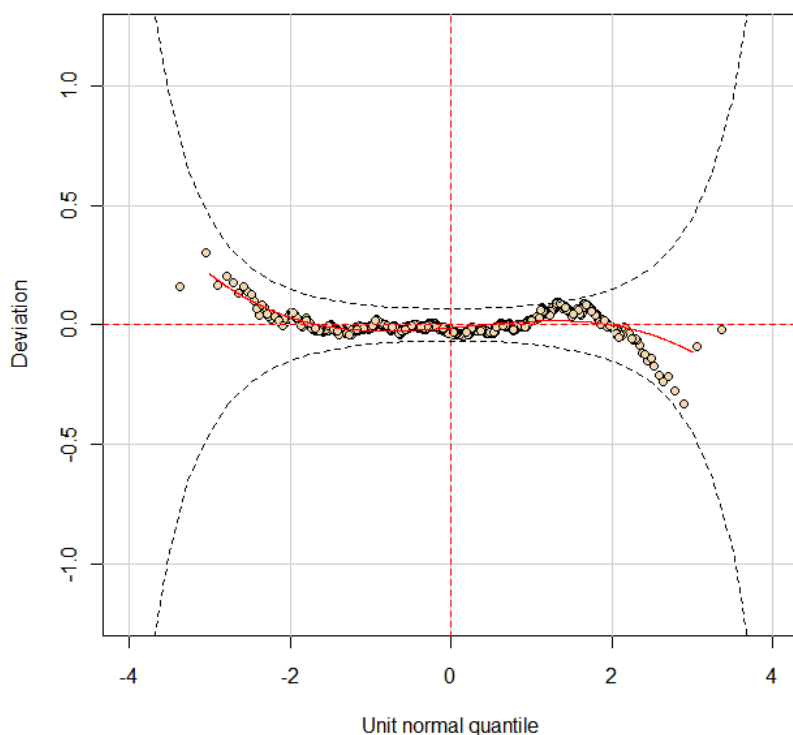


Figura 4.9: Gráfico *worm-plot* para o Modelo A.

As estimativas de máxima verossimilhança bem como os gráficos dos resíduos para os demais modelos considerados nesse trabalho podem ser encontrados na Seção B.2 do Apêndice B.

# Capítulo 5

## Abordagem Bayesiana

Neste capítulo propomos uma abordagem bayesiana baseada em métodos MCMC para fazer inferência sobre modelos de regressão não-lineares heterocedásticos, cuja componente de erro são variáveis aleatórias independentes assumindo diferentes distribuições simétricas. Para detectar observações influentes nos modelos utilizamos o método de análise de influência caso a caso com base na divergência de Kullback-Leibler.

A abordagem bayesiana do problema de estimação dos parâmetros de um modelo assume que estes parâmetros são variáveis aleatórias e qualquer informação inicial associada a eles pode ser modelada por uma função densidade de probabilidade *a priori*. Combinando estas densidades *a priori* com a função de verossimilhança dos dados, por meio do teorema de Bayes, chegamos à função densidade de probabilidade *a posteriori* (Paulino *et al.*, 2003). Denotando o vetor de parâmetros de um modelo por  $\boldsymbol{\theta}$ , a densidade *a priori* por  $\pi_0(\boldsymbol{\theta})$  e a função de verossimilhança associada a um conjunto de observações  $\mathcal{D}$  por  $L(\mathcal{D}|\boldsymbol{\theta})$ , temos que a função densidade de probabilidade

a *posteriori* pode ser expressa na forma

$$\pi(\boldsymbol{\theta}|\mathcal{D}) = \frac{L(\mathcal{D}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})}{\int_{\Theta} L(\mathcal{D}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}}, \quad \boldsymbol{\theta} \in \Theta. \quad (5.1)$$

A integral no denominador de (5.1) é uma integral múltipla sobre o domínio de definição dos parâmetros  $\Theta$  e representa a constante normalizadora da densidade *a posteriori*, assegurando que a integral de  $\pi(\boldsymbol{\theta}|\mathcal{D})$  para  $\boldsymbol{\theta} \in \Theta$  seja igual a 1, sendo assim somente função do conjunto de dados observados  $\mathcal{D}$ . Dessa forma, é comum adotar a notação de proporcionalidade para representar a função densidade de probabilidade *a posteriori* (Box & Tiao, 1973), com isso a forma usual do teorema de Bayes pode ser escrita como

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto L(\mathcal{D}|\boldsymbol{\theta})\pi_0(\boldsymbol{\theta}) \quad (5.2)$$

## 5.1 A função de verossimilhança

Para escrever a função de verossimilhança consideramos  $\mathcal{D} = \{(y_1, \mathbf{x}_1), \dots, (y_n, \mathbf{x}_n)\}$  o conjunto de  $n$  pares de observações em que  $y_i$  é a variável resposta que se relaciona com as covariáveis  $\mathbf{x}_i$  por meio de um modelo, como o modelo dado por (3.13). Considerando que os termos aleatórios  $\epsilon_i$  seguem uma distribuição com função densidade de probabilidade simétrica, a função de verossimilhança do modelo proposto pode ser escrita na forma

$$L(\mathcal{D}|\boldsymbol{\theta}) = \prod_{i=1}^n \sigma_i^{-1} g \left\{ \frac{(y_i - \mu_i(\boldsymbol{\beta}, \mathbf{x}_i))^2}{\sigma_i^2} \right\} \quad (5.3)$$

em que  $g(\cdot)$  é a função geradora de densidades,  $\sigma_i^2 = h(\tau_i)$  com  $\tau_i = \mathbf{z}_i^T \boldsymbol{\delta}$ ,  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T)^T$  o vetor de parâmetros desconhecidos e  $\epsilon_i = (y_i - \mu_i)/\sigma_i$ .

## 5.2 Distribuição *a priori*

O conceito de distribuição *a priori* é um tópico de grande importância da inferência bayesiana, pois ela representa o conhecimento prévio (experiência) de um especialista a respeito de uma quantidade de interesse antes das realizações dos experimentos. A obtenção detalhada de distribuições *a priori* é uma questão problemática, pois envolve descrever as crenças do especialista em uma forma matemática. Nesse sentido, é necessário adotar uma forma de análise que seja capaz de captar essa noção prévia do pesquisador sem exercer grandes efeitos sobre os dados.

Na literatura, existem diversas maneiras de descrever distribuições *a priori* tais como *priori* subjetivas (ou informativas), conjugadas e não-informativas. A utilização de distribuições *a priori* não-informativas é uma maneira de representar as crenças individuais de forma que essas informações sejam mínimas quando comparadas à informação fornecida pelos dados. Nesse contexto, podemos destacar a classe de distribuições *a priori* não informativas propostas por Jeffreys (Paulino *et al.*, 2003). Esta classe é invariante sob transformações porém, em muitos casos, tem a desvantagem de levar a distribuições *a priori* impróprias. Por outro lado, a atribuição de distribuições *a priori* subjetivas também pode ser complicado, visto que cada pesquisador tem uma interpretação pessoal. No entanto, se algum conhecimento específico a respeito dos parâmetros está disponível, ele pode ser usado para especificar a função densidade de probabilidade *a priori*. Contudo, nesse trabalho focamos nas distribuições *a priori* informativas, pois todos os modelos apresentados na Tabela 4.1 apresentam algum tipo de restrição nos parâmetros para que a relação hipsométrica seja biologicamente coerente. Dessa forma, essas restrições são impostas nas distribuições *a priori* escolhidas para cada parâmetro.

### 5.3 Distribuição *a posteriori*

Dada a verossimilhança  $L(\mathcal{D}|\boldsymbol{\theta})$  e uma distribuição *a priori* para o vetor de parâmetros  $\boldsymbol{\theta}$ , para qualquer inferência paramétrica o próximo passo é a obtenção da densidade *a posteriori* que é definida por (5.1). Para isso, vamos considerar o modelo dado por (4.1) e (4.2). Assumindo independência entre os parâmetros, a densidade *a priori* conjunta para  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T)^T$  é dada pelo produto das densidades *a priori* de cada parâmetro, ou seja,

$$\pi_0(\boldsymbol{\theta}) = \pi_0(\boldsymbol{\beta})\pi_0(\boldsymbol{\delta}) \quad (5.4)$$

A partir do histórico dos dados, atribuímos as seguintes distribuições *a priori* para os elementos de  $\boldsymbol{\beta}$  e  $\boldsymbol{\delta}$ :

$$\begin{aligned} \beta_j &\sim N(a_j, b_j); \quad a_j \text{ e } b_j \text{ conhecidos, } j = 1, 2, 3, \\ \alpha &\sim N(c_0, d_0); \quad c_0 \text{ e } d_0 \text{ conhecidos,} \\ \delta_2 &\sim N(e_0, g_0); \quad e_0 \text{ e } g_0 \text{ conhecidos,} \end{aligned} \quad (5.5)$$

em que  $N(a_j, b_j)$ ,  $N(c_0, d_0)$  e  $N(e_0, g_0)$  denotam, respectivamente, distribuições normais com médias  $a_j$ ,  $c_0$ ,  $e_0$  e variâncias  $b_j$ ,  $d_0$ ,  $g_0$ . Nos modelos de Stofel, Curtis e Petterson o parâmetro  $\beta_2$  deve ser obrigatoriamente não negativo, ao passo que no modelo de Prodan os parâmetros  $\beta_2$  e  $\beta_3$  devem ser não negativos. Para incorporar essas restrições consideramos que a função densidade *a priori*  $\pi_0(\boldsymbol{\beta})$  é dada por  $\pi_0(\beta_j) = f_0(\beta_j)I_{(\beta_j, \infty)}$  para  $j = 2, 3$ , sendo  $f_0(\cdot)$  a função densidade de probabilidade do parâmetro  $\beta_j$  e

$$I_{(\beta_j, \infty)} = \begin{cases} 1, & \text{se } \beta_j > 0 \\ 0, & \text{se } \beta_j \leq 0 \end{cases} \quad (5.6)$$

Com isso, a função densidade *a priori* conjunta pode ser escrita como

$$\pi_0(\boldsymbol{\theta}) = \left( \prod_{j=1}^3 \pi_0(\beta_j) \right) \pi_0(\alpha) \pi_0(\delta_2) \quad (5.7)$$

Combinando a função de verossimilhança (5.3) e a distribuição *a priori* (5.7), temos a densidade *a posteriori* conjunta de  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\delta}^T)^T$  dada por

$$\pi(\boldsymbol{\theta}|\mathcal{D}) \propto \prod_{i=1}^n \left[ \sigma_i^{-1} g \left\{ \frac{(y_i - \mu_i(\boldsymbol{\beta}, \mathbf{x}_i))^2}{\sigma_i^2} \right\} \right] \pi_0(\boldsymbol{\theta}) \quad (5.8)$$

Contudo, para se obter amostras representativas de (5.8) e calcular os estimadores bayesianos dos parâmetros de interesse utilizamos o algoritmo de simulação de Monte Carlo via Cadeias de Markov.

## 5.4 Critérios de seleção dos modelos

A determinação de um modelo é um problema fundamental na literatura estatística. No contexto bayesiano diversos critérios podem ser adotados para a seleção de um modelo que melhor se ajusta a um conjunto de dados. No entanto, nesse trabalho consideramos apenas os critérios apresentados nas Seções 5.4.1, 5.4.2, 5.4.3 e 5.4.4.

### 5.4.1 Critério de Informação Bayesiano (ou Schwarz) Esperado (EBIC)

O critério de seleção BIC proposto por Schwarz (1978) foi modificado por Carlin & Louis (2000) para ser aplicado no contexto bayesiano. Esse critério pondera entre o valor esperado *a posteriori* da função log-verossimilhança maximizada e o número

de parâmetros do modelo. O modelo mais versátil ao conjunto de dados é aquele que apresenta o menor valor de EBIC dado por

$$EBIC = E_{\theta|\mathcal{D}}[\log\{L(\mathcal{D}|\theta)\}] - \frac{1}{2}r \log(n). \quad (5.9)$$

#### 5.4.2 Critério de Informação de Akaike Esperado (EAIC)

O critério de informação de Akaike (AIC) (Akaike, 1972) também foi modificado para ser utilizado no contexto bayesiano, sendo dado por

$$EAIC = E_{\theta|\mathcal{D}}[\log\{L(\mathcal{D}|\theta)\}] - r, \quad (5.10)$$

em que  $E_{\theta|\mathcal{D}}[\log\{L(\mathcal{D}|\theta)\}]$  é o valor esperado com relação a densidade *a posteriori* da função log-verossimilhança,  $r$  é o número de parâmetros a serem estimados e  $n$  é o tamanho amostral.

Valores menores de EBIC ou EAIC indicam modelos melhores, porém o critério EBIC tende a selecionar modelos com menor número de parâmetros.

#### 5.4.3 Critério de Informação de Desvio (DIC)

O critério DIC foi introduzido por Spiegelhalter *et al.* (2002) como uma medida de comparação e adequação de modelos com base na distribuição *a posteriori* da estatística desvio. Esse critério pode ser considerado como uma generalização do critério AIC (Ntzoufras, 2009). O DIC é usual em problemas bayesianos de seleção de modelos para os quais as amostras da distribuição *a posteriori* dos parâmetros são obtidas via MCMC.

A estatística desvio pode ser definida como



$$D(\boldsymbol{\theta}) = -2 \log\{L(\mathcal{D}|\boldsymbol{\theta})\} + C(\mathcal{D}), \quad (5.11)$$

em que  $L(\mathcal{D}|\boldsymbol{\theta})$  é a função verossimilhança para um conjunto de observações  $\mathcal{D}$  dado o vetor de parâmetros desconhecidos  $\boldsymbol{\theta}$  e  $C(\mathcal{D})$  é uma função somente dos dados observados. Esta função se cancela, portanto não tem impacto na comparação dos modelos.

O critério DIC é definido como

$$DIC = \bar{D} + p_D = 2\bar{D} - D(\hat{\boldsymbol{\theta}}), \quad (5.12)$$

sendo  $D(\hat{\boldsymbol{\theta}})$  o desvio avaliado na média *a posteriori*, ou seja,  $D(E_{\boldsymbol{\theta}|\mathcal{D}}[\boldsymbol{\theta}])$  e  $p_D$  o número efetivo de parâmetros no modelo, que é dado por  $p_D = \bar{D} - D(\hat{\boldsymbol{\theta}})$  em que  $\bar{D} = E_{\boldsymbol{\theta}|\mathcal{D}}[D]$  é o desvio médio *a posteriori* que mede a qualidade do ajuste do modelo aos dados. Valores menores para DIC indicam modelos melhores.

#### 5.4.4 Critério da Densidade Preditiva Ordenada (CPO)

Um outro critério bastante utilizado para a seleção de modelos é o critério da densidade preditiva condicional ordenada (CPO) (Gelfand *et al.*, 1992). Para aplicar este critério consideremos  $L(\mathcal{D}|\boldsymbol{\theta}) = \prod_{k=1}^n f(y_k|\boldsymbol{\theta})$  a função de verossimilhança associada ao conjunto de observações  $\mathcal{D}$ , em que  $f(y_k|\boldsymbol{\theta})$  é a função densidade de probabilidade de  $y_k$ . Consideremos também  $L(\mathcal{D}_{(-i)}|\boldsymbol{\theta}) = \prod_{k=1, k \neq i}^n f(y_k|\boldsymbol{\theta})$  a função de verossimilhança associada ao conjunto de dados após retirada a  $i$ -ésima observação  $\mathcal{D}_{(-i)}$ . As distribuições *a posteriori* dado o conjunto de observações  $\mathcal{D}$  e  $\mathcal{D}_{(-i)}$  podem ser definidas, respectivamente, por  $\pi(\boldsymbol{\theta}|\mathcal{D}) \propto L(\mathcal{D}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$  e  $\pi(\boldsymbol{\theta}|\mathcal{D}_{(-i)}) \propto L(\mathcal{D}_{(-i)}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$ , em que  $\pi(\boldsymbol{\theta})$  é a distribuição *a priori* de  $\boldsymbol{\theta}$ . Dessa forma, considerando que dado  $\boldsymbol{\theta}$ , as variáveis  $y_i$  são independentes para  $i = 1, \dots, n$  temos que  $L(\mathcal{D}|\boldsymbol{\theta})/L(\mathcal{D}_{(-i)}|\boldsymbol{\theta}) = f(y_i|\boldsymbol{\theta})$  (Cho *et al.*, 2009) e com isso, podemos escrever o CPO <sub>$i$</sub>  para a  $i$ -ésima observação como

$$\text{CPO}_i = \frac{1}{E_{\boldsymbol{\theta}}[\{f(y_i|\boldsymbol{\theta})\}^{-1}|\mathcal{D}]}, \quad i = 1, \dots, n. \quad (5.13)$$

em que

$$E_{\boldsymbol{\theta}}[\{f(y_i|\boldsymbol{\theta})\}^{-1}|\mathcal{D}] = \int_{\Theta} \frac{\pi(\boldsymbol{\theta}|\mathcal{D})}{f(y_i|\boldsymbol{\theta})} d\boldsymbol{\theta}. \quad (5.14)$$

Valores grandes para  $\text{CPO}_i$  implicam em melhor ajuste do modelo. Uma estimativa para  $\text{CPO}_i$  pode ser obtida usando uma amostra gerada pelo MCMC da distribuição *a posteriori*  $\pi(\boldsymbol{\theta}|\mathcal{D})$ . Considerando uma amostra  $\{\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(R)}\}$  de tamanho  $R$  de  $\pi(\boldsymbol{\theta}|\mathcal{D})$  após o *burning*, uma aproximação para  $\text{CPO}_i$  pode ser obtida por meio da expressão (Cancho *et al.*, 2010)

$$\widehat{\text{CPO}}_i = \left\{ \frac{1}{R} \sum_{j=1}^R \frac{1}{f(y_i|\boldsymbol{\theta}^{(j)})} \right\}^{-1}. \quad (5.15)$$

Uma forma resumida do  $\text{CPO}_i$  é dada pela estatística  $B = \sum_{i=1}^n \log(\widehat{\text{CPO}}_i)$ , em que maiores valores para  $B$  indicam melhores ajustes.

## 5.5 Análise de influência caso a caso

Uma maneira comum de analisar a influência de uma observação no ajuste de um modelo é por meio da deleção de casos (Cook & Weisberg, 1992). Para medir a influência da observação  $y_i$  calculamos a divergência de Kullback-Leibler (divergência de K-L) entre as distribuições *a posteriori*  $\pi(\boldsymbol{\theta}|\mathcal{D})$  e  $\pi(\boldsymbol{\theta}|\mathcal{D}_{(-i)})$ , em que  $\mathcal{D}$  é o conjunto completo de observações e  $\mathcal{D}_{(-i)}$  é o conjunto após retirada a  $i$ -ésima observação. Essa medida, denotada por  $K(\pi, \pi_{(-i)})$ , é a divergência de K-L entre  $\pi$  e  $\pi_{(-i)}$  e pode ser calculada como segue

$$K(\pi, \pi_{(-i)}) = \int \pi(\boldsymbol{\theta}|\mathcal{D}) \log \left[ \frac{\pi(\boldsymbol{\theta}|\mathcal{D})}{\pi(\boldsymbol{\theta}|\mathcal{D}_{(-i)})} \right] d\boldsymbol{\theta}. \quad (5.16)$$

Como descrito em Cancho *et al.* (2010),  $K(\pi, \pi_{(-i)})$  mede o efeito de deletar a  $i$ -ésima observação a partir dos dados completos sobre a distribuição *a posteriori* conjunta de  $\boldsymbol{\theta}$ . A calibração de  $K(\pi, \pi_{(-i)})$  pode ser obtida resolvendo a equação  $K(\pi, \pi_{(-i)}) = K[B(0.5), B(p_i)] = -\log\{4p_i(1 - p_i)\}/2$  para  $p_i$ , em que  $B(p)$  denota a distribuição de Bernoulli com probabilidade de sucesso  $p$ . Após algumas manipulações algébricas é possível mostrar que  $p_i = 0,5 [1 + \sqrt{1 - \exp\{-2K(\pi, \pi_{(-i)})\}}]$ . Essa equação implica que  $0,5 \leq p_i \leq 1$ . Dessa forma, se  $p_i \gg 0,5$  o  $i$ -ésimo caso é considerado influente.

Temos ainda que, a expressão (5.16) pode ser escrita em termos de uma esperança *a posteriori*, ou seja

$$\begin{aligned} K(\pi, \pi_{(-i)}) &= \log E_{\boldsymbol{\theta}|\mathcal{D}}[\{f(y_i|\boldsymbol{\theta})\}^{-1}] + E_{\boldsymbol{\theta}|\mathcal{D}}[\log\{f(y_i|\boldsymbol{\theta})\}] \\ &= -\log(\text{CPO}_i) + E_{\boldsymbol{\theta}|\mathcal{D}}[\log\{f(y_i|\boldsymbol{\theta})\}] \end{aligned} \quad (5.17)$$

em que  $E_{\boldsymbol{\theta}|\mathcal{D}}(\cdot)$  denota o valor esperado com respeito a distribuição *a posteriori* conjunta  $\pi(\boldsymbol{\theta}|\mathcal{D})$ . Com isso, podemos obter uma estimativa de (5.17) gerando amostras da distribuição *a posteriori* de  $\boldsymbol{\theta}$  via MCMC. Considerando  $\{\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(R)}\}$  uma amostra de tamanho  $R$  de  $\pi(\boldsymbol{\theta}|\mathcal{D})$  temos

$$K(\widehat{\pi}, \widehat{\pi}_{(-i)}) = -\log(\widehat{\text{CPO}}_i) + \frac{1}{R} \sum_{j=1}^R \log\{f(y_i|\boldsymbol{\theta}^{(j)})\}. \quad (5.18)$$

# Capítulo 6

## Resultados da abordagem bayesiana

### 6.1 Aplicação

Para ilustrar a aplicabilidade da abordagem bayesiana utilizamos os dados de diâmetro-altura de *Eucalyptus sp.* discutidos na Seção 4.1. Nesse contexto, ajustamos novamente os modelos dados pelas relações hipsométricas apresentadas na Tabela 4.1, supondo para os erros dos modelos as distribuições normal, normal generalizada,  $t$  de Student e logística-II. Consideramos restrições nos parâmetros do tipo  $\beta_j > 0$  para  $j = 2, \dots, p$ . Para determinar o parâmetro de forma  $s$  da distribuição normal generalizada consideramos um conjunto de valores  $\{2, 3, 4, 5\}$  e aplicamos um procedimento de seleção que consiste em minimizar a função EBIC, dada por (5.9), para os valores  $s$  fixados. Encontramos o menor valor para EBIC quando  $s = 3$ . Para a distribuição  $t$  de Student esse procedimento não é capaz de selecionar o melhor número de graus de liberdade  $\nu$ , pois o valor de EBIC diminuiu quando  $\nu \rightarrow \infty$ . Diante disso, afim de comparar o ajuste obtido com a distribuição  $t$  de Student aos demais, fixamos  $\nu = 30$ . Para obter as amostras da distribuição *a posteriori* de cada modelo e calcular os estimadores bayesianos dos parâmetros utilizamos métodos MCMC, mais especificamente o amostrador de Gibbs, cuja implementação foi desenvolvida nos sistemas OpenBUGS

(Spiegelhalter *et al.*, 2007) e R (R Development Core Team, 2010). Adotamos distribuições *a priori* independentes com hiperparâmetros conhecidos como definidas em (5.5). Após a escolha dos hiperparâmetros das distribuições *a priori*, geramos duas cadeias paralelas com 60.000 iterações para cada parâmetro. Em cada cadeia, para cada parâmetro, as 10.000 primeiras iterações foram descartadas para eliminar o efeito dos valores iniciais (*burning*). Seleccionamos da segunda parte um valor a cada 100 valores gerados para evitar problemas de autocorrelação, totalizando uma amostra final de tamanho 1000. Para monitorar a convergência das cadeias utilizamos o método de Gelman & Rubin (1992). Para cada amostra calculamos os critérios DIC, EBIC, EAIC e B, os quais são utilizados para comparação dos modelos. Por fim, com as amostras de Gibbs, calculamos as medidas da divergência de Kullback-Leibler e com base nessas medidas analisamos a influência caso a caso para os dados considerados.

## 6.2 Seleção do modelo

A Figura 6.1 apresenta o gráfico com os valores do critério EBIC e os valores fixados para o parâmetro  $s$  da distribuição normal generalizada. Para cada um dos quatro modelos de hipsometria considerados encontramos o menor valor para o critério EBIC quando  $s = 3$ .

Na Tabela 6.1 apresentamos as estimativas de Monte Carlo dos critérios bayesianos EBIC, EAIC, B e DIC com base nas amostras de Gibbs. Na Figura 6.2 ilustramos os gráficos de índices de  $CPO_i$  para os modelos considerados.

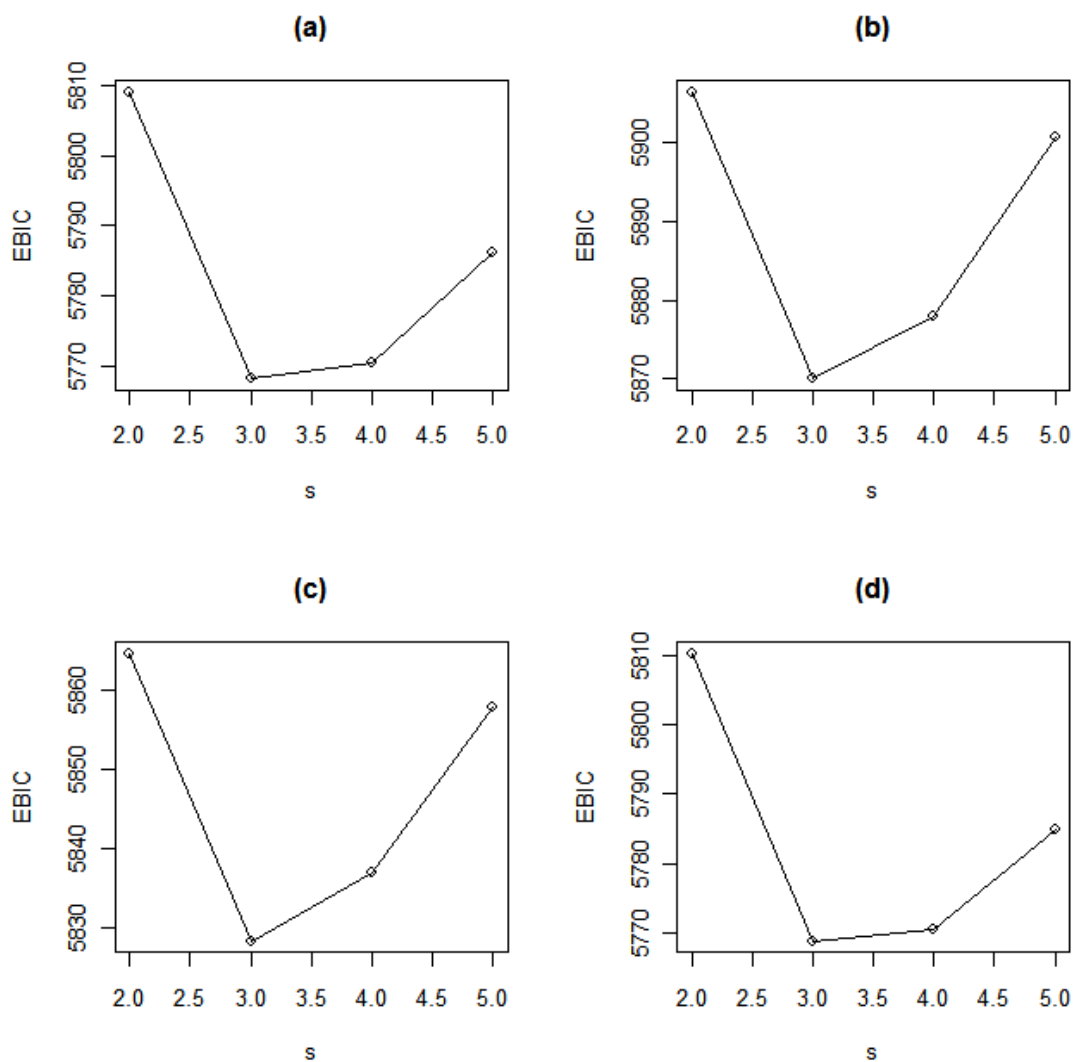


Figura 6.1: Valores fixados para o parâmetro  $s$  da distribuição normal generalizada contra o critério EBIC calculado para os modelos hipsométricos de (a) Stofel, (b) Curtis, (c) Petterson e (d) Prodan.

Na Figura 6.2 podemos notar que todos os modelos considerados apresentam ajustes semelhantes com cada uma das distribuições normal, normal generalizada,  $t$  de Student e logística-II. No entanto, a maior parte das observações estão melhor ajustadas pela distribuição normal generalizada com  $s = 3$  (Figura 6.2b). Ainda, analisando a Figura 6.2b, é possível notar que os modelos de Stofel e Prodan possuem maiores valores de  $CPO_i$  comparados aos modelos de Curtis e Petterson. A análise dos critérios apresentados na Tabela 6.1 indica que a distribuição normal generalizada com  $s = 3$  é a que apresenta melhor ajuste e, dentre os modelos hipsométricos considerados, a relação de

Tabela 6.1: Comparação dos modelos ajustados utilizando diferentes critérios bayesianos.

Distribuições	Critérios	Modelos			
		Stofel	Curtis	Petterson	Prodan
Normal	EBIC	5809,14	5906,43	5864,49	5809,60
	EAIC	5788,44	5885,72	5843,78	5783,72
	B	-2892,08	-2942,56	-2920,92	-2889,43
	DIC	5787,00	5885,00	5843,00	5781,00
Normal Generalizada ( $s = 3$ )	EBIC	5768,17	5870,03	5828,37	<b>5768,75</b>
	EAIC	5747,47	5849,33	5807,66	<b>5742,87</b>
	B	-2871,83	-2925,06	-2903,74	<b>-2870,34</b>
	DIC	5745,00	5848,00	5807,00	<b>5741,00</b>
$t$ de Student ( $\nu = 30$ )	EBIC	5823,83	5920,06	5877,85	5824,79
	EAIC	5803,12	5899,35	5857,14	5798,91
	B	-2899,46	-2948,18	-2927,26	-2897,64
	DIC	5802,00	5898,00	5856,0	5797,000
Logística-II	EBIC	5878,45	5973,54	5931,28	5878,87
	EAIC	5857,74	5952,83	5910,57	5852,99
	B	-2926,70	-2974,87	-2953,72	-2924,22
	DIC	5856,00	5951,00	5909,00	5851,00

Prodan é a que melhor representa a média dos dados. Com isso, o modelo mais versátil aos dados, de acordo com os critérios apresentados na Seção 5.4, é dado pela relação hipsométrica de Prodan com a distribuição normal generalizada com  $s = 3$  (Modelo B). As estimativas *a posteriori* dos parâmetros do Modelo B são apresentadas na Tabela 6.2. Para todos os parâmetros observamos valores de  $\hat{R}$  próximos de 1, indicando a convergência das amostras geradas. As Figuras 6.3 e 6.4 ilustram, respectivamente, o núcleo das densidades *a posteriori* e a trajetória de simulação de cada parâmetro após a seleção da amostra.

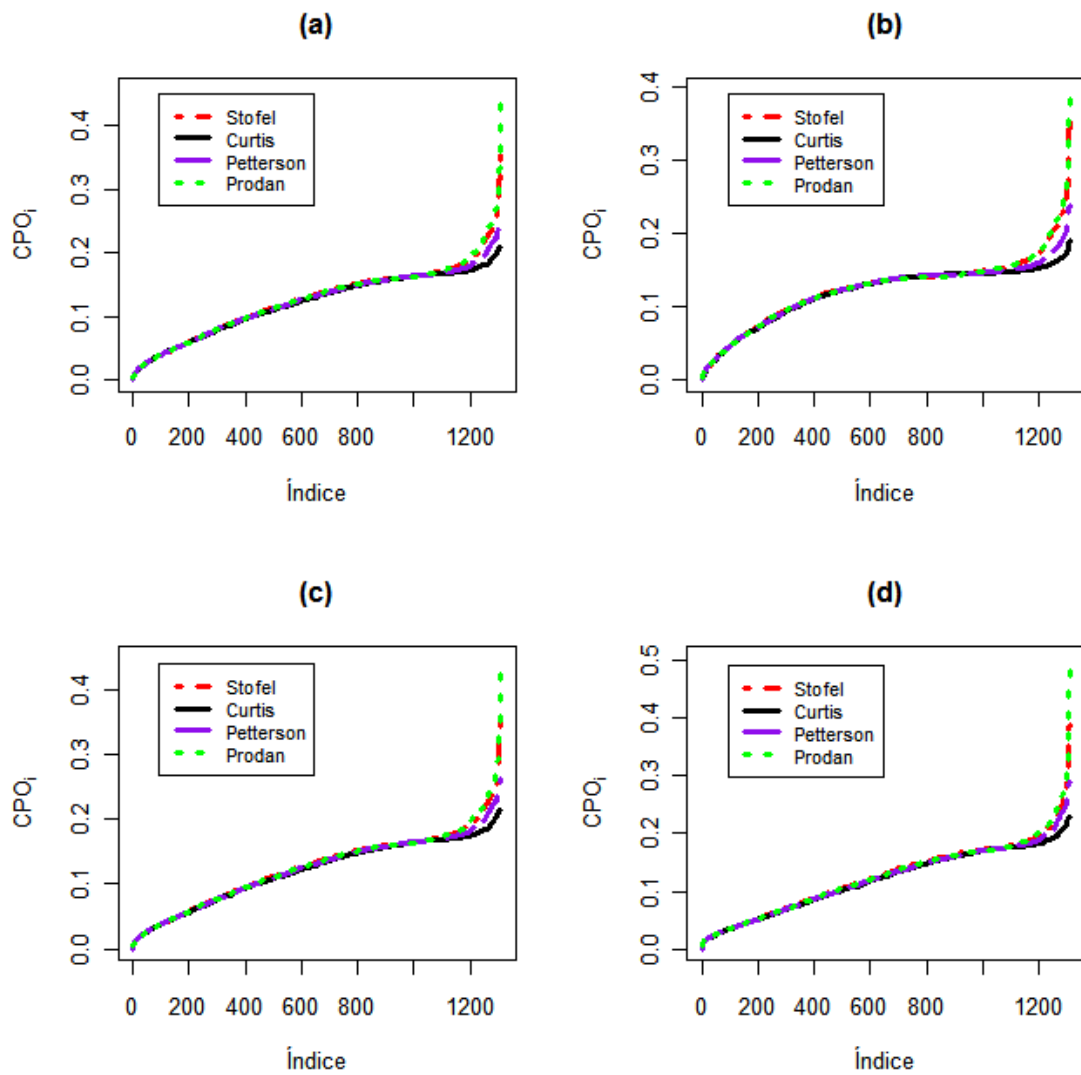


Figura 6.2: Gráfico de índices de  $CPO_i$  para as distribuições (a) normal (b) normal generalizada com  $s = 3$  (c)  $t$  de Student com  $\nu = 30$  e (d) logística-II.



Tabela 6.2: Sumário *a posteriori* para o Modelo B.

Parâmetros	Média	DP	I.C.(95%)	$\hat{R}$
$\beta_1$	-0,533	0,196	(-0,877 ; -0,152)	1,001
$\beta_2$	0,510	0,032	(0,450 ; 0,569)	1,000
$\beta_3$	0,015	0,001	(0,013 ; 0,017)	1,000
$\alpha$	16,253	0,646	(15,045 ; 17,600)	1,005
$\delta_2$	-0,023	0,002	(-0,027 ; -0,019)	1,003

DP - desvio padrão.

I.C.(95%) - intervalo de credibilidade 95%.

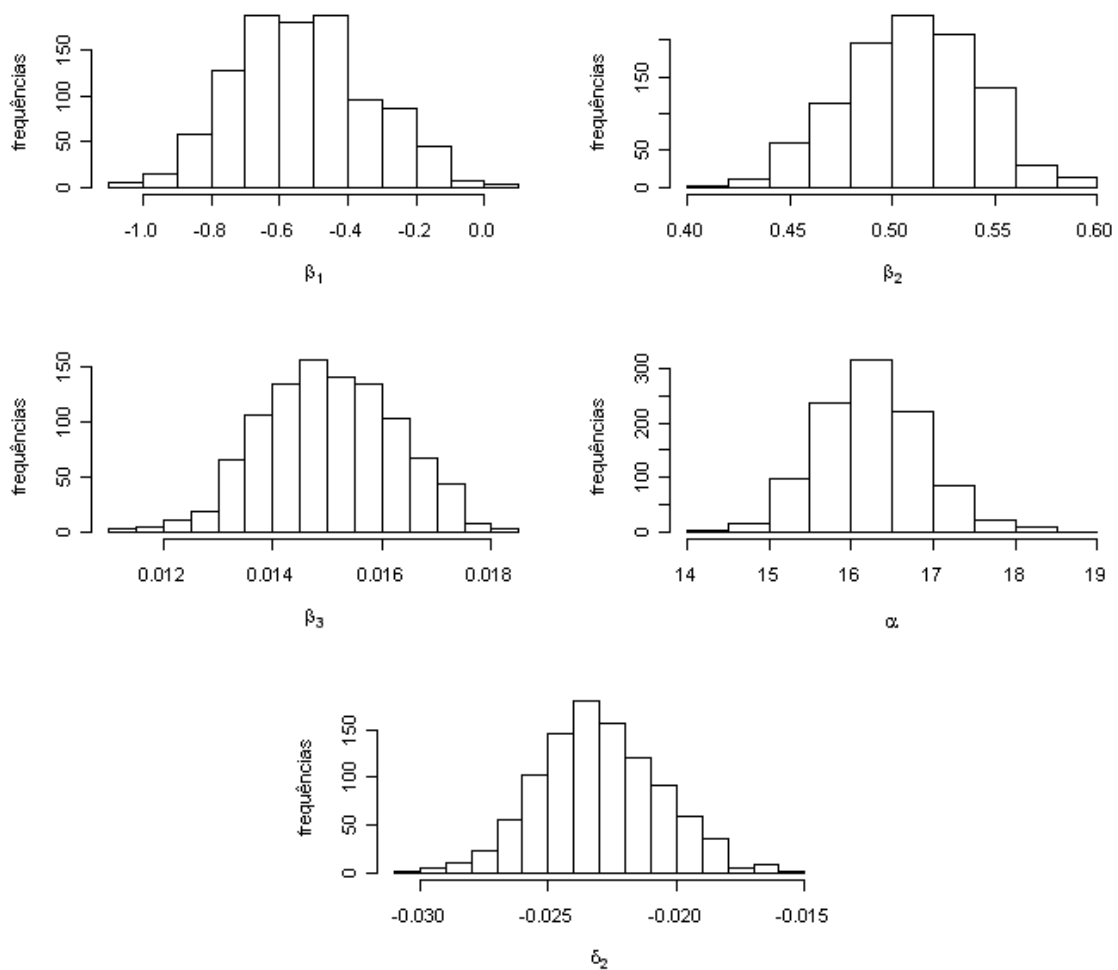


Figura 6.3: Histograma dos parâmetros para o Modelo B.

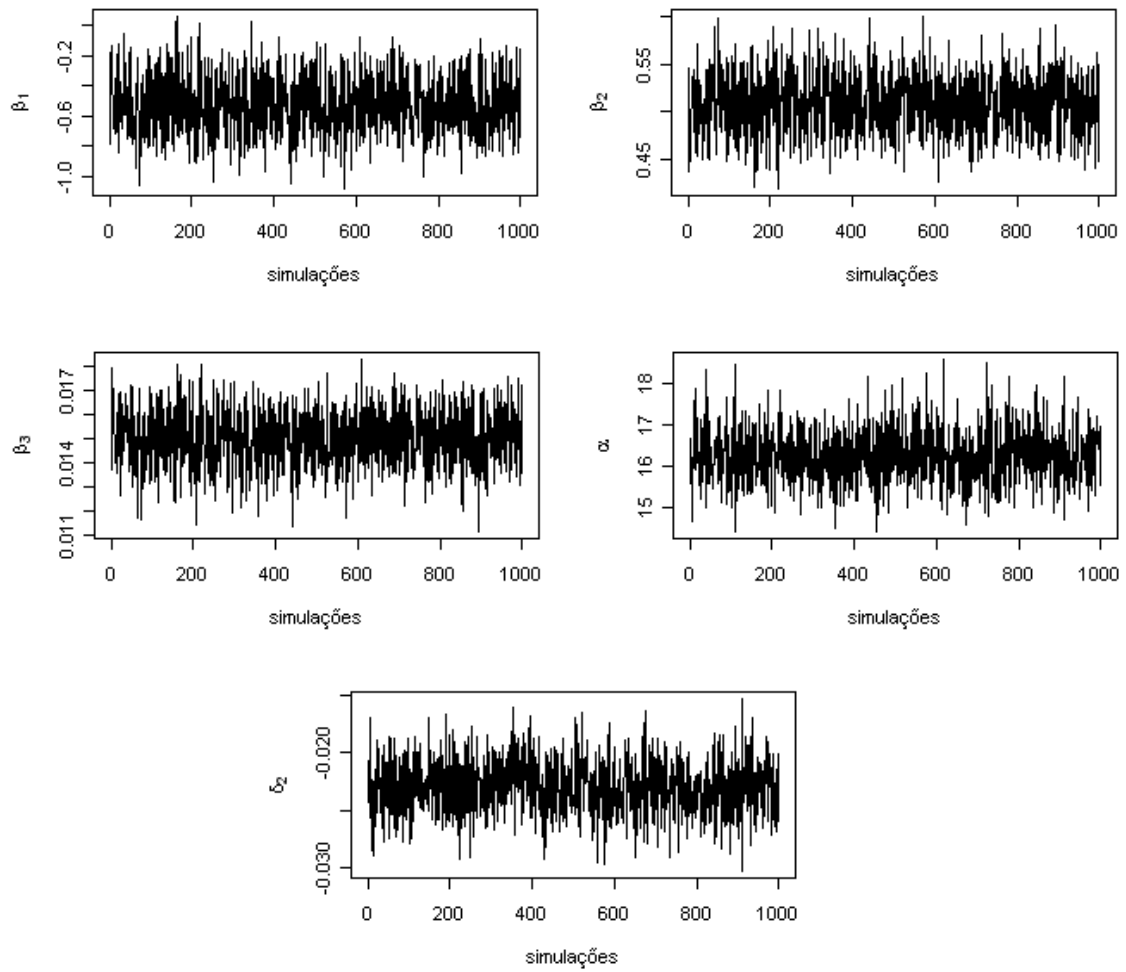


Figura 6.4: Simulação dos parâmetros para o Modelo B.

## 6.3 Influência de observações extremas

### 6.3.1 Dados simulados

Um dos objetivos desse trabalho é mostrar a necessidade de modelos que produzam estimativas robustas frente a presença de observações extremas nos dados. Com esse intuito, vamos considerar um conjunto de dados simulados. Consideramos o modelo

de Stofel apresentado na Tabela 4.1 dado por

$$y_i = \exp\{\beta_1 + \beta_2 x_i\} + \epsilon_i, \quad i = 1, \dots, 100. \quad (6.1)$$

sendo  $y_i = h_i$  e  $x_i = \ln(d_i)$ , em que  $h_i$  e  $d_i$  representam alturas e diâmetros de árvores, respectivamente. Além disso,  $\epsilon_i \sim \text{NG}(0, \sigma_i^2, 3)$  e  $\sigma_i^2$  é dado por (4.2). Os diâmetros  $d_i$  foram gerados a partir de uma distribuição uniforme no intervalo  $[5 ; 25]$ , com isso obtivemos as covariáveis  $x_i$ . Fixamos os valores dos parâmetros em  $\beta_1 = 2$ ,  $\beta_2 = 0,7$ ,  $\alpha = 16$  e  $\delta_2 = -0,05$ . Nos dados simulados, as alturas  $y_i$  variam entre 23,40 e 70,01 com média = 48,739 e desvio padrão = 13,828.

Selecionamos os casos 15, 38 e 79 e perturbamos a variável resposta da seguinte maneira:  $\tilde{y}_i = y_i + S_y$ , em que  $S_y$  são os desvios padrão das variáveis respostas  $y_i$ .

As estimativas MCMC foram obtidas conforme apresentado na Seção 6.1. Na Tabela 6.3 o conjunto de dados (a) denota os dados originais simulados sem perturbação enquanto que os conjuntos de dados (b)-(f) denotam os dados com casos perturbados. A Tabela 6.3 mostra que as estimativas *a posteriori* do modelo ajustado com a distribuição normal generalizada (NG) com  $s = 3$  são sensíveis à perturbação dos casos selecionados, em que  $\delta_2$  passa a ser não significativo nos dados (c). Com o intuito de mostrar a necessidade de modelos que produzam estimativas robustas para lidar com a presença de observações extremas nos dados, ajustamos o modelo (6.1) com erros independentes assumindo distribuição  $t$  de Student (tS). As estimativas MCMC foram obtidas de maneira análoga àquelas sob o modelo cujo erro tem distribuição normal generalizada. Aplicando um procedimento de seleção, que consiste em minimizar a função EBIC dada por (5.9), encontramos os graus de liberdade  $\nu$  da distribuição  $t$  de Student para cada versão (a)-(f) (sem e com perturbação) dos dados originais. Os valores de  $\nu$  são apresentados na Tabela 6.4, bem como as estimativas *a posteriori* obtidas a partir do ajuste da distribuição  $t$  de Student para cada versão perturbada dos dados

originais. Como esperado, as estimativas apresentadas na Tabela 6.4 são menos afetadas do que aquelas sob o ajuste da distribuição normal generalizada, principalmente para o parâmetro  $\delta_2$ .

Tabela 6.3: Dados simulados da NG. Médias e desvios padrão *a posteriori* para os parâmetros do modelo ajustado com a distribuição NG com  $s = 3$ .

Dados	Casos	$\beta_1$		$\beta_2$		$\alpha$		$\delta_2$	
		Média	DP	Média	DP	Média	DP	Média	DP
a	Nenhum	1,996	0,007	0,701	0,002	17,817	2,777	-0,051	0,004
b	15	2,204	0,057	0,627	0,021	15,253	1,769	0,012	0,005
c	38	2,130	0,044	0,655	0,016	15,675	1,827	-0,002*	0,006
d	79	1,999	0,006	0,700	0,002	22,035	1,529	-0,055	0,003
e	{15, 38}	2,295	0,062	0,595	0,023	15,389	1,778	0,019	0,005
f	{15, 38, 79}	2,245	0,055	0,621	0,020	23,739	1,706	0,004*	0,004

\* valor não significativo ao nível de 5%.

Tabela 6.4: Dados simulados da NG. Médias e desvios padrão *a posteriori* para os parâmetros do modelo ajustado com a distribuição tS.

Dados	$\beta_1$		$\beta_2$		$\alpha$		$\delta_2$		$\nu$
	Média	DP	Média	DP	Média	DP	Média	DP	
a	1,997	0,009	0,701	0,003	5,948	1,371	-0,049	0,005	21
b	1,996	0,009	0,701	0,003	4,199	1,042	-0,046	0,006	4
c	1,997	0,010	0,701	0,003	4,418	1,139	-0,046	0,006	4
d	1,997	0,008	0,701	0,003	6,915	1,674	-0,053	0,005	8
e	1,997	0,010	0,701	0,003	3,939	0,986	-0,045	0,006	3
f	1,998	0,009	0,701	0,003	5,439	1,616	-0,050	0,006	3

Tabela 6.5: Dados simulados. Comparação entre os ajustes da NG e da tS usando diferentes critérios bayesianos.

Dados	NG				tS			
	B	DIC	EAIC	EBIC	B	DIC	EAIC	EBIC
a	-150,114	302,500	303,997	314,418	-151,046	304,900	306,249	316,670
b	-280,350	532,600	534,260	544,681	-162,906	328,500	330,097	340,518
c	-259,776	491,000	492,745	503,166	-163,569	330,000	331,359	341,780
d	-191,906	372,000	373,574	383,995	-160,504	322,900	324,496	334,917
e	-288,337	558,800	560,652	571,073	-171,131	345,300	346,569	356,990
f	-309,285	592,700	594,080	604,501	-176,965	356,100	357,602	368,023

Notemos também que o grau de liberdade  $\nu$  funciona como um parâmetro de acomodação das observações extremas, uma vez que este diminui na presença das mesmas. Na Tabela 6.5 apresentamos as estimativas de Monte Carlo dos critérios B, DIC, EAIC e EBIC para cada versão perturbada dos dados originais sob os ajustes da distribuição normal generalizada e  $t$  de Student. Note que o modelo com erro  $t$  de Student apresenta um melhor ajuste na presença das observações extremas.

Considerando as amostras geradas da distribuição *a posteriori* dos parâmetros do modelo ajustado com distribuição normal generalizada e  $t$  de Student, calculamos as medidas da divergência de K-L e suas respectivas calibrações para cada um dos casos (a)-(f) (com e sem perturbação). Esses resultados são apresentados na Tabela 6.6 e indicam que sem perturbação nos dados (a), os casos selecionados não são influentes em ambos os modelos, pois obtivemos valores pequenos para  $K(\pi, \pi_{(-i)})$  e calibrações em torno de 0,61. Entretanto, após perturbação (b)-(f) as medidas  $K(\pi, \pi_{(-i)})$  aumentaram muito e as correspondentes calibrações ficaram próximas ou iguais a 1, indicando que estes casos são influentes quando considerado o modelo com distribuição normal generalizada. Por outro lado, para o modelo  $t$  de Student essas medidas variam pouco,

indicando que esse modelo é mais robusto do que o modelo com a normal generalizada na presença de observações extremas.

Tabela 6.6: Medidas de diagnóstico para os dados simulados.

Dados	Identificação dos casos	NG		tS	
		$K(\pi, \pi_{(-i)})$	Calibração	$K(\pi, \pi_{(-i)})$	Calibração
a	15	0,024	0,608	0,020	0,600
	38	0,001	0,526	0,002	0,533
	79	0,002	0,533	0,002	0,534
b	15	15,889	1,000	0,056	0,662
c	38	16,122	1,000	0,058	0,665
d	79	7,699	0,999	0,311	0,840
e	15	8,099	1,000	0,033	0,627
	38	2,606	0,999	0,033	0,627
f	15	8,110	1,000	0,040	0,638
	38	2,186	0,997	0,040	0,638
	79	5,129	0,999	0,079	0,692

Nas Figuras 6.5 e 6.6 apresentamos os gráficos de índices  $K(\pi, \pi_{(-i)})$  para os dados ajustados com distribuição normal generalizada e  $t$  de Student, respectivamente. Claramente, a divergência de K-L tem um bom desempenho para identificar os casos influentes, apresentando valores maiores de  $K(\pi, \pi_{(-i)})$  para os casos perturbados quando comparados aos outros casos.

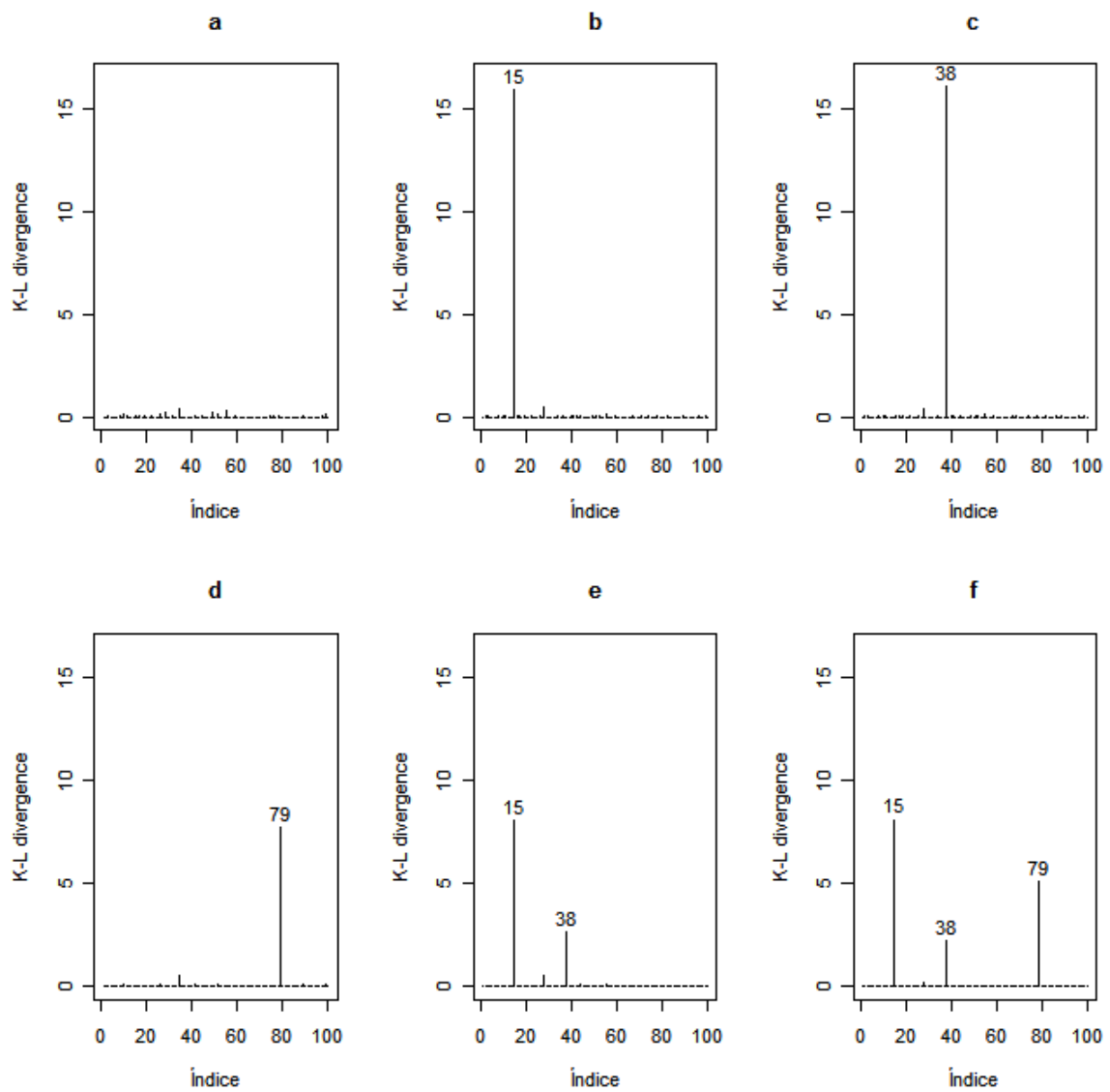


Figura 6.5: Dados simulados da NG. Gráfico de índices de  $K(\pi, \pi_{(-i)})$  a partir do ajuste da NG.

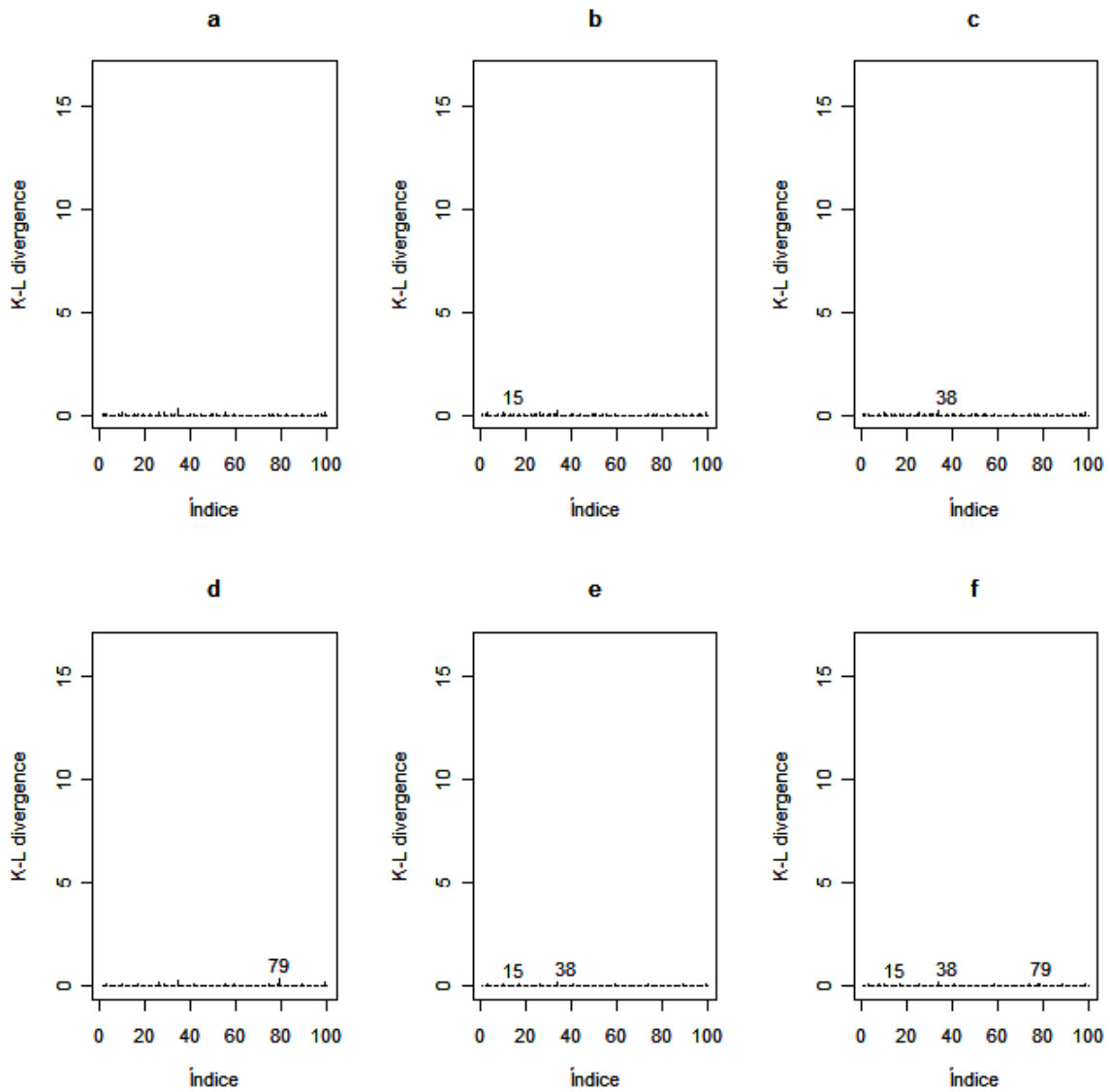


Figura 6.6: Dados simulados da NG. Gráfico de índices de  $K(\pi, \pi_{(-i)})$  a partir do ajuste da tS.

### 6.3.2 Dados de diâmetro-altura de *Eucalyptus sp.*

Nesta seção investigamos a influência das observações no ajuste do modelo escolhido na Seção 6.2 (Modelo B). Com as amostras de Gibbs estimamos as medidas da divergência de K-L para cada uma das observações. Esses resultados são ilustrados na Figura 6.7, sendo que as observações de índices 153, 505 e 988 apresentam valores maiores se comparadas às demais observações.



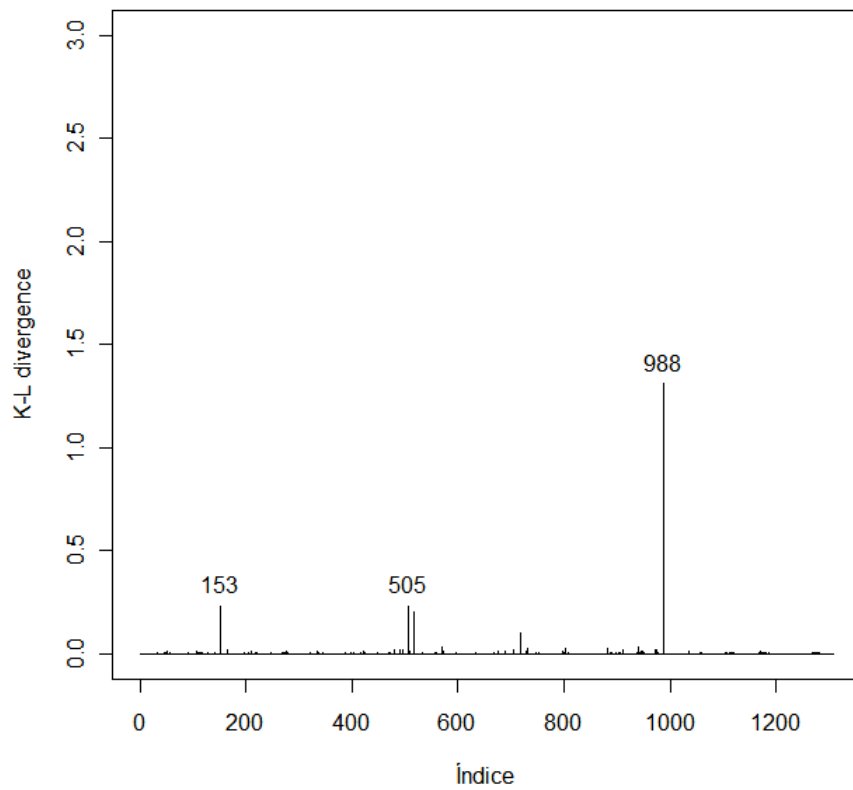


Figura 6.7: Dados de diâmetro-altura de *Eucalyptus sp.*. Gráfico de índices de  $K(\pi, \pi_{(-i)})$  para o Modelo B.

Para verificarmos se essas observações são influentes estimamos a calibração da medida da divergência de K-L. Essas estimativas são apresentadas na Tabela 6.7 conjuntamente com as respectivas estimativas da divergência de K-L, indicando que a observação 988 é influente.

Tabela 6.7: Identificação de casos influentes para os dados de diâmetro-altura de *Eucalyptus sp.*

Identificação do caso			Influência caso a caso	
Observação	Diâmetro	Altura	$K(\pi, \pi_{(-i)})$	Calibração
153	8,72	19,00	0,231	0,804
505	22,63	23,70	0,232	0,804
988	5,28	12,20	1,311	0,982

Na Tabela 6.8 apresentamos as novas estimativas dos parâmetros após a exclusão do caso influente.

Tabela 6.8: Estimativas para os dados de diâmetro-altura de *Eucalyptus sp.* após retirado o caso influente.

Parâmetros	Média	I.C.(95%)
$\beta_1$	-0,292	(-0,749 ; 0,187)
$\beta_2$	0,476	(0,404 ; 0,554)
$\beta_3$	0,016	(0,013 ; 0,019)
$\alpha$	16,330	(14,948 ; 17,357)
$\delta_2$	-0,025	(-0,029 ; -0,019)

Notamos uma mudança apenas na estimativa do parâmetro  $\beta_1$  quando excluimos a observação influente. Observamos que a estimativa deste parâmetro mostrou-se sensível em relação ao ponto influente, observando-se a ocorrência de mudança de sinal no valor do parâmetro. Os valores para os critérios de seleção EBIC, EAIC, B e DIC são, respectivamente, 5763,05 , 5737,17 , -2866,39 e 5735,00. Comparados com os valores da Tabela 6.1 percebemos uma variação muito pequena, indicando que o modelo não é afetado pelo caso influente.

# Capítulo 7

## Conclusão

Nesse trabalho tratamos de duas técnicas de ajuste para os modelos não-lineares heterocedásticos simétricos: a abordagem clássica utilizando os modelos GAMLSS e a abordagem bayesiana usando densidades *a priori* informativas conjuntamente com técnicas de simulação de Monte Carlo em Cadeia de Markov (MCMC). Tratamos também de uma análise de influência com base na divergência de Kullback-Leibler.

Para ilustrar a aplicação de ambos os métodos, clássico e bayesiano, utilizamos os dados de diâmetro-altura de *Eucalyptus sp.* Ajustamos os modelos dados pelas relações hipsométricas de Stofel, Curtis, Petterson e Prodan, nos quais supomos erros com distribuições simétricas e uma estrutura heterocedástica. Tanto o método de máxima verossimilhança irrestrito quanto o método bayesiano apresentaram ajustes satisfatórios e semelhantes, validando as duas técnicas. Dentre os modelos estudados, o modelo de Prodan conjuntamente com a distribuição normal generalizada com  $s = 3$  mostrou-se o modelo mais versátil aos dados em ambos os casos.

No entanto, devido ao pequeno tamanho das amostras que geralmente são utilizadas em situações práticas, as estimativas irrestritas resultam, em muitos casos, em valores inconsistentes biologicamente, ou seja, estimativas negativas que não permitem predi-

zer os valores das alturas das árvores quando considerados diâmetros muito pequenos ou muito grandes (Moreira *et al.*, 2007). Nessas situações, para resolver o problema a solução é incorporar restrições nos parâmetros dos modelos. Esse fato nos mostra que a metodologia bayesiana torna-se mais viável, mostrando-se de fácil aplicação e entendimento. As restrições presentes nos parâmetros dos modelos podem ser facilmente modeladas por meio da função densidade de probabilidade *a priori* adotada.

Podemos constatar no entanto, que o uso da abordagem bayesiana na prática fornece inferências mais precisas, o que não é possível observar em todos os casos com métodos clássicos irrestritos. Verificamos ainda, que a divergência de K-L detectou de maneira eficiente os pontos influentes no modelo, o que pudemos comprovar por meio do estudo de simulação e dos dados reais.

# Referências Bibliográficas

- Akaike, H. (1972). Use of an information theoretic quantity for statistical model identification. In *Proc. 5th Hawaii Int. Conf. System Sciences*, pages 249–250.
- Arellano-Valle, R. (1994). *Distribuições elípticas: propriedades, inferência e aplicações a modelos de regressão*. Ph.D. thesis, Tese de Doutorado, Instituto de Matemática e Estatística-USP, São Paulo, Brasil.
- Berkane, M. & Bentler, P. (1986). Moments of elliptically distributed random variates. *Statistics and Probability Letters*, **4**(6), 333–335.
- Blake, I. & Thomas, J. (1968). On a class of processes arising in linear estimation theory. *Information Theory, IEEE Transactions on*, **14**(1), 12–16.
- Box, G. E. P. & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*, volume 588. Addison-Wesley Reading, MA.
- Cambanis, S., Huang, S. & Simons, G. (1981). On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, **11**(3), 368–385.
- Cancho, V. G., Dey, D. K., Lachos, V. H. & Andrade, M. G. (2010). Bayesian nonlinear regression models with scale mixtures of skew normal distributions: Estimation and case influence diagnostics. *Computational Statistics and Data Analysis*, **55**, 588–602.
- Carlin, B. P. & Louis, T. A. (2000). *Bayes and empirical Bayes methods for data analysis*. New York, NY: Chapman and Hall.

- Chmielewski, M. (1981). Elliptically symmetric distributions: A review and bibliography. *International Statistical Review/Revue Internationale de Statistique*, **49**(1), 67–74.
- Cho, H., Ibrahim, J. G., Sinha, D. & Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics*, **65**(1), 116–124.
- Chu, K. (1973). Estimation and decision for linear systems with elliptical random processes. *IEEE Transactions on Automatic Control*, **18**, 499–505.
- Cook, R. D. & Weisberg, S. (1992). *Residuals and influence in regression*. Chapman and Hall New York.
- Cordeiro, G. (2004). Corrected likelihood ratio tests in symmetric nonlinear regression models. *Journal of Statistical Computation and Simulation*, **74**(8), 609–620.
- Cordeiro, G. & Andrade, M. (2009). Transformed symmetric models. *Statistical Modelling*, (artigo submetido).
- Cordeiro, G., Ferrari, S., Uribe-Opazo, M. & Vasconcellos, K. (2000). Corrected maximum-likelihood estimation in a class of symmetric nonlinear regression models. *Statistics and Probability Letters*, **46**(4), 317–328.
- Cox, D. R. & Hinkley, D. V. (1974). *Theoretical statistics*. Chapman & Hall/CRC.
- Cox, D. R. & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series B (Methodological)*, **30**(2), 248–275.
- Curtis, R. O. (1967). Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *Forest Science*, **13**(4).
- Cysneiros, F. (2004). *Métodos Restritos e Validação de Modelos Simétricos de Regressão*. Ph.D. thesis, Tese de doutorado, Departamento de Estatística, Universidade de São Paulo, Brasil.

- Cysneiros, F. & Vanegas, L. (2008). Residuals and their statistical properties in symmetrical nonlinear models. *Statistics & Probability Letters*, **78**(18), 3269–3273.
- Cysneiros, F. J. A., Paula, G. A. & Galea, M. (2005). Modelos Simétricos Aplicados. *São Pedro: 9a Escola de Modelos de Regressão*.
- Cysneiros, F. J. A., Paula, G. A. & Galea, M. (2007). Heteroscedastic symmetrical linear models. *Statistics & Probability Letters*, **77**(11), 1084–1090.
- Cysneiros, F. J. A., Cordeiro, G. M. & Cysneiros, H. M. A. (2010). Corrected maximum likelihood estimators in heteroscedastic symmetric nonlinear models. *Journal of Statistical Computation and Simulation*, **80**(4), 451–461.
- Do Couto, H. T. Z. & Bastos, N. L. M. (1987). Modelos de equações de volume e relações hipsométricas para plantações de Eucalyptus no Estado de São Paulo. *IPEF*, (37), 33–44.
- Dominguez-Molina, J. A., González-Farías, G. & Rodríguez-Dagnino, R. M. (2003). A practical procedure to estimate the shape parameter in the generalized Gaussian distribution. *technique report I-01-18\_eng.pdf*, available through [http://www.cimat.mx/reportes/enlinea/I-01-18\\_eng.pdf](http://www.cimat.mx/reportes/enlinea/I-01-18_eng.pdf).
- Fang, K., Kotz, S. & Ng, K. (1990). *Symmetric multivariate and related distributions*, volume 378. Chapman and Hall London.
- Ferrari, S. & Uribe-Opazo, M. (2001). Corrected likelihood ratio tests in a class of symmetric linear regression models. *Brazilian Journal of Probability and Statistics*, **15**, 49–67.
- Franco, E. J., Scolforo, J. R. S., Mello, J. M. & Lima, J. T. (2009). Eficiência na estimativa do peso seco para árvores individuais e definição do ponto ótimo de amostragem para determinação da densidade básica de Eucalyptus camaldulensis. *Ciência Florestal*, **8**(1).

- Gelfand, A. E., Dey, D. K. & Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods. In *Bayesian statistics: proceedings of the Fourth Valencia International Meeting, April 15-20, 1991*, volume 4, pages 147–167. Oxford University Press, USA.
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, **7**(4), 457–472.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). Continuous univariate distributions, Volume II, 2nd edn. *Wiley, New York*.
- Kelker, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, **32**(4), 419–430.
- Kuha, J. (2004). AIC and BIC: Comparisons of Assumptions and Performance. *Sociological Methods & Research*, **33**(2), 188.
- Lachos, V. H., Bandyopadhyay, D. & Garay, A. M. (2011). Heteroscedastic nonlinear regression models based on scale mixtures of skew-normal distributions. *Statistics & Probability Letters*.
- Lange, k. (2004). *Optimization*. Springer Texts in Statistics, Springer.
- Lange, K., Little, R. & Taylor, J. (1989). Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, pages 881–896.
- Little, R. (1988). Robust estimation of the mean and covariance matrix from data with missing values. *Applied Statistics*, **37**(1), 23–38.
- McGraw, D. K. & Wagner, J. F. (1968). Elliptically symmetric distributions. *Information Theory, IEEE Transactions on*, **14**(1), 110–120.



- Moreira, M. F. B., Loibel, S., Thiersch, C. R. & Andrade, M. G. (2007). O Uso de Métodos Bayesianos Empíricos para Ajustar Relações Hipsométricas com o Modelo de Curtis com Restrições nos Parâmetros. *Rev. Bras. Biom*, **25**(3), 23–38.
- Muirhead, R. (1980). The effects of symmetric distributions on some standard procedures involving correlation coefficients. *Multivariate Statistical Theory Analysis (ed. RP Gupta)*, North-Holland, pages 143–159.
- Muirhead, R. (1982). *Aspects of multivariate statistical theory*. Wiley Online Library. ISBN 0471094420.
- Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics*, **32**(7), 685–694.
- Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. John Wiley & Sons Inc.
- Paulino, C. D., Turkman, M. A. A. & Murteira, B. (2003). *Estatística Bayesiana*. Lisboa: Fundação Calouste Gulbenkian.
- R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rao, B. (1990). Remarks on univariate symmetric distributions. *Statistics and Probability Letters*, **10**, 307–315.
- Russo, C. M., Paula, G. A. & Aoki, R. (2009). Influence diagnostics in nonlinear mixed-effects elliptical models. *Computational Statistics & Data Analysis*, **53**(12), 4143–4156.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, **6**(2), 461–464.
- Spiegelhalter, D., Thomas, A., Best, N. & Lunn, D. (2007). *OpenBUGS User Manual*. MRC Biostatistics Unit, Cambridge, version 3.0.2 edition.

- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4), 583–639.
- Stasinopoulos, D. M. & Rigby, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS) in R. *Journal of Statistical Software*, **23**(7), 1–46.
- Stasinopoulos, M., Rigby, B. & Akantziliotou, C. (2008). *Instructions on how to use the gamlss package in R Second Edition*.
- Taylor, J. (1992). Properties of modelling the error distribution with an extra shape parameter. *Computational Statistics and Data Analysis*, **13**(1), 33–46.
- Uribe-Opazo, M., Ferrari, S. & Cordeiro, G. (2008). Improved score tests in symmetric linear regression models. *Communications in Statistics-Theory and Methods*, **37**(2), 261–276.
- van Buuren, S. & Fredriks, M. (2001). Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, **20**(8), 1259–1277.
- Varanasi, M. K. & Aazhang, B. (1989). Parametric generalized Gaussian density estimation. *The Journal of the Acoustical Society of America*, **86**, 1404–1415.
- Yamaguchi, K. (1990). Generalized EM algorithm for model with contaminated error term. In *Proceedings of the Seven Japan and Korea Joint Conference of Statistics*, pages 107–114.

# Apêndice A

## Características de algumas distribuições simétricas

Nesta seção apresentamos as demais distribuições simétricas citadas durante o desenvolvimento dessa dissertação bem como algumas de suas características. Para todas essas distribuições  $u = [(y - \mu)/\sigma]^2$ , em que  $Y \sim S(\mu, \sigma^2)$  na reta real. Para as distribuições que satisfazem condições como a existência dos momentos temos que  $E(Y) = \mu$  e  $Var(Y) = \xi\sigma^2$ .

Tabela A.1: Algumas distribuições pertencentes à classe simétrica.

Distribuição	Função geradora de	$\xi$	Curtose
	densidades $g(u)$		
Cauchy	$\frac{1}{\pi(1+u)}$	não existe	não existe
$t$ -Student generalizada	$\frac{s^{r/2}}{B(1/2, r/2)}(s+u)^{-(r+1)/2}$ , $s, r > 0$	$\frac{s}{(r-2)}$ , $r > 2$	$3 + \frac{6}{(r-4)}$ , $r > 4$
Logística- $I$	$\frac{c}{e^u(1+e^{-u})^2}$	0, 79569	2, 385165
Logística generalizada	$\frac{\alpha}{B(m, m)} \left\{ \frac{e^{-\alpha\sqrt{u}}}{(1+e^{-\alpha\sqrt{u}})^2} \right\}^m$ , $\alpha, m > 0$	$\frac{2\psi'(m)}{\alpha(m)}$	$3 + \frac{\psi'''(m)}{2\psi'(m)^2}$
Exponencial dupla	$\frac{1}{2} \exp\{-\sqrt{u}\}$	2	6

$B(\cdot, \cdot)$  é a função beta;  $c \approx 1,484300029$ .  
 $\psi'(\cdot)$ ,  $\psi'''(\cdot)$  a primeira e terceira derivadas da função digama.

Tabela A.2: Continuação da Tabela A.1.

Distribuição	Função geradora de densidades $g(u)$	$\xi$	Curtose
Exponencial potência	$\frac{\Gamma(1 + \frac{1+k}{2})2^{1+(1+k)/2}}{\exp\{-\frac{1}{2}u^{1/(1+k)}\}}, \quad -1 < k \leq 1$	$2^{k+1} \left[ \frac{\Gamma\left\{\frac{3(1+k)}{2}\right\}}{\Gamma\left(\frac{1+k}{2}\right)} \right]$	$\frac{\Gamma\left\{\frac{5}{2}(1+k)\right\}}{\Gamma^2\left\{\frac{3}{2}(1+k)\right\}}$
Kotz	$\frac{r^{(2N-1)/2}}{\Gamma\left(\frac{2N-1}{2}\right)} u^{N-1} e^{-ru}, \quad r > 0, N \geq 1$	$\frac{2N-1}{2r}$	$\frac{2N+1}{2N-1}$
Kotz generalizada	$\frac{sr^{(2N-1)/2s}}{\Gamma\left(\frac{2N-1}{2s}\right)} u^{N-1} e^{-ru^s}, \quad r, s > 0, N \geq 1$	$\frac{\Gamma\left(\frac{2N-1}{2s}\right)}{r^{1/s} \Gamma\left(\frac{2N-1}{2s}\right)}$	$\frac{\Gamma\left(\frac{2N-1}{2s}\right) \Gamma\left(\frac{2N+3}{2s}\right)}{\Gamma^2\left(\frac{2N+1}{2s}\right)}$
Normal contaminada	$(1-\epsilon) \frac{\exp\{-u/2\}}{\sqrt{2\pi}} + \epsilon \frac{\exp\{-u/(2\sigma^2)\}}{\sqrt{2\pi}\sigma}, \quad \sigma > 0, 0 \leq \epsilon \leq 1$	$1 + \epsilon(\sigma^2 - 1)$	$3 \frac{\{1 + \epsilon(\sigma^4 - 1)\}}{\{1 + \epsilon(\sigma^2 - 1)\}^2}$

$\Gamma(\cdot)$  é a função gama.

# Apêndice B

## Demais resultados da abordagem clássica

### B.1 Resultados dos testes de heterocedasticidade

Nesta seção apresentamos os resultados obtidos para os testes da razão de verossimilhanças e escore discutidos na Seção 3.2.1 para os modelos dados pelas relações hipsométricas de Stofel, Curtis, Petterson e Prodan com as distribuições normal, normal generalizada com  $s = 3$ ,  $t$  de Student com 30 g.l. e logística-II.

Tabela B.1: Valor das estatísticas dos testes e p-valor (entre parênteses) para os modelos ajustados com a distribuição normal.

Estatística	Modelos			
	Stofel	Curtis	Petterson	Prodan
$\xi_{RV}$	48,230 (0,000)	6,582 (0,021)	20,220 (0,000)	52,169 (0,000)
$\xi_{SR}$	28,687 (0,000)	5,438 (0,039)	13,667 (0,000)	30,217 (0,000)

Tabela B.2: Valor das estatísticas dos testes e p-valor (entre parênteses) para os modelos ajustados com a distribuição normal generalizada com  $s = 3$ .

Modelos				
Estatística	Stofel	Curtis	Petterson	Prodan
$\xi_{RV}$	65,463	8,567	26,087	70,340
	(0,000)	(0,007)	(0,000)	(0,000)
$\xi_{SR}$	24,006	5,655	12,353	24,804
	(0,000)	(0,035)	(0,001)	(0,000)

Tabela B.3: Valor das estatísticas dos testes e p-valor (entre parênteses) para os modelos ajustados com a distribuição  $t$  de Student com 30 g.l.

Modelos				
Estatística	Stofel	Curtis	Petterson	Prodan
$\xi_{RV}$	46,007	3,296*	15,079	49,777
	(0,000)	(0,139)	(0,000)	(0,000)
$\xi_{SR}$	46,053	8,405	22,009	47,320
	(0,000)	(0,008)	(0,000)	(0,000)

\* valor não significativo ao nível 5%.

Tabela B.4: Valor das estatísticas dos testes e p-valor (entre parênteses) para os modelos hipsométricos ajustados com a distribuição logística-II.

Modelos				
Estatística	Stofel	Curtis	Petterson	Prodan
$\xi_{RV}$	39,030	5,786	17,755	42,783
	(0,000)	(0,032)	(0,000)	(0,000)
$\xi_{SR}$	34,195	5,672	15,448	35,284
	(0,000)	(0,035)	(0,000)	(0,000)

## B.2 Estimativas de máxima verossimilhança e gráficos dos resíduos via GAMLSS

- Distribuição normal

Tabela B.5: Estimativas de máxima verossimilhança para o modelo de Stofel ajustado com a distribuição normal.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	1,334	0,638	5,999	-0,022
Erro-padrão	0,031	0,011	0,284	0,003

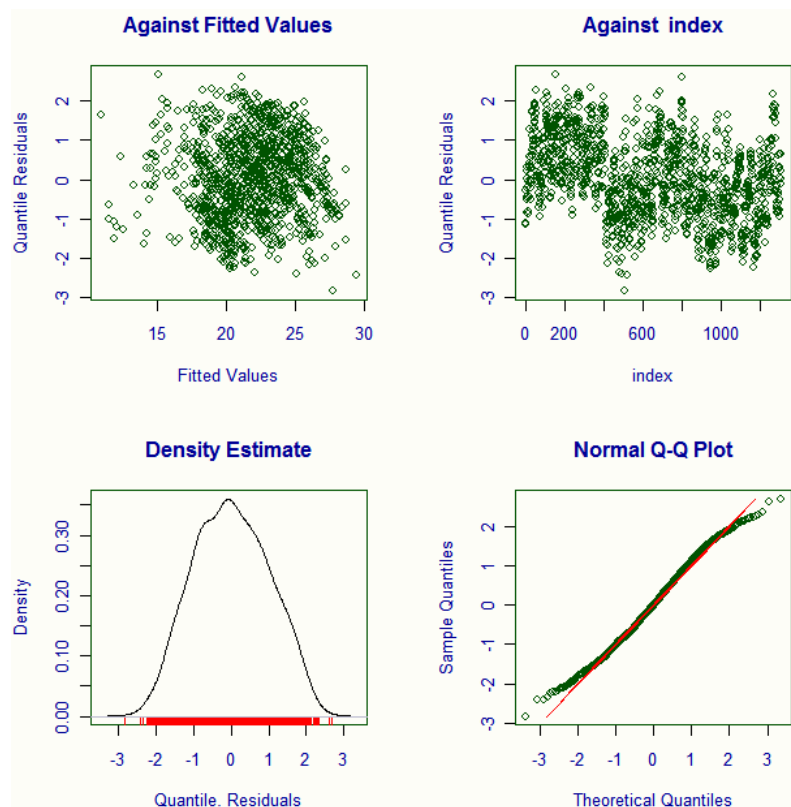


Figura B.1: Quantis residuais do modelo de Stofel ajustado com a distribuição normal.



Tabela B.6: Estimativas de máxima verossimilhança para o modelo de Curtis ajustado com a distribuição normal.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	3,705	9,403	5,614	-0,008
Erro-padrão	0,017	0,273	0,268	0,003

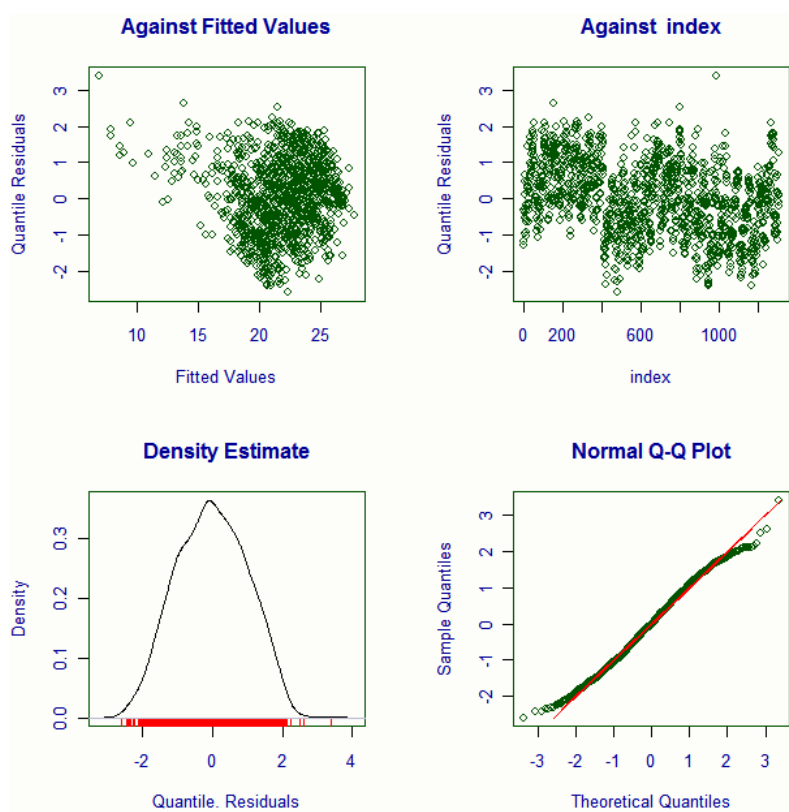


Figura B.2: Quantis residuais do modelo de Curtis ajustado com a distribuição normal.

Tabela B.7: Estimativas de máxima verossimilhança para o modelo de Petterson ajustado com a distribuição normal.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	0,284	1,124	5,799	-0,014
Erro-padrão	0,002	0,030	0,295	0,003

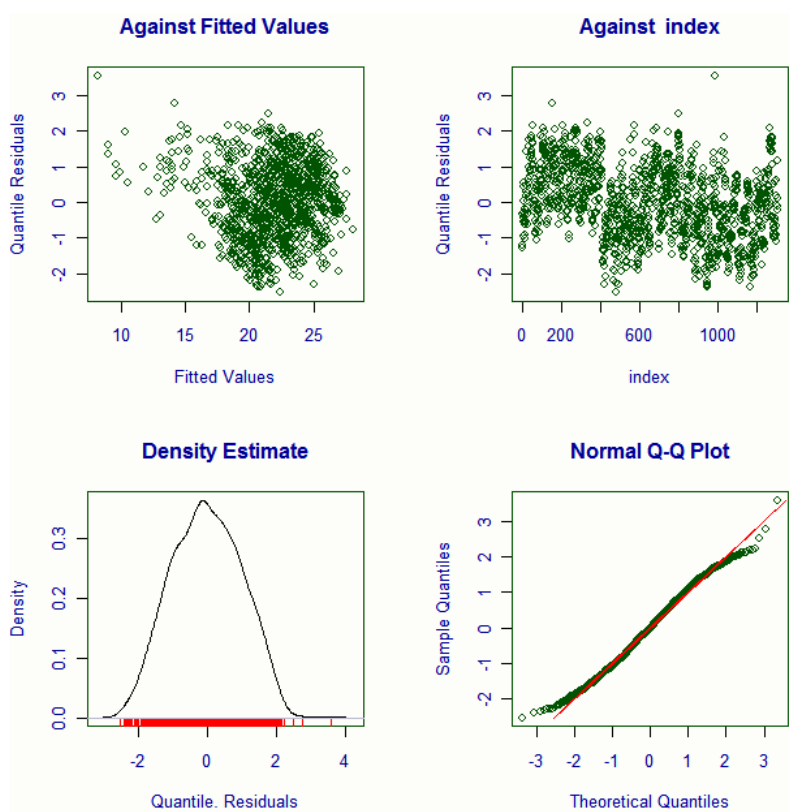


Figura B.3: Quantis residuais do modelo de Petterson ajustado com a distribuição normal.

Tabela B.8: Estimativas de máxima verossimilhança para o modelo de Prodan ajustado com a distribuição normal.

	Parâmetros				
	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\delta_2$
Estimativa	-0,636	0,535	0,014	6,039	-0,023
Erro-padrão	0,266	0,044	0,002	0,369	0,003

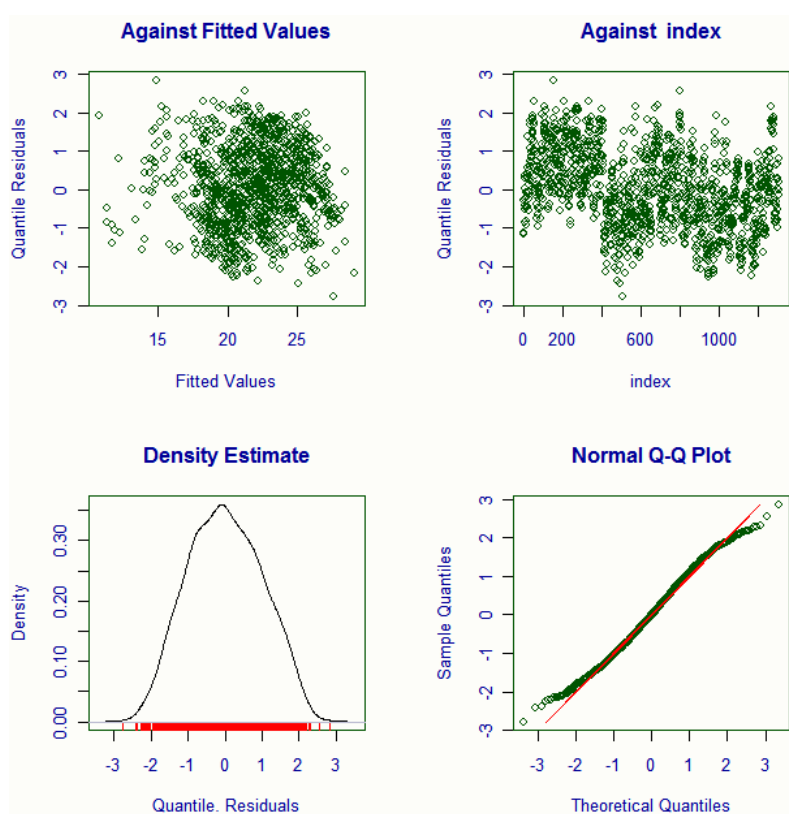


Figura B.4: Quantis residuais do modelo de Prodan ajustado com a distribuição normal.

- Distribuição normal generalizada com parâmetro de forma  $s = 3$

Tabela B.9: Estimativas de máxima verossimilhança para o modelo de Stofel ajustado com a distribuição normal generalizada ( $s = 3$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	1,360	0,629	15,943	-0,022
Erro-padrão	0,028	0,010	0,613	0,002

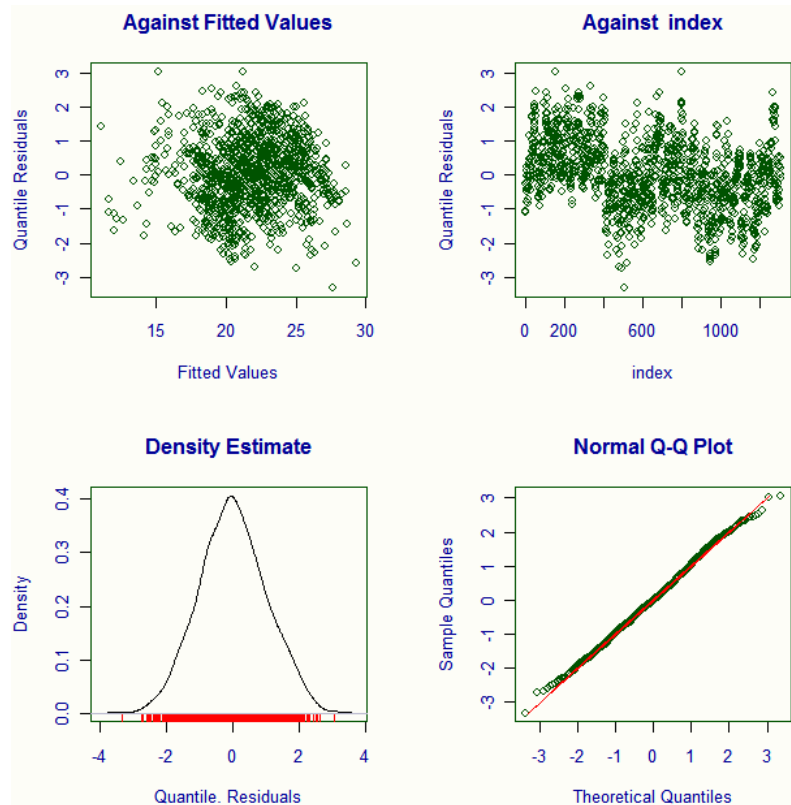


Figura B.5: Quantis residuais do modelo de Stofel ajustado com a distribuição normal generalizada ( $s = 3$ ).

Tabela B.10: Estimativas de máxima verossimilhança para o modelo de Curtis ajustado com a distribuição normal generalizada ( $s = 3$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	3,702	9,370	14,928	-0,006
Erro-padrão	0,018	0,275	0,571	0,002

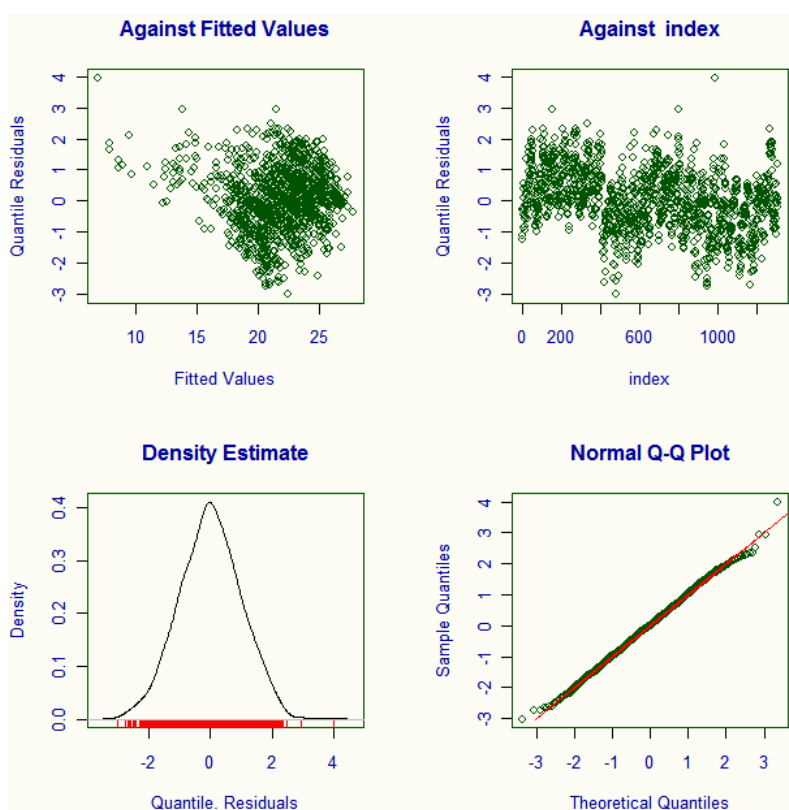


Figura B.6: Quantis residuais do modelo de Curtis ajustado com a distribuição normal generalizada ( $s = 3$ ).

Tabela B.11: Estimativas de máxima verossimilhança para o modelo de Petterson ajustado com a distribuição normal generalizada ( $s = 3$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	0,285	1,106	15,318	-0,013
Erro-padrão	0,002	0,029	0,637	0,002

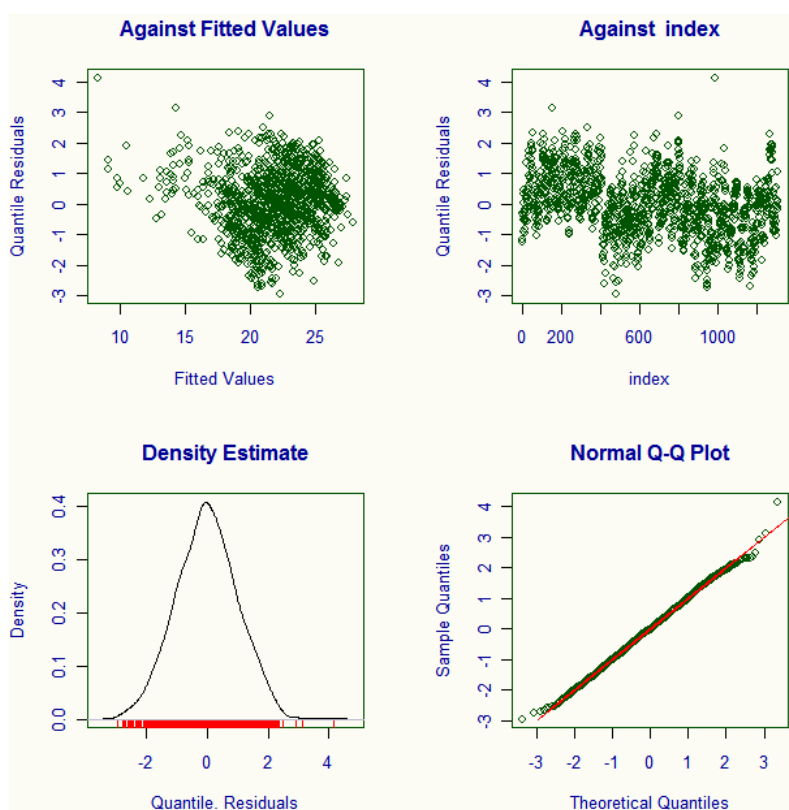


Figura B.7: Quantis residuais do modelo de Petterson ajustado com a distribuição normal generalizada ( $s = 3$ ).

- Distribuição  $t$  de Student com 30 graus de liberdade

Tabela B.12: Estimativas de máxima verossimilhança para o modelo de Stofel ajustado com a distribuição  $t$  de Student (30 g.l.).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	1,329	0,639	5,749	-0,022
Erro-padrão	0,032	0,011	0,283	0,003

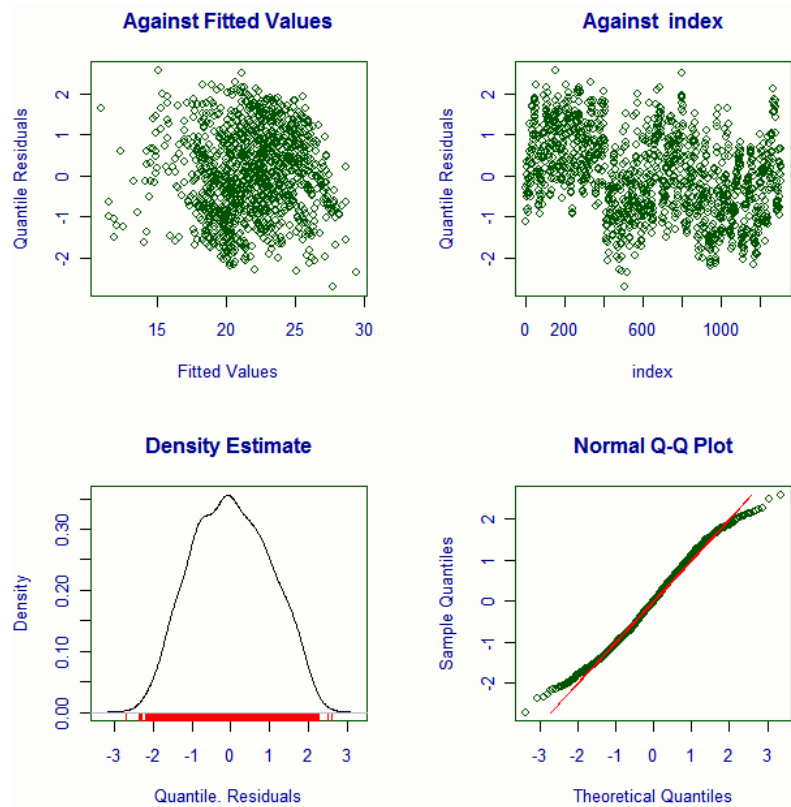


Figura B.8: Quantis residuais do modelo de Stofel ajustado com a distribuição  $t$  de Student (30 g.l.).

Tabela B.13: Estimativas de máxima verossimilhança para o modelo de Curtis ajustado com a distribuição  $t$  de Student (30 g.l.).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	3,730	9,819	4,985	-0,004*
Erro-padrão	0,018	0,285	0,236	0,003

\* valor não significativo ao nível 5%.

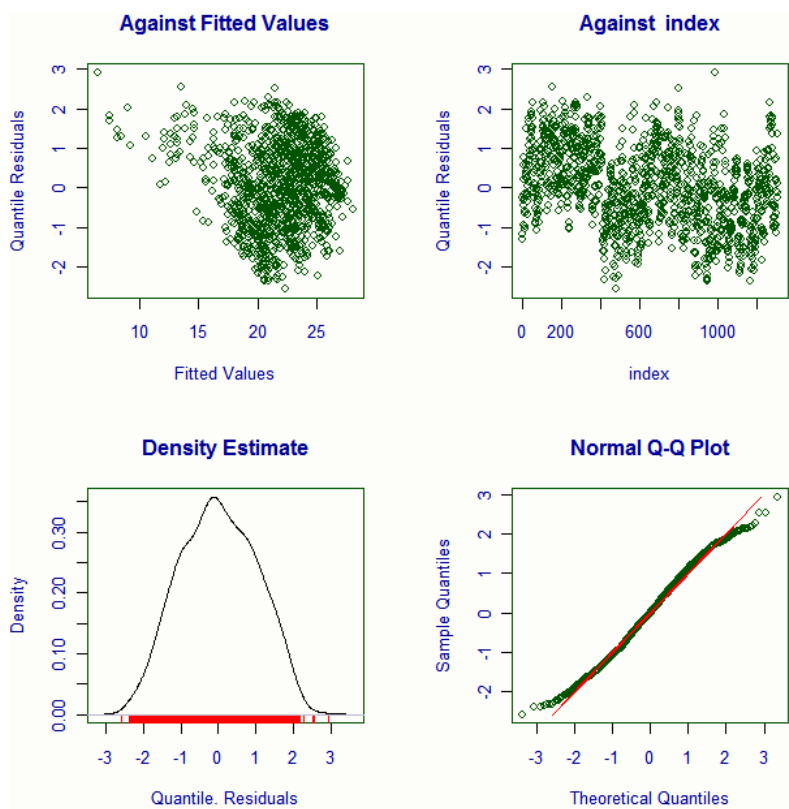


Figura B.9: Quantis residuais do modelo de Curtis ajustado com a distribuição  $t$  de Student (30 g.l.).



Tabela B.14: Estimativas de máxima verossimilhança para o modelo de Petterson ajustado com a distribuição  $t$  de Student (30 g.l.).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	0,282	1,148	4,991	-0,011
Erro-padrão	0,002	0,030	0,243	0,003

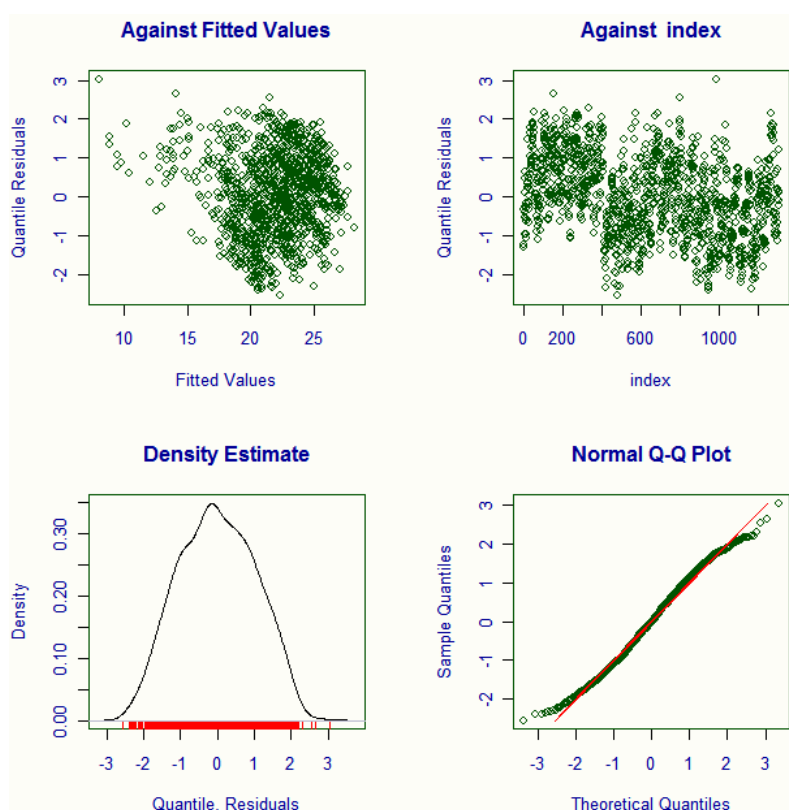


Figura B.10: Quantis residuais do modelo de Petterson ajustado com a distribuição  $t$  de Student (30 g.l.).

Tabela B.15: Estimativas de máxima verossimilhança para o modelo de Prodan ajustado com a distribuição  $t$  de Student (30 g.l.).

	Parâmetros				
	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\delta_2$
Estimativa	-0,576*	0,526	0,014	5,718	-0,023
Erro-padrão	0,281	0,046	0,002	0,354	0,003

\* valor não significativo ao nível 5%.

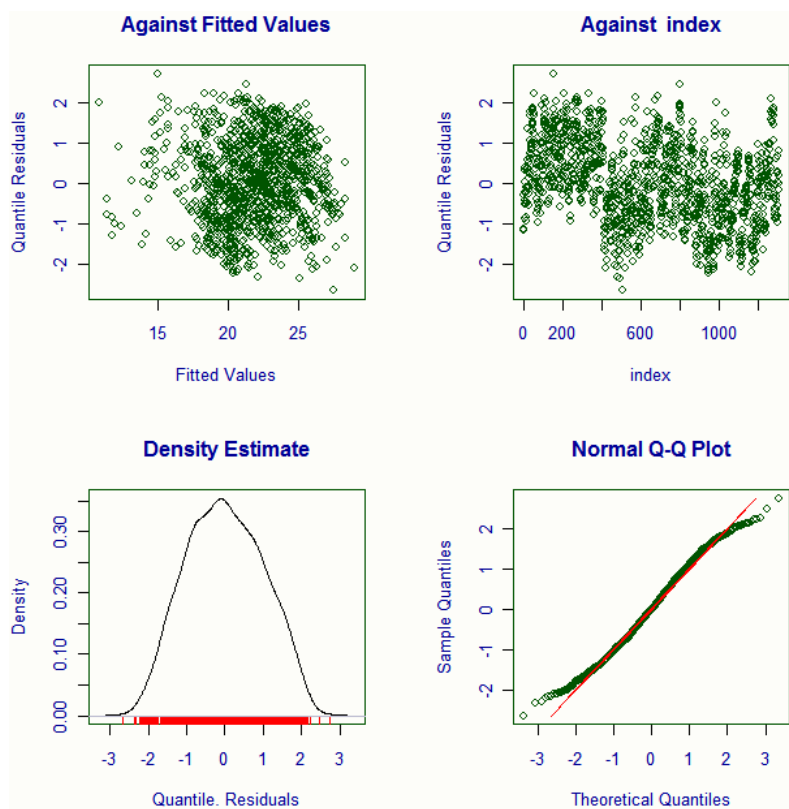


Figura B.11: Quantis residuais do modelo de Prodan ajustado com a distribuição  $t$  de Student (30 g.l.).

- Distribuição logística-II

Tabela B.16: Estimativas de máxima verossimilhança para o modelo de Stofel ajustado com a distribuição logística-II.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	1,318	0,643	2,055	-0,022
Erro-padrão	0,033	0,012	0,113	0,003

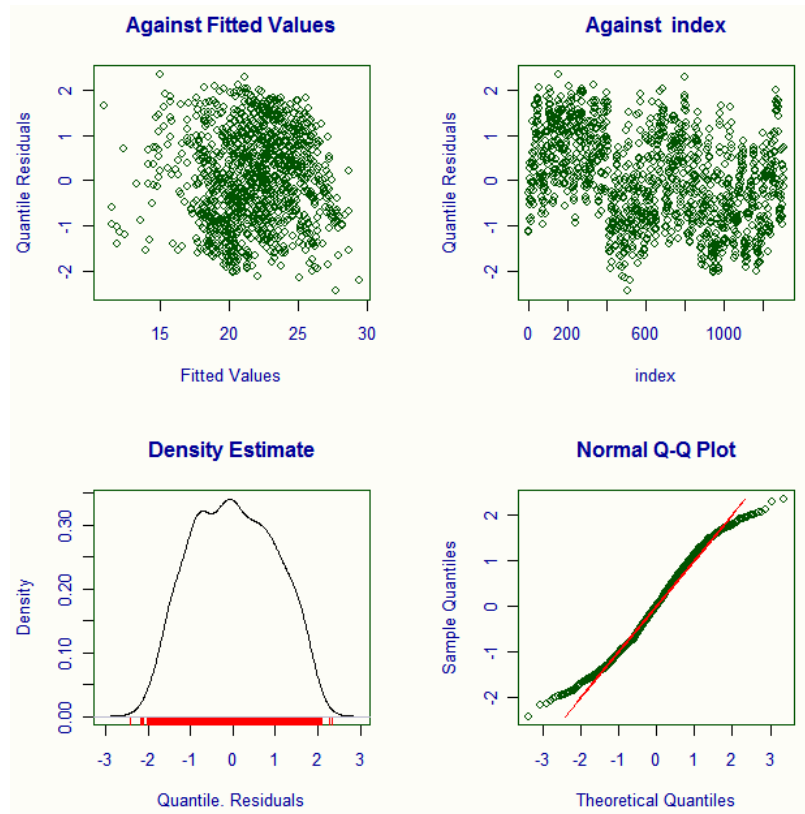


Figura B.12: Quantis residuais do modelo de Stofel ajustado com a distribuição logística-II.

Tabela B.17: Estimativas de máxima verossimilhança para o modelo de Curtis ajustado com a distribuição logística-II.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	3,711	9,502	1,960	-0,009
Erro-padrão	0,017	0,282	0,112	0,004

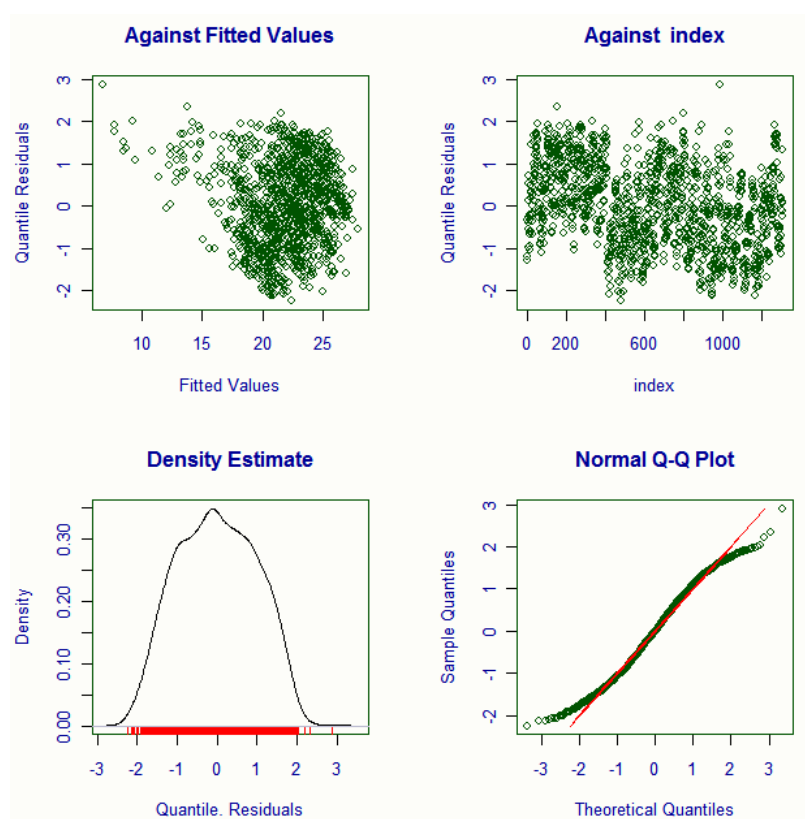


Figura B.13: Quantis residuais do modelo de Curtis ajustado com a distribuição logística-II.

Tabela B.18: Estimativas de máxima verossimilhança para o modelo de Petterson ajustado com a distribuição logística-II.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Estimativa	0,283	1,129	2,042	-0,016
Erro-padrão	0,002	0,031	0,123	0,004

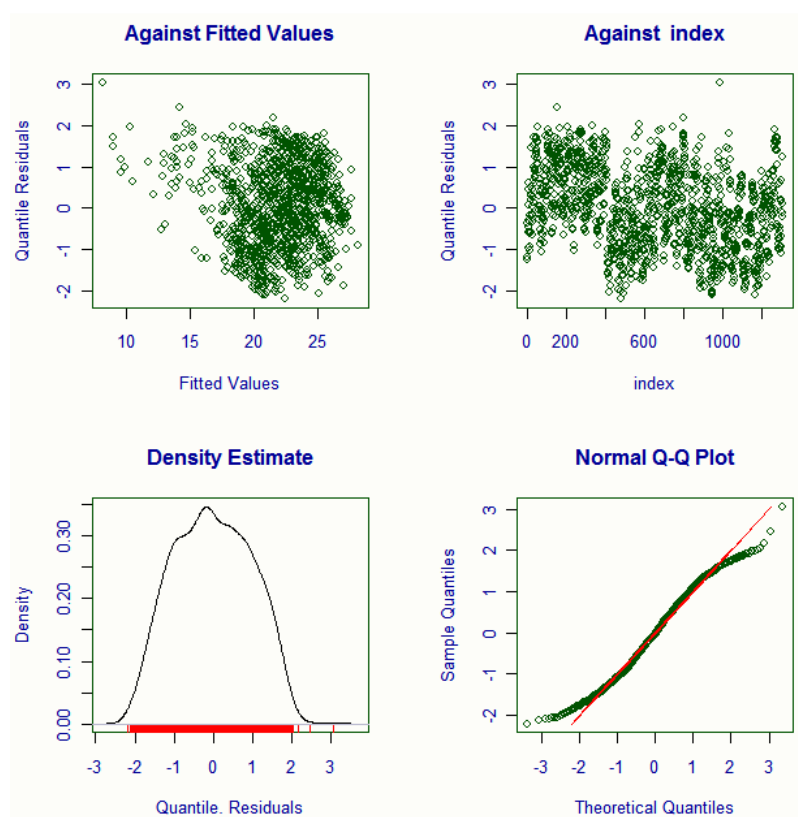


Figura B.14: Quantis residuais do modelo de Petterson ajustado com a distribuição logística-II.

Tabela B.19: Estimativas de máxima verossimilhança para o modelo de Prodan ajustado com a distribuição logística-II.

	Parâmetros				
	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\delta_2$
Estimativa	-0,585*	0,532	0,014	2,101	-0,024
Erro-padrão	0,296	0,047	0,002	0,143	0,003

\* valor não significativo ao nível 5%.

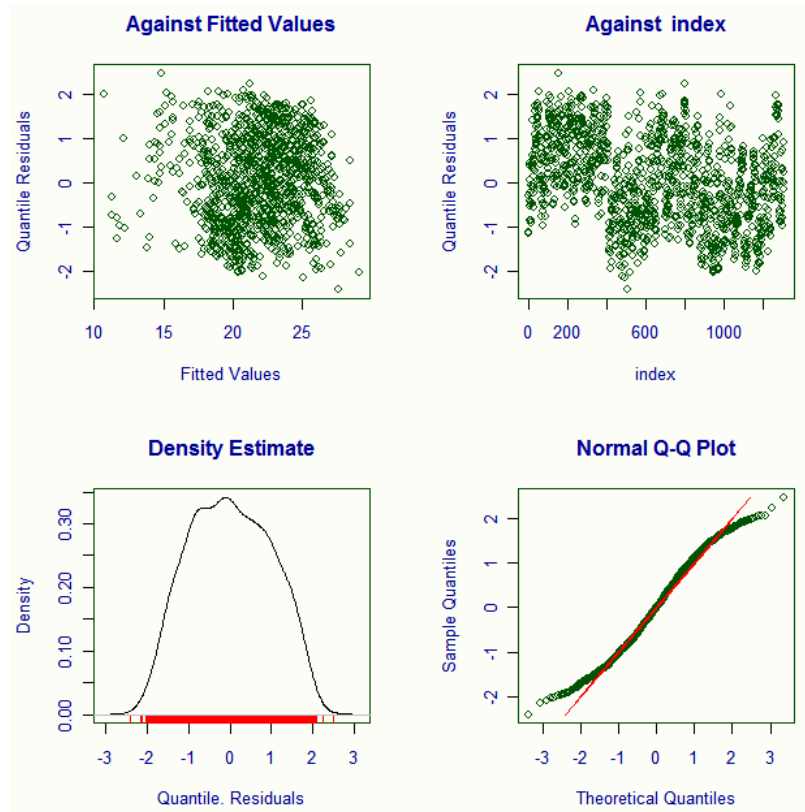


Figura B.15: Quantis residuais do modelo de Prodan ajustado com a distribuição logística-II.

# Apêndice C

## Demais resultados da abordagem bayesiana

### C.1 Estimativas bayesianas

- Distribuição normal

Tabela C.1: Estimativas *a posteriori* para o modelo de Stofel ajustado com a distribuição normal.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	1,334	0,637	6,017	-0,022
Desvio padrão	0.031	0,011	0,300	0,003

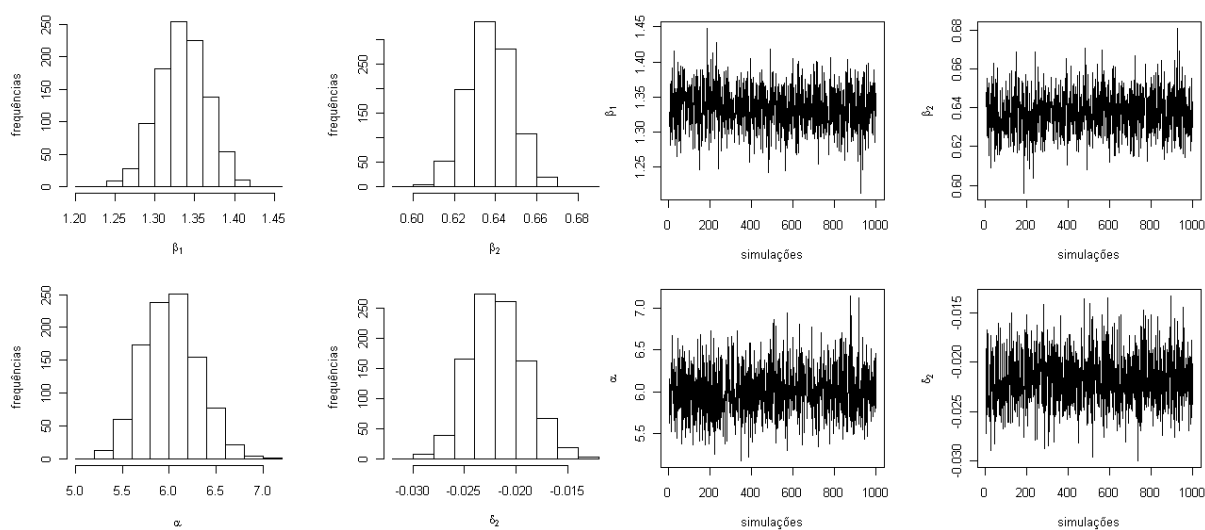


Figura C.1: Histograma e Simulação dos parâmetros para o modelo de Stofel ajustado com a distribuição normal.

Tabela C.2: Estimativas *a posteriori* para o modelo de Curtis ajustado com a distribuição normal.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	3,705	9,402	5,642	-0,008
Desvio padrão	0,017	0,276	0,276	0,003



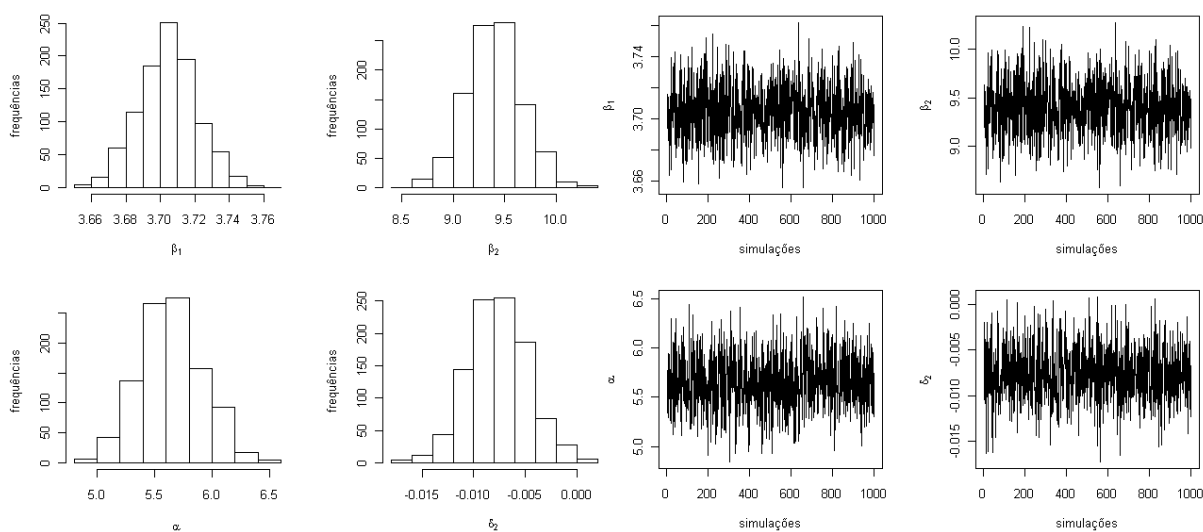


Figura C.2: Histograma e Simulação dos parâmetros para o modelo de Curtis ajustado com a distribuição normal.

Tabela C.3: Estimativas *a posteriori* para o modelo de Petterson ajustado com a distribuição normal.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	0,284	1,124	5,805	-0,014
Desvio padrão	0,002	0,029	0,292	0,003

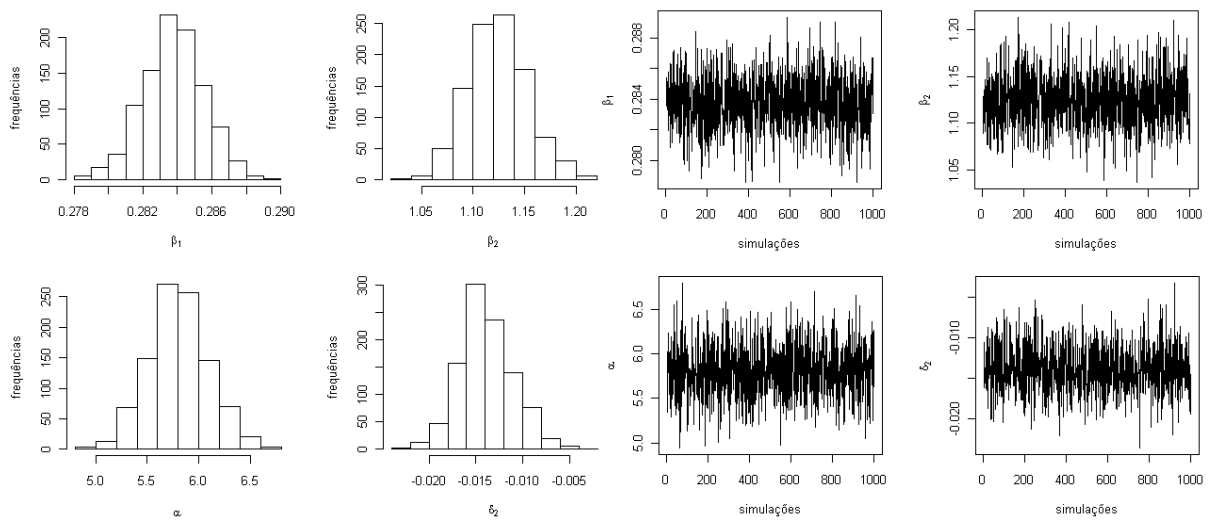


Figura C.3: Histograma e Simulação dos parâmetros para o modelo de Petterson ajustado com a distribuição normal.

Tabela C.4: Estimativas *a posteriori* para o modelo de Prodan ajustado com a distribuição normal.

Parâmetros					
	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\delta_2$
Média	-0,463	0,508	0,015	6,070	-0,023
Desvio padrão	0,173	0,026	0,001	0,285	0,003

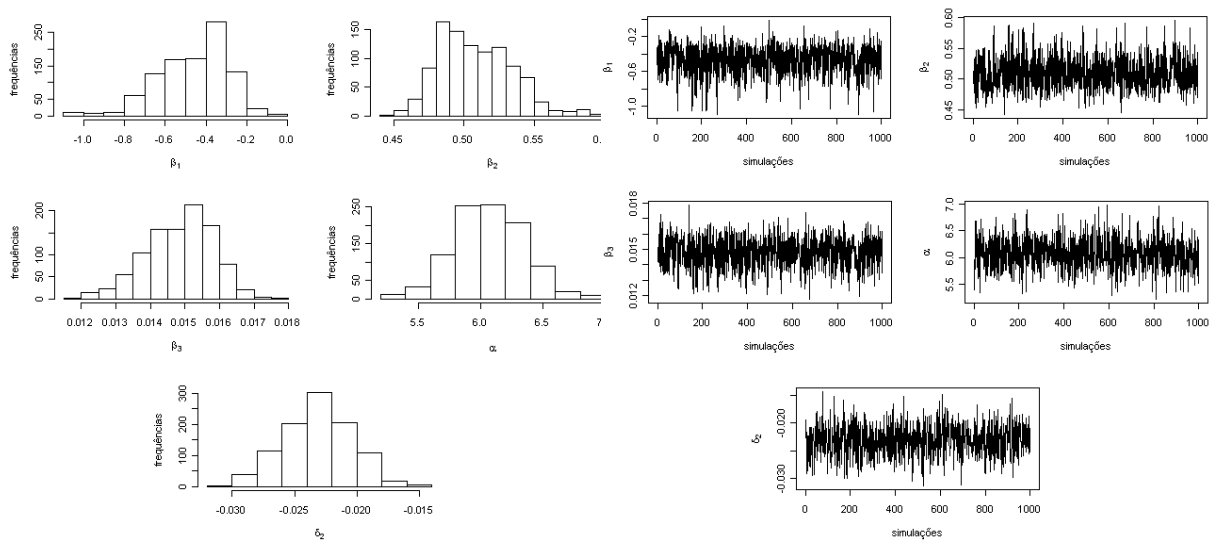


Figura C.4: Histograma e Simulação dos parâmetros para o modelo de Prodan ajustado com a distribuição normal.

- Distribuição normal generalizada com parâmetro de forma  $s = 3$

Tabela C.5: Estimativas *a posteriori* para o modelo de Stofel ajustado com a distribuição normal generalizada ( $s = 3$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	1,357	0,629	15,963	-0,021
Desvio padrão	0,027	0,010	0,624	0,002

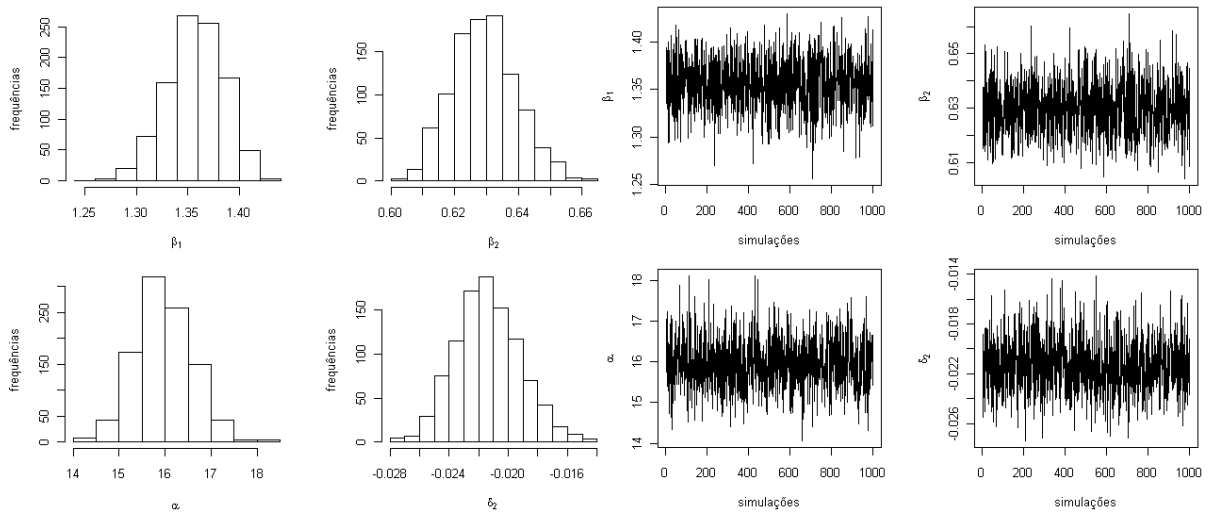


Figura C.5: Histograma e Simulação dos parâmetros para o modelo de Stofel ajustado com a distribuição normal generalizada ( $s = 3$ ).

Tabela C.6: Estimativas *a posteriori* para o modelo de Curtis ajustado com a distribuição normal generalizada ( $s = 3$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	3,690	9,186	14,939	-0,007
Desvio padrão	0,017	0,270	0,589	0,002

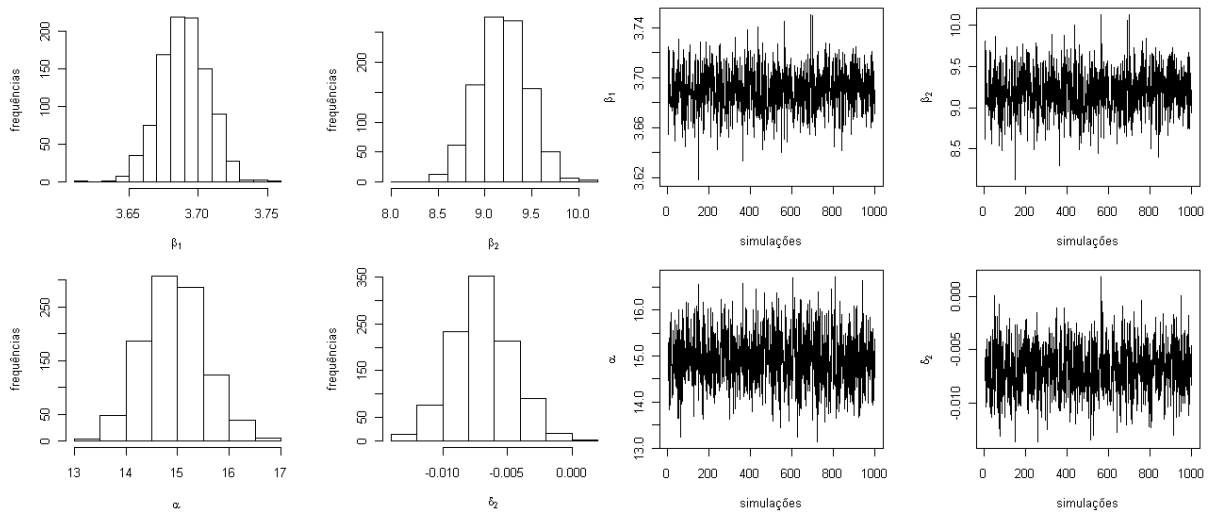


Figura C.6: Histograma e Simulação dos parâmetros para o modelo de Curtis ajustado com a distribuição normal generalizada ( $s = 3$ ).

Tabela C.7: Estimativas *a posteriori* para o modelo de Petterson ajustado com a distribuição normal generalizada ( $s = 3$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	0,285	1,107	15,334	-0,013
Desvio padrão	0,002	0,029	0,583	0,002

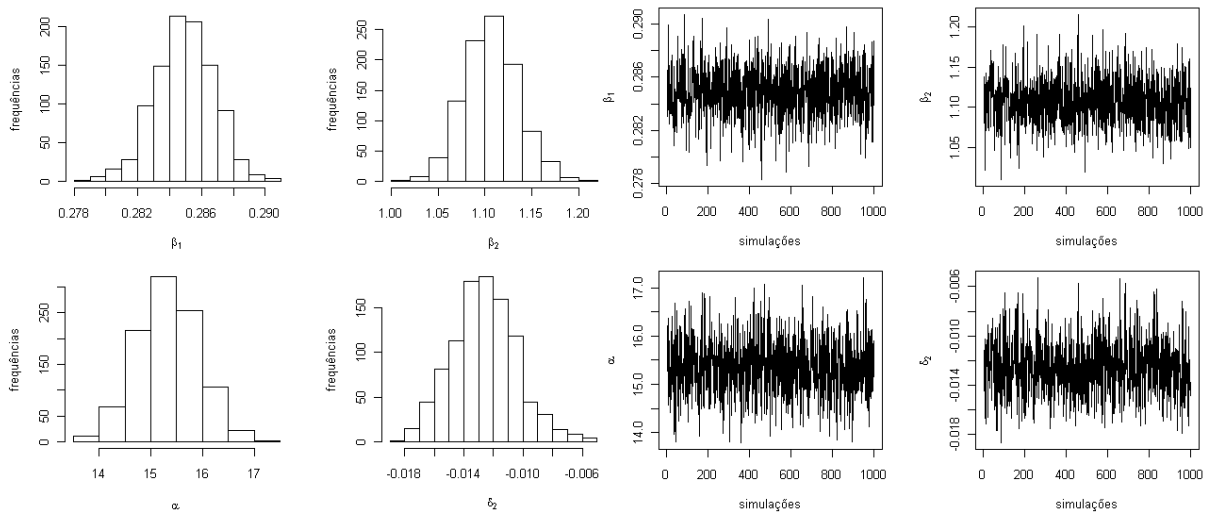


Figura C.7: Histograma e Simulação dos parâmetros para o modelo de Petterson ajustado com a distribuição normal generalizada ( $s = 3$ ).

- Distribuição  $t$  de Student com 30 graus de liberdade

Tabela C.8: Estimativas *a posteriori* para o modelo de Stofel ajustado com a distribuição  $t$  de Student ( $\nu = 30$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	1,326	0,640	5,756	-0,022
Desvio padrão	0,033	0,012	0,288	0,003

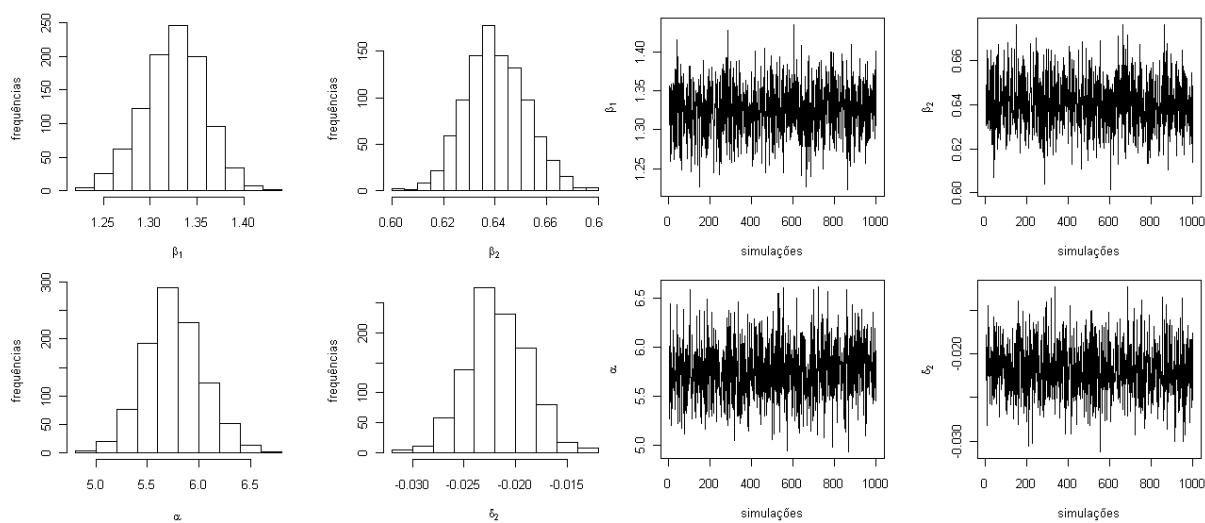


Figura C.8: Histograma e Simulação dos parâmetros para o modelo de Stofel ajustado com a distribuição  $t$  de Student ( $\nu = 30$ ).

Tabela C.9: Estimativas *a posteriori* para o modelo de Curtis ajustado com a distribuição  $t$  de Student ( $\nu = 30$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	3,705	9,412	5,429	-0,008
Desvio padrão	0,017	0,272	0,273	0,003

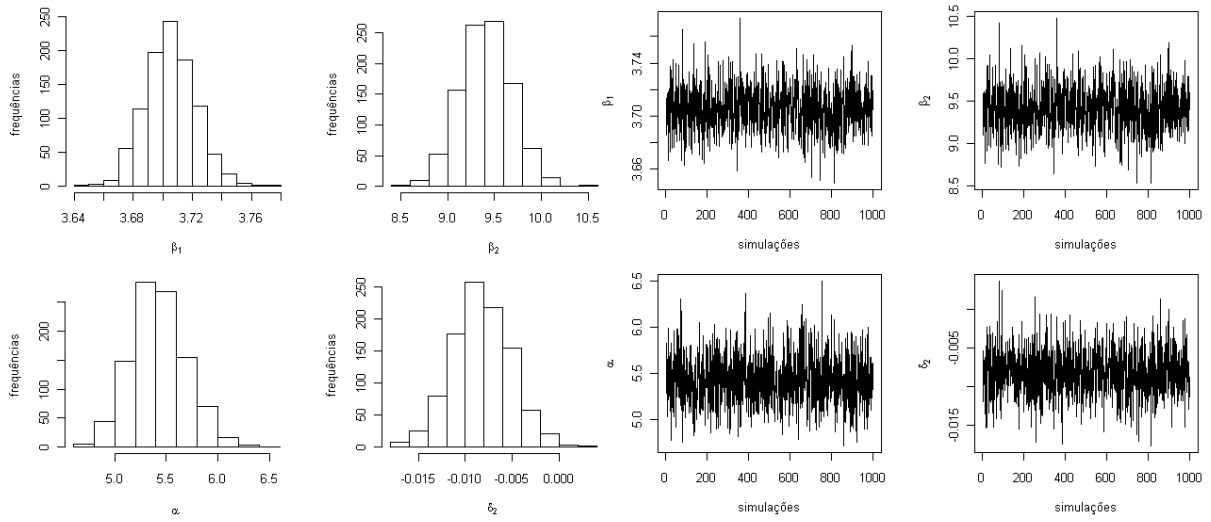


Figura C.9: Histograma e Simulação dos parâmetros para o modelo de Curtis ajustado com a distribuição  $t$  de Student ( $\nu = 30$ ).

Tabela C.10: Estimativas *a posteriori* para o modelo de Petterson ajustado com a distribuição  $t$  de Student ( $\nu = 30$ ).

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	0,284	1,127	5,577	-0,015
Desvio padrão	0,002	0,028	0,283	0,003



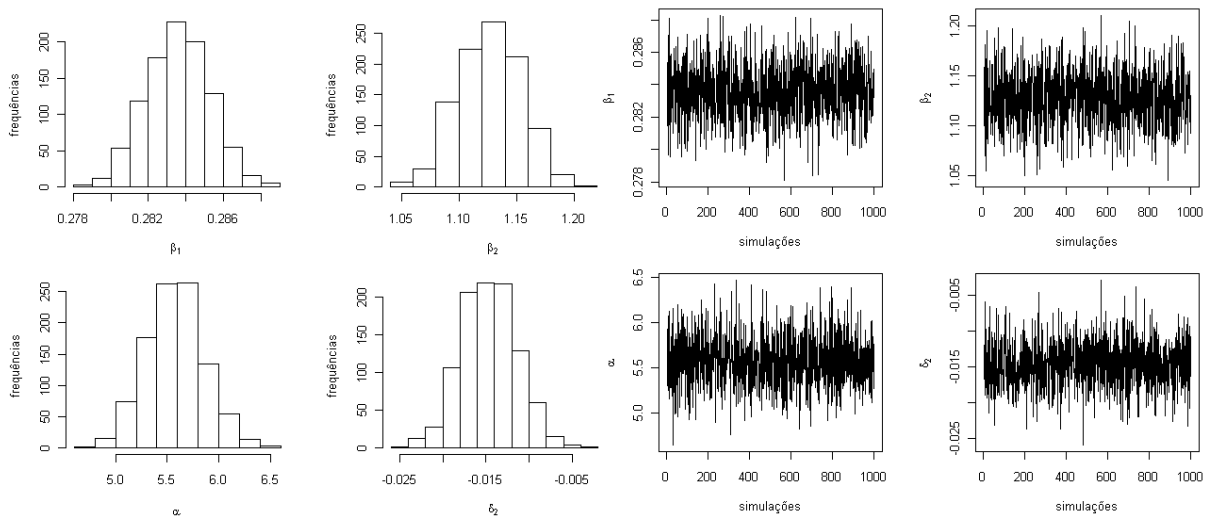


Figura C.10: Histograma e Simulação dos parâmetros para o modelo de Petterson ajustado com a distribuição  $t$  de Student ( $\nu = 30$ ).

Tabela C.11: Estimativas *a posteriori* para o modelo de Prodan ajustado com a distribuição  $t$  de Student ( $\nu = 30$ ).

Parâmetros					
	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\delta_2$
Média	-0,494	0,515	0,015	5,834	-0,023
Desvio padrão	0,235	0,035	0,001	0,301	0,003

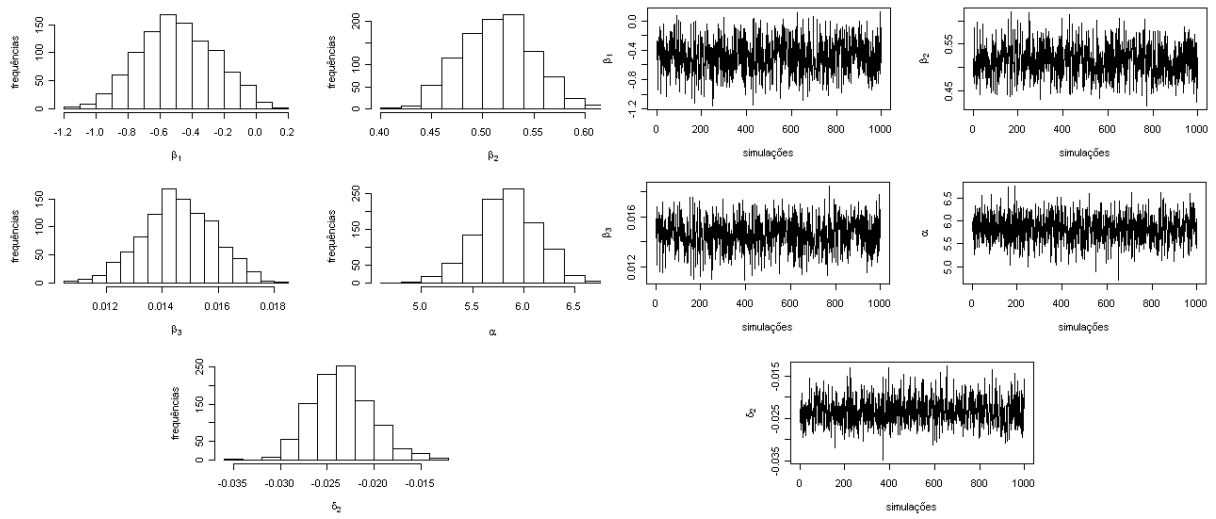


Figura C.11: Histograma e Simulação dos parâmetros para o modelo de Prodan ajustado com a distribuição  $t$  de Student ( $\nu = 30$ ).

- Distribuição logística-II

Tabela C.12: Estimativas *a posteriori* para o modelo de Stofel ajustado com a distribuição logística-II.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	1,313	0,645	2,081	-0,022
Desvio padrão	0,032	0,011	0,112	0,003

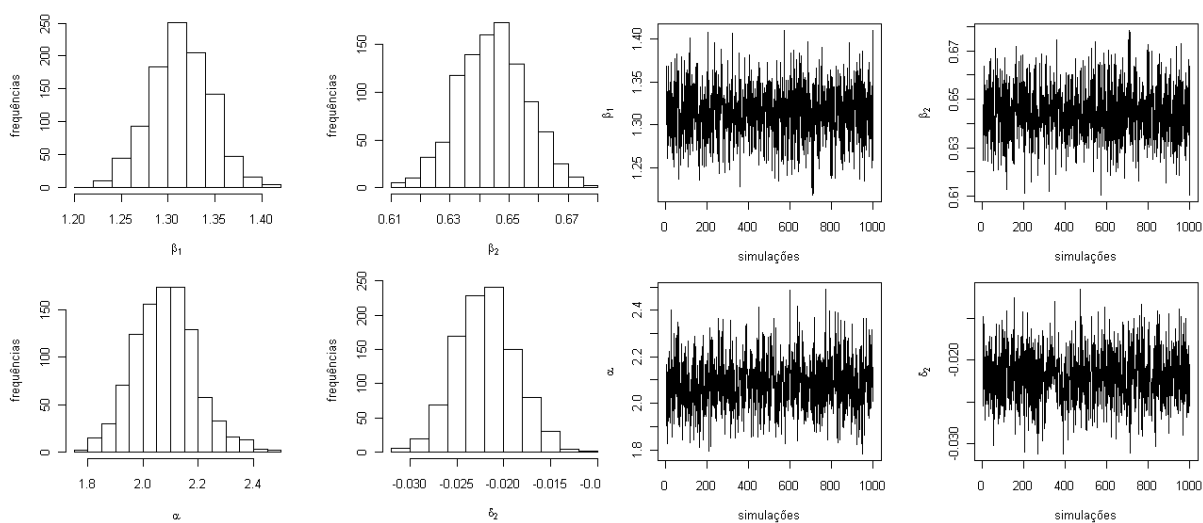


Figura C.12: Histograma e Simulação dos parâmetros para o modelo de Stofel ajustado com a distribuição logística-II.

Tabela C.13: Estimativas *a posteriori* para o modelo de Curtis ajustado com a distribuição logística-II.

Parâmetros				
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	3,712	9,516	1,968	-0,009
Desvio padrão	0,017	0,274	0,120	0,004

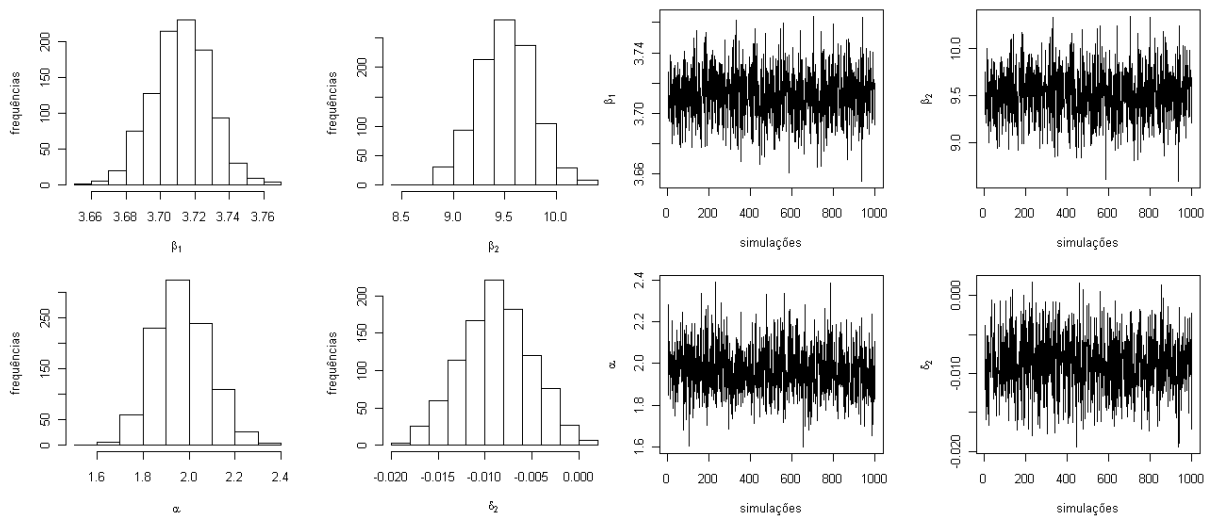


Figura C.13: Histograma e Simulação dos parâmetros para o modelo de Curtis ajustado com a distribuição logística-II.

Tabela C.14: Estimativas *a posteriori* para o modelo de Petterson ajustado com a distribuição logística-II.

	Parâmetros			
	$\beta_1$	$\beta_2$	$\alpha$	$\delta_2$
Média	0,283	1,130	2,054	-0,016
Desvio padrão	0,002	0,030	0,115	0,004

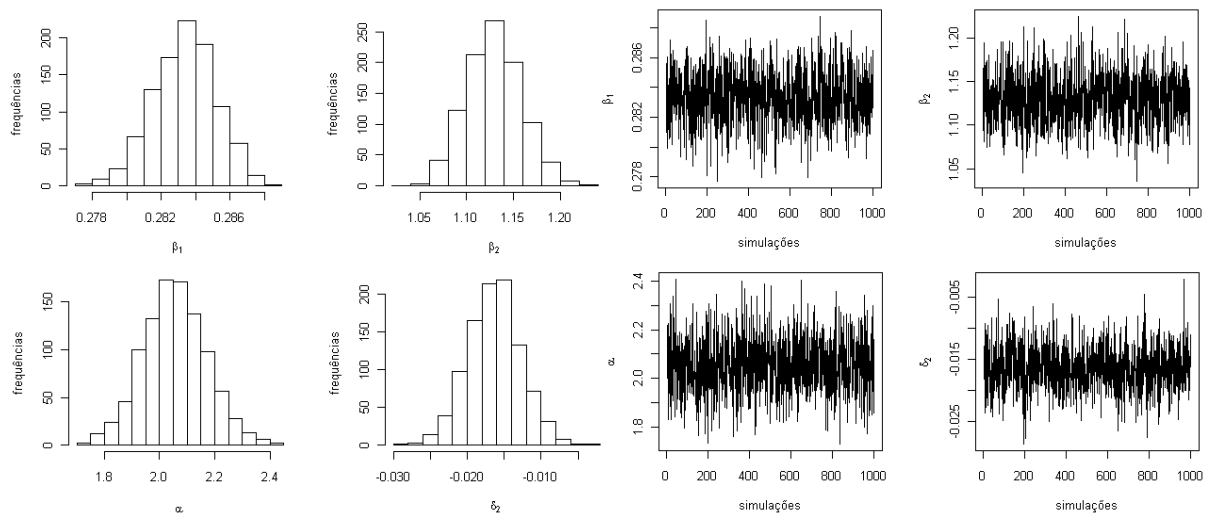


Figura C.14: Histograma e Simulação dos parâmetros para o modelo de Petterson ajustado com a distribuição logística-II.

Tabela C.15: Estimativas *a posteriori* para o modelo de Prodan ajustado com a distribuição logística-II.

	Parâmetros				
	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha$	$\delta_2$
Média	-0,447	0,510	0,015	2,140	-0,025
Desvio padrão	0,218	0,035	0,001	0,113	0,003

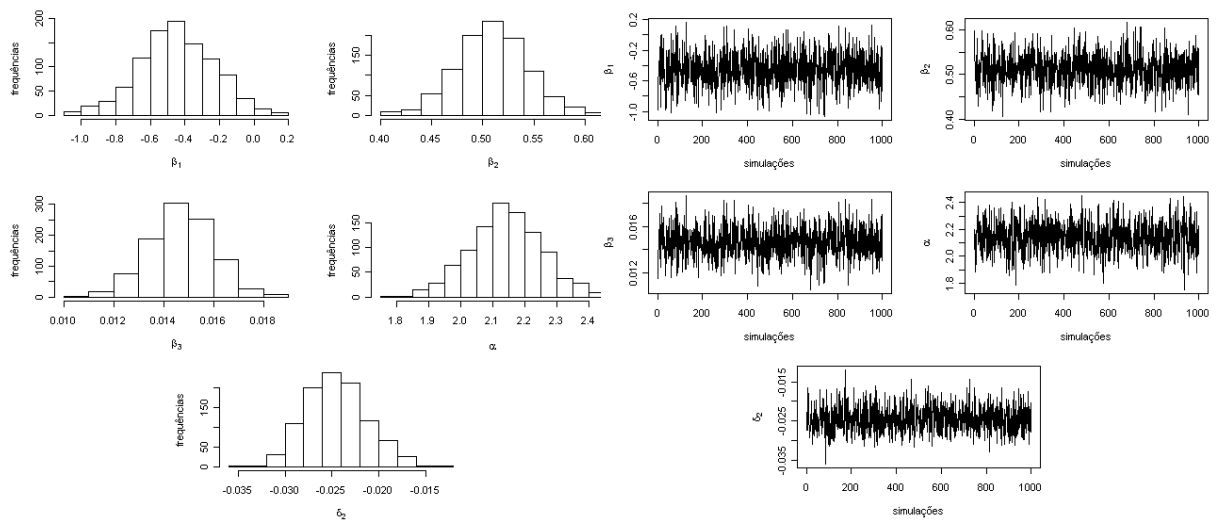


Figura C.15: Histograma e Simulação dos parâmetros para o modelo de Prodan ajustado com a distribuição logística-II.