

---

Modelos não lineares assimétricos com efeitos mistos

*Marcos Antonio Alves Pereira*

---



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

MARCOS ANTONIO ALVES PEREIRA

**MODELOS NÃO LINEARES ASSIMÉTRICOS COM EFEITOS MISTOS**

Tese apresentada ao Departamento de Estatística – DEs/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientadora: Profa. Dra. Cibele Maria Russo Novelli

**São Carlos**  
**Setembro de 2019**



UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA  
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

MARCOS ANTONIO ALVES PEREIRA

**NONLINEAR MIXED-EFFECTS MODELS WITH SCALE MIXTURE OF SKEW-NORMAL  
DISTRIBUTIONS**

Doctoral dissertation submitted to the Department of Statistics – DEs-UFSCar and to the Institute of Mathematics and Computer Sciences – ICMC-USP, in partial fulfillment of the requirements for the PhD degree Statistics – Interagency Program Graduate in Statistics UFSCar-USP.

Advisor: Profa. Dra. Cibele Maria Russo Novelli

**São Carlos  
September 2019**





# UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia  
Programa Interinstitucional de Pós-Graduação em Estatística

---

## Folha de Aprovação

---

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Tese de Doutorado do candidato Marcos Antonio Alves Pereira, realizada em 02/08/2019:

---

Profa. Dra. Cibeles Maria Russo Noveli  
USP

---

Profa. Dra. Viviana Giampaoli  
IME-USP

---

Profa. Dra. Larissa Avila Matos  
UNICAMP

---

Prof. Dr. Cristian Marcelo Villegas Lobos  
ESALQ/USP

---

Prof. Dr. Marinho Gomes de Andrade Filho  
USP

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Viviana Giampaoli e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ão) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

---

Profa. Dra. Cibeles Maria Russo Noveli



*Este trabalho é dedicado às minhas filhas Mônica Sofia e Manuella Christina,  
e aos meus familiares e amigos que sempre torceram por mim.*



# AGRADECIMENTOS

---

---

Agradeço primeiramente a Deus por todas as coisas boas que aconteceram, acontecem e virão a acontecer na minha vida, em especial, na minha vida acadêmica.

Gostaria de agradecer à Universidade Federal do Piauí (UFPI), por ter permitido meu afastamento, de forma integral, das atividades de docente no *campus* de Floriano, à minha orientadora professora Cibele Russo pela dedicação e confiança, e ao professor Gustavo Pereira por ter dedicado seu tempo para esclarecer diversas dúvidas. Agradecer aos professores do mestrado Betsabé Blás e Francisco Cribari, e do doutorado Mário de Castro e José Galvão Leite pelos bons ensinamentos que transmitiram.

Quero agradecer também aos meus amigos Karina Emboaba, Rogério Santana, Juan Pablo Bustamante, Hélio Azevedo e Marcelo Andrade por me proporcionarem muitas conversas produtivas e que certamente influenciaram positivamente no desenvolvimento deste trabalho, e aos demais amigos que de alguma forma contribuíram com esta pesquisa.

Agradeço também aos professores Cristian Villegas, Viviana Giampaoli, Larissa Ávila e Marinho Gomes, integrantes da comissão examinadora, pelas relevantes contribuições ao trabalho, além de agradecer aos funcionários do ICMC/USP Monique da Conceição, do DEs-UFSCar Isabel Araújo e Cellius Martinez, pelo apoio em sanar dúvidas sobre assuntos administrativos.

Envio um agradecimento especial ao professor Walter Spinelli por ter me ajudado durante a graduação em Matemática, sempre me inspirando e me mostrando as maravilhas que esta ciência proporcionou e proporciona à humanidade.

Meus sinceros agradecimentos.

***Marcos Antonio Alves Pereira***



*“Nossas virtudes e nossos defeitos são inseparáveis, assim como a força e a matéria.  
Quando se separam, o homem deixa de existir.”  
(Nikola Tesla)*



# RESUMO

PEREIRA, M. A. A. **Modelos não lineares assimétricos com efeitos mistos**. 2019. 150 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Este trabalho tem como objetivo desenvolver modelos de regressão não lineares assimétricos com efeitos mistos, que proporcionam alternativas ao uso da distribuição normal e outras distribuições simétricas, para evitar a sensibilidade nas estimativas a observações atípicas e assimetria. Modelos não lineares com efeitos mistos são explorados em várias áreas do conhecimento, principalmente quando os dados estão correlacionados, como dados longitudinais, medidas repetidas e dados multiníveis, em particular, por sua flexibilidade em lidar com medidas de áreas como biologia e farmacocinética. No entanto, existem dificuldades em obter estimadores explícitos para os parâmetros nesses modelos. Atualmente muitos estudos estão sendo desenvolvidos com a família misturas de escala da distribuição skew-normal (SMSN) que abrange distribuições com caudas leves e pesadas, como a skew-normal, skew-Student-t, skew-normal contaminada e skew-slash, como também as versões simétricas destas distribuições. Neste trabalho são apresentados modelos de regressão não lineares com efeitos mistos em que as componentes aleatórias têm distribuições pertencentes a família SMSN. Para estimação dos parâmetros é utilizada uma solução numérica via algoritmo EM e suas extensões, e algoritmo de Newton-Raphson. Análises em conjuntos de dados reais são realizadas com essa nova proposta, como o estudo da cinética de drogas em seres humanos, além de análises de diagnóstico, por meio de análise de resíduos e diagnóstico de influência. São conduzidos estudos de simulações para verificar as propriedades de máxima verossimilhança dos estimadores.

**Palavras-chave:** modelos não lineares, assimetria, efeitos mistos, misturas de escala skew-normal, algoritmo EM.



# ABSTRACT

PEREIRA, M. A. A. **Nonlinear mixed-effects models with scale mixture of skew-normal distributions**. 2019. 150 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

This work aims to develop asymmetric nonlinear regression models with mixed-effects, which provide alternatives to the use of normal distribution and other symmetric distributions, in order to avoid the sensitivity in the estimates to atypical observations and asymmetry. Nonlinear models with mixed-effects are explored in several areas of knowledge, especially when data are correlated, such as longitudinal data, repeated measures and multilevel data, in particular, for their flexibility in dealing with measures of areas such as biology and pharmacokinetics. However, there are difficulties in obtaining explicit estimators for the parameters in these models. At present, many studies have been developed with the family scale mixtures of skew-normal distribution (SMSN) that encompasses distributions with light and heavy tails, such as skew-normal, skew-Student-t, skew-contaminated normal and skew-slash, as well as symmetrical versions of these distributions. In this work, nonlinear regression models with mixed-effects are presented in which the random components have distributions belonging to the SMSN family. For the parameters estimation, a numerical solution via the EM algorithm and its extensions and Newton-Raphson algorithm is obtained. Analyzes for real data sets are performed with this new proposal, such as the study of drug kinetics in humans, as well as diagnostic analyzes, through residual analysis and influence diagnostics. Simulation studies are conducted to verify the maximum likelihood properties of the estimators.

**Keywords:** nonlinear model, skewness, mixed-effects, scale mixtures of skew-normal, EM algorithm.



# LISTA DE ILUSTRAÇÕES

---

---

Figura 1 – Interpretação dos parâmetros $Asym$ , $xmid$ e $scal$ em um modelo logístico. . . . .	30
Figura 2 – Valores observados do diâmetro do tronco de cinco laranjeiras. . . . .	30
Figura 3 – Boxplots dos resíduos ordinários por árvore de laranjeira para todo o conjunto de dados. . . . .	31
Figura 4 – Curva ajustada para a média, sem considerar a correlação entre as laranjeiras. . . . .	32
Figura 5 – Boxplots dos resíduos ordinários por árvore de laranjeira. . . . .	33
Figura 6 – Curvas ajustadas para cada laranjeira. . . . .	33
Figura 7 – Concentração de cefamandol contra o tempo após a administração. . . . .	34
Figura 8 – A função biexponencial como combinação linear de duas exponenciais. . . . .	35
Figura 9 – Perfis ajustados sob a distribuição normal por indivíduo dos dados de cefamandol. . . . .	36
Figura 10 – Histogramas dos valores preditos da variável resposta por medição dos dados de cefamandol. . . . .	37
Figura 11 – Interpretação do parâmetro $lke$ , $lka$ e $lcl$ em um modelo de compartimentos. . . . .	38
Figura 12 – Concentração de <i>theophylline</i> contra o tempo após a administração. . . . .	39
Figura 13 – Perfis ajustados sob a distribuição normal por indivíduo dos dados de <i>theophylline</i> . . . . .	40
Figura 14 – Histogramas dos valores preditos da variável resposta por medição dos dados de <i>theophylline</i> . . . . .	41
Figura 15 – Densidades da distribuição skew-normal padrão univariada com diferentes valores para o parâmetro de assimetria. . . . .	44
Figura 16 – Densidades de distribuições univariadas pertencentes a família SMSN. . . . .	54
Figura 17 – Contornos da distribuição skew-normal bivariada. . . . .	55
Figura 18 – Contornos da distribuição skew-t bivariada. . . . .	55
Figura 19 – Contornos da distribuição skew-normal contaminada bivariada. . . . .	55
Figura 20 – Perfis ajustados sob o modelo skew-normal contaminada por árvore de laranjeira. . . . .	70
Figura 21 – Perfis ajustados sob o modelo skew-normal contaminada por indivíduo dos dados de cefamandol. . . . .	73
Figura 22 – Perfis ajustados sob o modelo skew-normal contaminada por indivíduo dos dados de <i>theophylline</i> . . . . .	75
Figura 23 – Razões de verossimilhanças baseadas nas verossimilhanças perfiladas para os dados de <i>theophylline</i> . . . . .	77

Figura 24 – Valores preditos e observados da concentração de <i>theophylline</i> em dois tempos.	78
Figura 25 – Distâncias de Mahalanobis dos modelos ajustados aos dados de crescimento de laranjeiras. . . . .	91
Figura 26 – Gráficos dos resíduos condicionados das medições por árvore dos dados de crescimento de laranjeiras. . . . .	91
Figura 27 – Gráficos dos resíduos quantílicos aleatorizados para os dados das laranjeiras.	92
Figura 28 – Gráficos dos resíduos quantílicos aleatorizados e envelopes simulados para os dados das laranjeiras. . . . .	92
Figura 29 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação ponderação de casos com $c^* = 0,5$ . . . . .	93
Figura 30 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na variável resposta com $c^* = 0,5$ .	94
Figura 31 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na matriz de dispersão dos efeitos aleatórios com $c^* = 0,5$ . . . . .	94
Figura 32 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na matriz de dispersão dos erros com $c^* = 0,5$ . . . . .	95
Figura 33 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na matriz de dispersão dos erros na direção da árvore 4. . . . .	96
Figura 34 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na variável resposta na direção da árvore 1. . . . .	96
Figura 35 – Distâncias de Mahalanobis dos modelos ajustados aos dados do antibiótico cefamandol. . . . .	98
Figura 36 – Gráfico de resíduos condicionados das medições por indivíduos dos dados de cefamandol. . . . .	99
Figura 37 – Gráficos dos resíduos quantílicos aleatorizados para os dados de cefamandol.	99
Figura 38 – Gráficos dos resíduos quantílicos aleatorizados e envelopes simulados para os dados de cefamandol. . . . .	100
Figura 39 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação ponderação de casos com $c^* = 1$ .	101
Figura 40 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na variável resposta com $c^* = 0,5$ .	101
Figura 41 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na matriz de dispersão dos efeitos aleatórios com $c^* = 0,5$ . . . . .	102
Figura 42 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na matriz de dispersão dos erros com $c^* = 0,5$ . . . . .	102
Figura 43 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na variável resposta na direção do indivíduo 3. . . . .	103
Figura 44 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação ponderação de casos na direção do indivíduo 4. . . . .	104
Figura 45 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na matriz de dispersão dos efeitos aleatórios na direção do indivíduo 5. . . . .	104

Figura 46 – Distância de Mahalanobis dos dados de <i>theophylline</i> . . . . .	106
Figura 47 – Gráfico de resíduos condicionados das medições por indivíduos dos dados de <i>theophylline</i> . . . . .	107
Figura 48 – Gráficos dos resíduos quantílicos aleatorizados para os dados de <i>theophylline</i> . . . . .	107
Figura 49 – Gráficos dos resíduos quantílicos aleatorizados e envelopes simulados para os dados de <i>theophylline</i> . . . . .	108
Figura 50 – Gráfico dos resíduos quantílicos aleatorizados e envelopes simulados de ajustes sob distribuições simétricas para os dados de <i>theophylline</i> . . . . .	109
Figura 51 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação ponderação de casos com $c^* = 0,5$ . . . . .	110
Figura 52 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na variável resposta com $c^* = 0,5$ . . . . .	110
Figura 53 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na matriz de dispersão dos efeitos aleatórios com $c^* = 0,7$ . . . . .	111
Figura 54 – Gráficos de $\mathbf{m}(0)$ sob esquema de perturbação na matriz de dispersão dos erros com $c^* = 0,7$ . . . . .	111
Figura 55 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação ponderação de casos na direção do indivíduo 7. . . . .	112
Figura 56 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na matriz de dispersão dos erros na direção do indivíduo 5. . . . .	113
Figura 57 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na variável resposta na direção do indivíduo 3. . . . .	114



# LISTA DE TABELAS

---

---

Tabela 1 – Coeficientes do ajuste para todas as laranjeiras. . . . .	31
Tabela 2 – Coeficientes do ajuste para cada árvore de laranjeira. . . . .	32
Tabela 3 – Resultados sobre a distância de Mahalanobis por distribuição pertencente à classe SMSN. . . . .	53
Tabela 4 – Estimativa dos parâmetros (Estim.) com erros padrões (E.P.) para os modelos N-SMSN, SN-NLME, ST-NLME, SSL-NLME e SNC-NLME ajustados aos dados de crescimento de laranjeiras. . . . .	69
Tabela 5 – Valores AIC e BIC para os modelos ajustados aos dados de crescimento de laranjeiras. . . . .	69
Tabela 6 – Resultados do teste da razão de verossimilhanças para testar $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ aplicados aos dados de crescimento de laranjeiras. . . . .	70
Tabela 7 – Valores absolutos do desvio relativo entre valores observados e preditos no modelo SNC-NMLE e dados de crescimento de laranjeiras. . . . .	71
Tabela 8 – Estimativa dos parâmetros (Estim.) com erros padrões (E.P.) para os modelos N-SMSN, SN-NLME, ST-NLME, SSL-NLME e SNC-NLME ajustados aos dados do cefamandol. . . . .	72
Tabela 9 – Valores AIC e BIC para os modelos ajustados aos dados de cefamandol. . . . .	72
Tabela 10 – Resultados do teste da razão de verossimilhanças para testar $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ aplicados aos dados de cefamandol. . . . .	73
Tabela 11 – Estimativa dos parâmetros (Estim.) com erros padrões (E.P.) para os modelos N-SMSN, SN-NLME, ST-NLME, SSL-NLME e SNC-NLME ajustados aos dados de <i>theophylline</i> . . . . .	74
Tabela 12 – Valores AIC e BIC para os modelos ajustados aos dados de <i>theophylline</i> . . . . .	75
Tabela 13 – Resultados do teste da razão de verossimilhanças para testar $H_0 : \boldsymbol{\lambda} = \mathbf{0}$ aplicados aos dados de <i>theophylline</i> . . . . .	76
Tabela 14 – Intervalos de confiança aproximados de 95% para o parâmetros de assimetria $\lambda_1$ obtidos com a estatística de razão de verossimilhanças para os dados de <i>theophylline</i> . . . . .	76
Tabela 15 – Valores absolutos do desvio relativo entre valores observados e preditos no modelo SNC-NMLE e dados de <i>theophylline</i> . . . . .	78
Tabela 16 – Média, viés, EQM e CP dos modelos ajustados sob as distribuições SN, ST e SNC por meio de 1000 amostras simuladas de Monte Carlo com $n = 5$ e $m_i = 7$ . . . . .	80

Tabela 17 – Média, viés, EQM e CP dos modelos ajustados sob as distribuições SN, ST e SNC por meio de 1000 amostras simuladas de Monte Carlo com $n = 6$ e $m_i = 11$ . . . . .	80
Tabela 18 – Média, viés, EQM e PC dos modelos ajustados sob as distribuições SN, ST e SNC por meio de 1000 amostras simuladas de Monte Carlo com $n = 9$ e $m_i = 6$ . . . . .	81
Tabela 19 – Média, viés, EQM e PC dos modelos ajustados sob as distribuições SN, ST e SNC por meio de 1000 amostras simuladas de Monte Carlo com $n = 12$ e $m_i = 11$ . . . . .	82
Tabela 20 – Percentuais de mudanças nas estimativas dos parâmetros quando árvores foram excluídas para os dados de crescimento de laranjeiras. . . . .	97
Tabela 21 – Percentuais de mudanças nas estimativas dos parâmetros quando árvores foram excluídas para os dados de cefamandol. . . . .	105
Tabela 22 – Percentuais de mudanças nas estimativas dos parâmetros quando indivíduos foram excluídos para os dados de <i>theophylline</i> . . . . .	115
Tabela 23 – Estimativas dos parâmetros como erros padrões (E.P.) do modelo SN-NLME ajustados aos dados de crescimento de laranjeiras. . . . .	124
Tabela 24 – Estimativas dos parâmetros como erros padrões (E.P.) do modelo SN-NLME ajustados aos dados de cefamandol. . . . .	125
Tabela 25 – Estimativas dos parâmetros como erros padrões (E.P.) do modelo SN-NLME ajustados aos dados de <i>theophylline</i> . . . . .	126
Tabela 26 – Média, viés e erro quadrático médio dos ajustes sob o modelo skew-normal baseadas em 1000 amostras Monte Carlo com $\lambda_b = 0,1$ e $\lambda_b = 0,5$ . . . . .	127

# SUMÁRIO

---

---

1	INTRODUÇÃO	23
1.1	Análise de diagnóstico	25
1.2	Suporte computacional	26
1.3	Produtos da tese	26
1.4	Objetivos e organização do trabalho	26
2	MOTIVAÇÃO	29
2.1	Dados de crescimento de laranjeiras	29
2.2	Dados do antibiótico cefamandol	34
2.3	Dados do antiasmático <i>theophylline</i>	38
3	FAMÍLIA DE DISTRIBUIÇÕES MISTURAS DE ESCALA SKEW-NORMAL	43
3.1	Momentos	47
3.2	Distribuição marginal e independência	48
3.3	Casos particulares	50
3.4	Representação hierárquica	56
3.5	Algoritmo EM em modelos SMSN	57
4	MODELO NÃO LINEAR MISTO COM MISTURAS DE ESCALA SKEW-NORMAL	59
4.1	Estimação por máxima verossimilhança	61
4.1.1	<i>Predição dos efeitos aleatórios e da variável resposta</i>	65
4.2	Estatística da razão de verossimilhanças	67
4.2.1	<i>Teste de hipóteses</i>	67
4.2.2	<i>Intervalo de confiança aproximado</i>	67
4.3	Aplicações	67
4.3.1	<i>Dados de crescimento de laranjeiras</i>	68
4.3.2	<i>Dados do antibiótico cefamandol</i>	71
4.3.3	<i>Dados do antiasmático theophylline</i>	74
4.4	Simulações	78
5	ANÁLISE DE DIAGNÓSTICO	83
5.1	Análise de resíduos	83

5.2	Diagnóstico de influência . . . . .	84
5.2.1	<i>Influência local</i> . . . . .	84
5.2.1.1	<i>Esquemas de perturbação</i> . . . . .	87
5.2.2	<i>Dados de crescimento de laranjeiras</i> . . . . .	90
5.2.3	<i>Dados de antibiótico cefamandol</i> . . . . .	98
5.2.4	<i>Dados do antiasmático theophylline</i> . . . . .	106
6	MODELO NÃO LINEAR MISTO SKEW-NORMAL . . . . .	117
6.1	Estimação por máxima verossimilhança . . . . .	120
6.2	Aplicações . . . . .	123
6.2.1	<i>Dados de crescimento de laranjeiras</i> . . . . .	123
6.2.2	<i>Dados do antibiótico cefamandol</i> . . . . .	124
6.2.3	<i>Dados do antiasmático theophylline</i> . . . . .	125
6.3	Simulações . . . . .	126
7	CONSIDERAÇÕES FINAIS . . . . .	129
7.1	Conclusões . . . . .	129
7.2	Perspectivas de trabalhos futuros . . . . .	131
	REFERÊNCIAS . . . . .	133
	APÊNDICE A           ALGORITMOS EM E NEWTON-RAPHSON . . . . .	139
A.1	Algoritmo EM . . . . .	139
A.2	Algoritmo de Newton-Raphson . . . . .	140
	APÊNDICE B           RESULTADOS ADICIONAIS DO CAPÍTULO 3 . . . . .	143
B.1	Matriz de informação de Fisher observada . . . . .	143
	APÊNDICE C           RESULTADOS ADICIONAIS DO CAPÍTULO 5 . . . . .	145
C.1	Envelopes simulados . . . . .	145
C.2	Matriz hessiana . . . . .	146
	APÊNDICE D           RESULTADOS ADICIONAIS DO CAPÍTULO 6 . . . . .	149
D.1	Matriz hessiana . . . . .	149

---

## INTRODUÇÃO

---

Modelos com efeitos mistos têm aplicação em problemas estatísticos em diversas áreas do conhecimento, em particular quando os dados são correlacionados, como dados longitudinais, medidas repetidas e multiníveis. Os modelos não lineares com efeitos mistos (NLMEs) recentemente têm sido amplamente explorados em áreas como Ciências Biológicas, Economia, Engenharia e Farmacocinética. Modelos para dados com medidas repetidas ganharam popularidade devido à sua flexibilidade na estrutura de covariâncias e por permitir correlação não constante entre as observações, além da capacidade de lidar tanto com dados balanceados como desbalanceados. Os modelos não lineares são utilizados geralmente quando existe uma interpretação física nas relações entre dados e parâmetros de um determinado problema que claramente não podem ser explicadas pela linearidade. Contudo, nestes modelos, existem dificuldades na obtenção de estimadores explícitos para os parâmetros. Os efeitos aleatórios e os erros, em geral, são apresentados sob suposição de normalidade, como é tratado em Lindstrom e Bates ([LINDSTROM; BATES, 1990](#)), mas a suposição de normalidade pode ocasionar a falta de precisão nas estimativas da variação entre os sujeitos, quando há presença de observações atípicas e de assimetria e, portanto, não ser realista.

Uma forma de minimizar problemas nas estimações dos parâmetros relacionados a presença de observações atípicas e assimetria nos dados consiste na transformação de variáveis, a fim de alcançar a normalidade. Porém, Azzalini e Capitanio ([AZZALINI; CAPITANIO, 1999](#)) apontaram alguns problemas, a exemplo da dificuldade de interpretação dos resultados, especialmente quando cada variável é transformada usando uma função diferente e também quando a homocedasticidade multivariada é necessária, isso geralmente requer uma transformação que é diferente da transformação para normalidade.

Desta forma, em tratamentos de dados em que a suposição de normalidade não é satisfeita, algumas metodologias robustas alternativas foram desenvolvidas, como por exemplo a utilização de distribuições com caudas pesadas. Russo et al. ([RUSSO; PAULA; AOKI, 2009](#)) desenvol-

veram um modelo NLME com distribuições elípticas (EL) e Meza et al. (MEZA; OSORIO; CRUZ, 2012) desenvolveram um modelo NLME com uma subclasse de distribuições elípticas, denominada de misturas de escala da distribuição normal (SMN), introduzida por Andrews e Mallows (ANDREWS; MALLOWS, 1974). Estudos de diagnóstico e influência local também têm ganhado importância em modelos NLME, nessa perspectiva podemos citar trabalhos como Tang et al. (TANG; WEI; ZHANG, 2006), Lee Xu (LEE; XU, 2004), Zhong et al. (ZHONG *et al.*, 2003), Beckman et al. (BECKMAN; NACHTSHEIM; COOK, 1987), Paula (PAULA, 1996), Tang et al. (TANG; WEI; WANG, 2001) e Matos et al. (MATOS *et al.*, 2013).

Para tratar dados que não tenham uma relação linear, levando-se em consideração a assimetria e sem efeitos aleatórios, algumas propostas foram desenvolvidas, como podemos verificar em Cancho et al. (CANCHO; LACHOS; ORTEGA, 2010) que desenvolveram um modelo de regressão não linear (NLM) com a distribuição skew-normal com a abordagem clássica e bayesiana, e que foi posteriormente estendido por Garay et al. (GARAY; LACHOS; ABANTO-VALLE, 2011) para distribuições assimétricas com caudas leves e pesadas, e em Ferreira e Lachos (FERREIRA; LACHOS, 2016) que desenvolveram um modelo não linear sob a classe de distribuições *skew-scale mixtures of normal* (SSMN).

Daremos atenção a modelos lineares com efeitos mistos, pois iremos apresentar modelos não lineares nos efeitos fixos, mas que são lineares nos efeitos aleatórios. Assim, podemos citar trabalhos como Laird e Ware (LAIRD; WARE, 1982) que apresentaram um modelo linear com efeitos mistos (LME) em que as componentes aleatórias são normalmente distribuídas, Savalli et al. (SAVALLI; PAULA; CYSNEIROS, 2006) apresentaram um modelo LME sob a classe de distribuições com contornos elípticos e Lachos et al. (LACHOS; GHOSH; ARELLANO-VALLE, 2010) que apresentaram um modelo linear misto sob a classe de distribuições skew-normal/independente (SNI). No contexto de acomodação de observações destoantes ou atípicas em modelos lineares mistos, alguns pesquisadores deram importantes contribuições, tais como Richardson e Welsh (RICHARDSON; WELSH, 1995) que discutiram a aplicação de métodos robustos para máxima verossimilhança restrita, Richardson (RICHARDSON, 1997) que propõe uma estimação de influência limitada para reduzir a ação de observações atípicas, tanto na variável resposta quanto na variável explicativa e Gill (GILL, 2000) que sugere robustificar o logaritmo da função de verossimilhança e aplica o procedimento proposto para estimação robusta em dados longitudinais.

Sob a perspectiva Bayesiana, Lachos et al. (LACHOS; BANDYOPADHYAY; DEY, 2011) apresentaram modelos lineares e não lineares com efeitos mistos para acomodar dados com censura sob a classe de distribuições normal independente ou *normal/independent* (NI) no tratamento de dados de carga viral (HIV). Huang e Dagne (HUANG; DAGNE, 2011) apresentaram um modelo NLME Bayesiano com erros de medida, assumindo que os efeitos aleatórios seguem a distribuição skew-normal, porém não encontrou diferença significativa quando comparou com o modelo em que os efeitos aleatórios são normalmente distribuídos. Lu e Huang (LU;

HUANG, 2014) propuseram um modelo Bayesiano não linear nos efeitos fixos e aleatórios para dados longitudinais, levando-se em conta a heterogeneidade e assimetria, com as distribuições normal, skew-normal e skew-t. Com a classe de distribuições elípticas, De la Cruz e Branco (CRUZ; BRANCO, 2009) apresentaram um modelo não linear Bayesiano para tratar dados com medidas repetidas. Liu e Wu (LIU; WU, 2007) propuseram dois métodos de verossimilhança aproximada para modelos NLME semiparamétricos, sob normalidade e com erros de medida nas covariáveis.

Neste trabalho apresentamos modelos para ajustar dados que não possuem relação linear com efeitos mistos e levando-se em consideração a assimetria. Desenvolvemos um modelo flexível em que a variável resposta é condicionada ao efeito aleatório (LACHOS; GHOSH; ARELLANO-VALLE, 2010) e estas componentes aleatórias seguem uma distribuição pertencente a família misturas de escala da distribuição skew-normal (SMSN) (BRANCO; DEY, 2001; KIM, 2008; ZELLER; LACHOS; VILCA-LABRA, 2011). Denominamos este modelo de modelo não linear com efeitos mistos com misturas de escala da distribuição skew-normal ou *nonlinear mixed-effects models with scale mixture of skew-normal distributions* SMSN-NLME. A família SMSN engloba distribuições com caudas leves e pesadas, como a skew-normal (SN), skew-Student-t (ST), skew-normal contaminada (SNC), skew-slash (SSL), como também as distribuições normal (N), Student-t (T), normal contaminada (CN) e slash (SL), que são as versões simétricas destas distribuições. Os erros seguem uma distribuição pertencente a família SMN, que é um caso particular da família SMSN. A família SMSN de distribuições possui uma representação hierárquica que possibilita a implementação de um algoritmo EM (DEMPSTER; LAIRD; RUBIN, 1977) e de suas extensões, e do algoritmo Newton-Raphson para a estimação dos parâmetros do modelo pelo método de máxima verossimilhança, por este motivo optamos pela abordagem clássica e também porque não temos informações prévias sobre os parâmetros. No Apêndice A são apresentados resumos sobre os algoritmos EM e Newton-Raphson.

Adicionalmente, desenvolvemos um modelo em que a variável resposta não está condicionada ao efeito aleatório e sob a distribuição skew-normal (ANDREWS; MALLOWS, 1974), cujas estimativas também são obtidas por máxima verossimilhança.

## 1.1 Análise de diagnóstico

Com o objetivo de avaliar a qualidade do modelo não linear sob a família SMSN de distribuições, vamos realizar um estudo de análise de diagnóstico em conjunto de dados reais. Em modelagem estatística a verificação da coerência das suposições, por meio de alguma medida apropriada, é de fundamental importância e recebe contribuições de diversos autores. A análise de diagnóstico é feita, em geral, por meio de análise de resíduos e diagnóstico de influência. O estudo de análise de resíduos tem como finalidade verificar a distância entre os pontos ajustados dos pontos previstos. No estudo de diagnóstico de influência, as técnicas de influência global

baseadas, por exemplo, na exclusão de casos e técnicas de influência local desenvolvidas por Cook (COOK, 1986). A abordagem de influência local é baseada no trabalho proposto por Cook (COOK, 1986), Poon e Poon (POON; POON, 1999) e Zhu e Lee (ZHU; LEE, 2001), aplicados em vários cenários e esquemas de perturbação, considerando as distribuições marginais e representações hierárquicas.

## 1.2 Suporte computacional

As aplicações, simulações e gráficos com respeito a teoria apresentada neste trabalho foi implementada no *software* R versão 3.5.2 (R Core Team, 2018), que é uma linguagem de programação e um ambiente de desenvolvimento integrado, de uso livre, para cálculos matemáticos, estatísticos e construção de gráficos, disponível no site [www.r-project.org](http://www.r-project.org). As bibliotecas do *software* R que utilizamos foram *nlme* (PINHEIRO *et al.*, 2018), *Runuram* (LEYDOLD; HÖRMANN, 2017), *mixsmsn* (PRATES; CABRAL; LACHOS, 2013), *EMMIXskew* (WANG; NG; MCLACHLAN., 2018), *sn* (AZZALINI, 2018), *hnp* (MORAL; HINDE; DEMÉTRIO, 2017), *mvtnorm* (GENZ; BRETZ, 2009), *numDeriv* (GILBERT; VARADHAN, 2016), *lattice* (SARKAR, 2008) e *ggplot2* (WICKHAM, 2016).

## 1.3 Produtos da tese

As pesquisas realizadas e apresentadas nesta tese permitiram a publicação dos seguintes trabalhos

- PEREIRA, Marcos Antonio Alves; RUSSO, Cibele Maria. Nonlinear mixed-effects models with scale mixture of skew-normal distributions. *Journal of Applied Statistics*, v. 46, n. 9, p. 1602-1620, 2019.
- PEREIRA, Marcos Antonio Alves; RUSSO, Cibele Maria. Nonlinear mixed-effects models with scale mixture of skew-normal distributions. *Proceedings of the 32nd International Workshop on Statistical Modelling*, p. 142-145, 2017.
- PEREIRA, Marcos Antonio Alves; RUSSO, Cibele Maria. *Modelos não lineares assimétricos com efeitos mistos*. Pôster apresentado na XV Escola de Modelos de Regressão. Goiânia. 2017.

## 1.4 Objetivos e organização do trabalho

O objetivo geral deste trabalho é o desenvolver procedimentos de estimação na classe dos modelos não lineares com efeitos mistos sob a família de distribuições misturas de escala skew-normal multivariada, baseados em técnicas de estimação frequentistas. Estes modelos são

não lineares nos efeitos fixos, porém lineares nos efeitos aleatórios. As estimativas de máxima verossimilhança são obtidas por meio de processos iterativos. Destacamos os seguintes objetivos específicos

- I Propor modelos não lineares com efeitos mistos para tratar dados correlacionados e levando-se em consideração assimetria, ou seja, com este modelo é possível ajustar dados em que a variável resposta segue uma distribuição simétrica ou assimétrica.
- II Introduzir uma classe de distribuições assimétricas multivariadas que possua uma estrutura conveniente para realizar inferências em modelos estatísticos, em especial, modelos não lineares mistos.
- III Propor procedimentos de estimação no modelo não linear misto com uma família de distribuições assimétricas pelo método de máxima verossimilhança via algoritmo do tipo EM. Esses modelos são lineares nos efeitos aleatórios.
- IV Realizar estudo de diagnóstico no modelo não linear misto, por meio da análise de influência local e global, e análise de resíduos.
- V Apresentar um modelo não linear misto em que a variável resposta não está condicionada ao efeito aleatório, cujas estimações são feitas pelo método de máxima verossimilhança via algoritmo do tipo EM e sob a distribuição skew-normal.

#### • **Organização do trabalho**

Apresentaremos no Capítulo 2 algumas motivações que impulsionaram o desenvolvimento deste trabalho, com aplicações em conjuntos de dados reais. No Capítulo 3 abordaremos a teoria já desenvolvida sobre a família SMSN de distribuições de probabilidade. No Capítulo 4 apresentaremos e discutiremos sobre modelos não lineares com efeitos mistos sob a família SMSN de distribuições, em que a inferência sobre o vetor de parâmetros é baseada na distribuição marginal da variável resposta. No Capítulo 4 são apresentados também procedimentos de estimação dos parâmetros, via algoritmo do tipo EM, estimações dos efeitos aleatórios e predições, além de aplicações nos conjuntos de dados apresentados no Capítulo 2 e estudo de simulação. Discutiremos brevemente a análise de resíduos e ferramentas para diagnóstico de influência no Capítulo 5. No Capítulo 6 apresentaremos um modelo não linear com efeitos mistos sob a distribuição skew-normal, no qual distribuição marginal da variável resposta não é condicionada ao efeito aleatório, como abordado no Capítulo 4. Finalmente no Capítulo 7, faremos algumas conclusões e apresentaremos algumas perspectivas de trabalhos futuros.



---

## MOTIVAÇÃO

---

Os modelos não lineares com efeitos mistos possuem grande aplicabilidade em problemas em que os dados são correlacionados. Estes tipos de problemas são encontrados em diversas áreas do conhecimento. Em geral, os problemas abordados com essa metodologia utilizam funções não lineares com parâmetros não conhecidos, com o objetivo de explicar ou descrever fenômenos estudados, e a estas funções ou a estes parâmetros são introduzidos efeitos aleatórios. Em especial, são comuns na literatura, análises de dados não lineares da área de Farmacocinética. Neste Capítulo, apresentamos exemplos de ajustes de dados reais correlacionados com estrutura não linear, previamente discutidos por outros autores e supondo que os erros e efeitos aleatórios são normalmente distribuídos.

### 2.1 Dados de crescimento de laranjeiras

A medida da circunferência do tronco de cinco laranjeiras (em *mm*) de um bosque no sul do estado da Califórnia nos Estados Unidos foram medidas ao longo do tempo (em dias). Os dados, colhidos em sete ocasiões diferentes, são apresentados em Pinheiro e Bates (PINHEIRO; BATES, 2000) e podem também ser encontrados na biblioteca *nlme* (PINHEIRO *et al.*, 2018) do *software R* (R Core Team, 2018) com o nome `Orange` (*datasets*).

Para ajustar estes dados, podemos considerar um modelo de regressão não linear logístico, considerando  $y_{ij}$  como a medida da circunferência do tronco, em milímetros, no tempo  $x_{ij}$ , em dias, dado pela expressão a seguir

$$y_{ij} = \frac{\text{Asym}}{1 + \exp[-(x_{ij} - \text{xmid})/\text{scal}]} + \varepsilon_{ij}, \quad i = 1, \dots, 5, \quad j = 1, \dots, 7, \quad (2.1)$$

em que  $\varepsilon_{ij}$  são as entradas do vetor aleatório  $\boldsymbol{\varepsilon}_i$ , normalmente distribuído, de dimensão  $7 \times 1$ .

Os elementos do vetor de parâmetros  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top = (\text{Asym}, \text{xmid}, \text{scal})^\top$  do modelo logístico podem ser interpretados como

- Asym: é a circunferência assintótica média do tronco ( $\alpha_1$ ),
- xmid: é a média da idade em que a árvore atinge a metade de sua circunferência assintótica ( $\alpha_2$ ) e
- scal: é a média da escala de crescimento ( $\alpha_3$ ).

Esta função é linear no parâmetro Asym e não linear nos parâmetros xmid e scal.

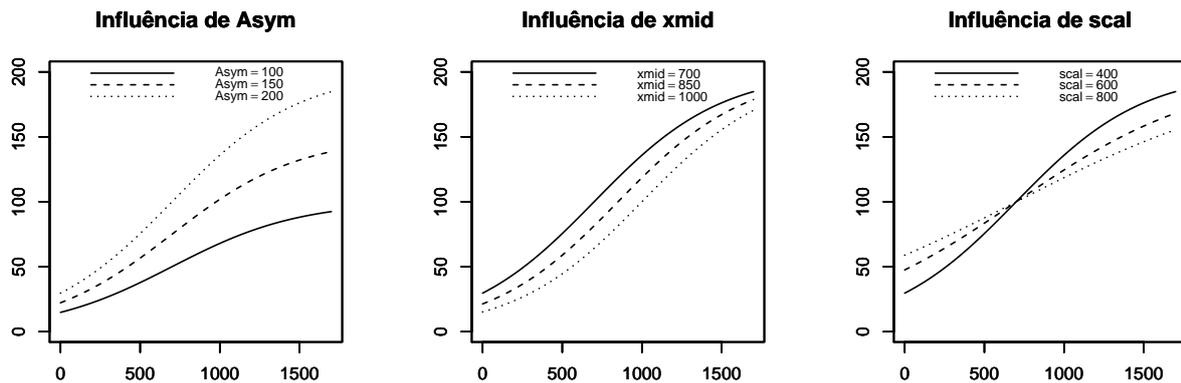


Figura 1 – Interpretação dos parâmetros Asym, xmid e scal em um modelo logístico.

Com os gráficos da Figura 1 podemos verificar as interpretações dos parâmetros, pois com o aumento do parâmetro Asym, observamos um aumento nos valores das observações finais, com o aumento do parâmetro xmid a curva sofre uma translação para baixo, indicando que a circunferência assintótica sofrerá uma redução, e com alterações no valor do parâmetro scal, a escala de crescimento também é alterada. Em cada gráfico, os demais parâmetros permaneceram constantes.

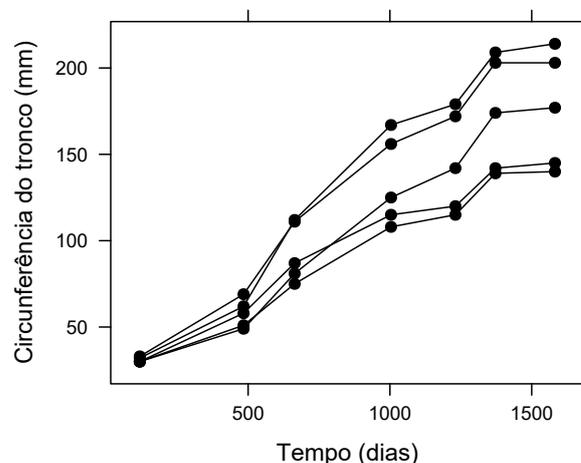


Figura 2 – Valores observados do diâmetro do tronco de cinco laranjeiras.

Podemos notar no gráfico de medidas observadas ao longo do tempo na Figura 2 que o crescimento da medida da circunferência do tronco é diferente em cada uma das cinco árvores.

Ajustando os dados com um modelo não linear de efeitos fixos, conforme (2.1), por meio da função `nls` do *software* R, teremos os coeficientes estimados

Tabela 1 – Coeficientes do ajuste para todas as laranjeiras.

Asym	xmid	scal
192,7	728,8	353,5

Na Tabela 1 podemos observar que o ajuste sob o modelo (2.1) são estimados apenas 3 parâmetros.

Os resíduos ordinários na Figura 3 tendem a ser negativos para as árvores 1 e 3 e positivos para as árvores 2 e 4.

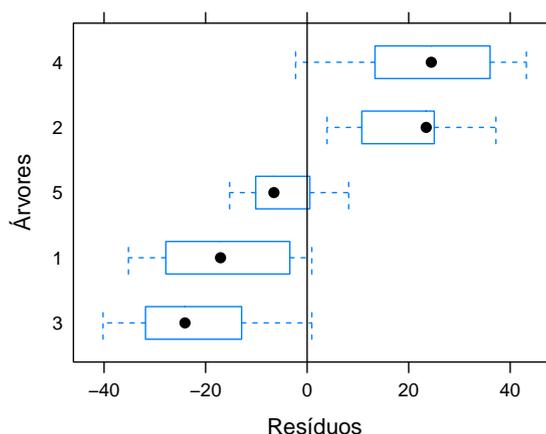


Figura 3 – Boxplots dos resíduos ordinários por árvore de laranjeira para todo o conjunto de dados.

Obviamente a curva ajustada com todos os dados, curva em preto na Figura 4, não é adequada para a maioria das árvores, cujos valores observados estão assinalados em cinza. Assim, analisando os gráficos das Figuras 3 e 4 é evidente a necessidade de um ajuste em que o "efeito árvore" é considerado.

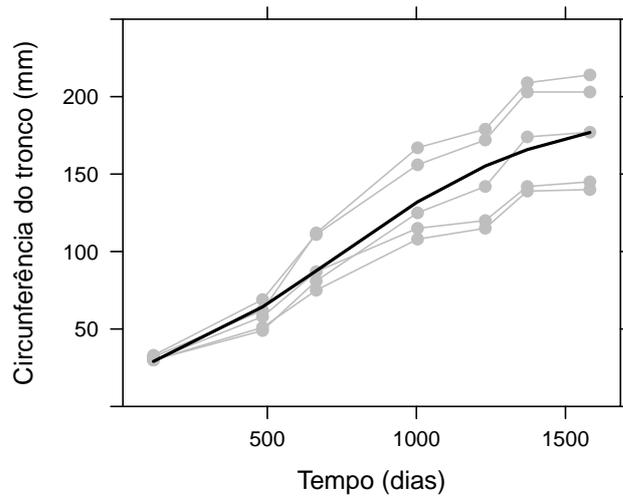


Figura 4 – Curva ajustada para a média, sem considerar a correlação entre as laranjeiras.

Fazendo ajuste com um modelo não linear levando em consideração o efeito aleatório de cada árvore, por meio da função `nlsList` da biblioteca *nlme* (mais detalhes sobre o modelo estão em (PINHEIRO *et al.*, 2018)), teremos os coeficientes estimados conforme apresentados na Tabela 2.

Tabela 2 – Coeficientes do ajuste para cada árvore de laranjeira.

Árvore	Asym	xmid	scal
1	154.1629	627.1923	362.5709
2	218.9983	700.3450	332.4934
3	158.8294	734.8421	400.9549
4	225.3038	710.7051	303.1438
5	207.2579	861.3037	379.9465

Uma desvantagem deste modelo é a estimação de 15 parâmetros para considerar os efeitos individuais de cada árvore.

No gráfico de resíduos ordinários mostrados na Figura 5, verificamos que não é mais evidente o "efeito de árvore", observado no ajuste anterior. Neste exemplo destacamos que a correlação intra-indivíduos deve ser considerada quando analisamos este tipo de dados. Tal abordagem é bastante conhecida e utilizada na literatura.

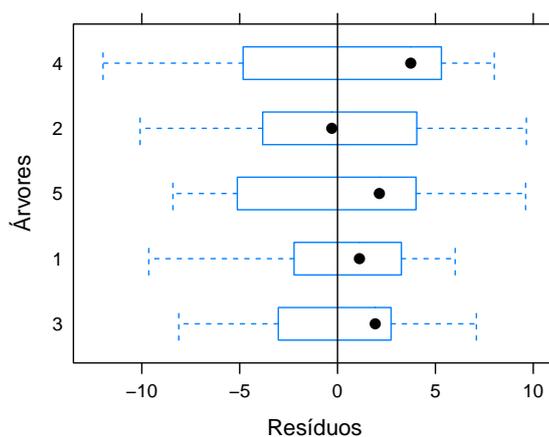


Figura 5 – Boxplots dos resíduos ordinários por árvore de laranjeira.

A Figura 6 mostra gráficos, gerados com o auxílio da biblioteca *lattice* (SARKAR, 2008), dos valores observados, em cinza, e ajustados, em preto, para cada árvore de laranjeira.

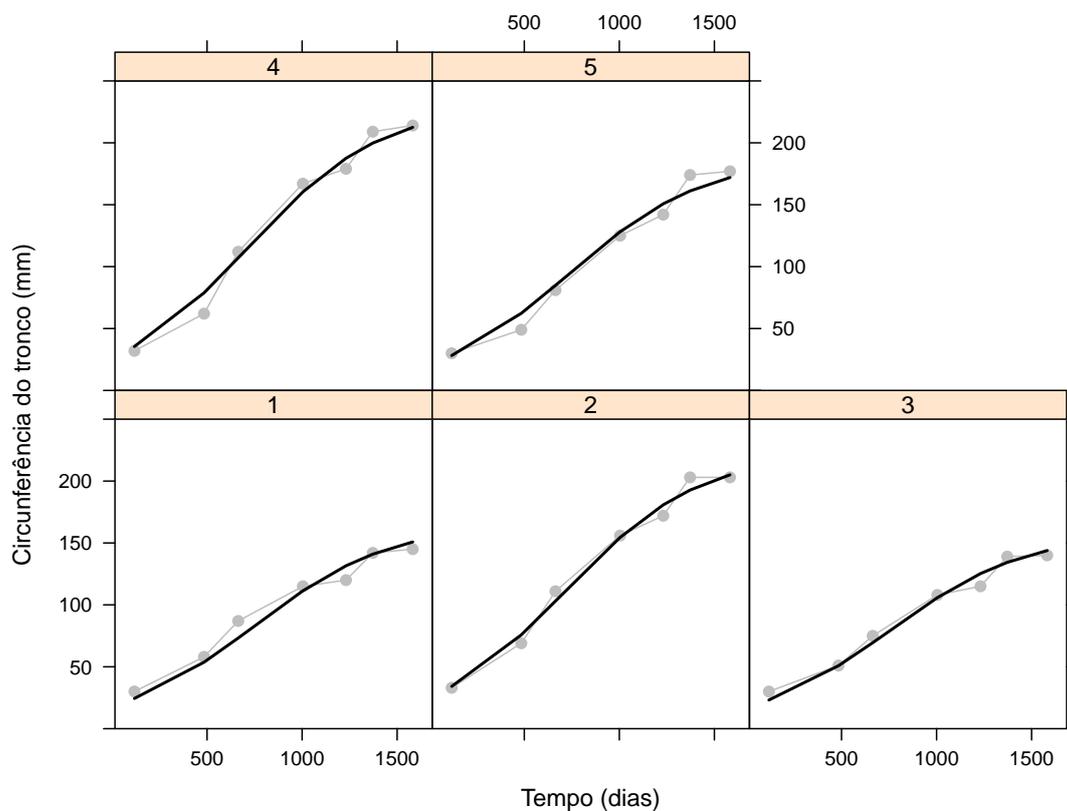


Figura 6 – Curvas ajustadas para cada laranjeira.

Nos exemplos a seguir vamos destacar, além da necessidade de efeitos aleatórios em modelos não lineares, a assimetria nestes efeitos, que é um dos objetivos principais deste trabalho.

## 2.2 Dados do antibiótico cefamandol

Davidian e Giltinan (DAVIDIAN; GILTINAN, 1995) descrevem dados obtidos durante um estudo para investigar a cinética da droga cefamandol. As concentrações plasmáticas (em  $mcg/ml$ ) da droga foram medidas em 6 voluntários saudáveis em 14 tempos (em minutos) diferentes após uma dose intravenosa de  $15\text{ mg/kg}$  de massa corporal de cefamandol. Os dados estão apresentados na Figura 7 e também podem ser encontrados na biblioteca *nlme* (PINHEIRO *et al.*, 2018) com o nome `Cefdamandole` (*datasets*).

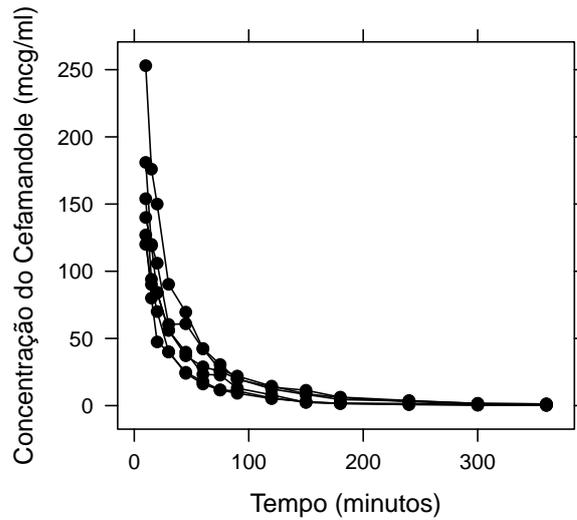


Figura 7 – Concentração de cefamandol contra o tempo após a administração.

Segundo Gibaldi e Perrier (GIBALDI; PERRIER, 1982), um método comum de modelagem de dados farmacocinéticos é representar o corpo humano como um sistema de compartimentos em que a droga é transferida de acordo com a cinética de primeira ordem (PINHEIRO; BATES, 2000). Neste modelo de compartimentos, a concentração da droga ao longo do tempo nos diferentes compartimentos é determinada por um sistema linear de equações diferenciais, cuja solução pode ser expressa como uma combinação linear de termos exponenciais. A função não linear utilizada para relacionar a concentração do cefamandol com o tempo é composto por dois compartimentos, ou seja, a concentração  $\mathbf{y}_i = (y_1, \dots, y_{14})^\top$  no tempo  $\mathbf{t}_i = (t_1, \dots, t_{14})^\top$  como uma combinação linear de duas exponenciais

$$f(\boldsymbol{\alpha}, t_{ij}) = \alpha_1 e^{-t_{ij} \alpha_2'} + \alpha_3 e^{-t_{ij} \alpha_4'}, \quad i = 1, \dots, 6, \quad j = 1, \dots, 14, \quad (2.2)$$

com  $\alpha_2' = \log \alpha_2$ ,  $\alpha_4' = \log \alpha_4$ ,  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top$ ,  $\alpha_2 > 0$  e  $\alpha_4 > 0$ . A função não linear apresentada acima é não identificável, pois os parâmetros  $\alpha_2$  e  $\alpha_4$  podem ser trocados sem alterar o valor de  $f(\boldsymbol{\alpha}, t_{ij})$ , o que pode ser corrigido se considerarmos  $\alpha_2 > \alpha_4$ .

Ajustando os dados com um modelo não linear considerando efeitos aleatórios associados aos parâmetros  $\alpha_2$  e  $\alpha_4$  e sob a distribuição normal para a variável resposta, teremos o seguinte

modelo

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 6, \quad (2.3)$$

em que  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) = (f(\boldsymbol{\alpha}, t_{i1}), \dots, f(\boldsymbol{\alpha}, t_{i14}))^\top$ ,  $\mathbf{Z}_i$ , de dimensão  $14 \times 2$ , é uma matriz de constantes conhecidas, e  $\boldsymbol{\varepsilon}_i \stackrel{\text{i.i.d.}}{\sim} N_{14}(\mathbf{0}, \sigma^2 \mathbf{I}_{14})$  e  $\mathbf{b}_i \stackrel{\text{i.i.d.}}{\sim} N_2(\mathbf{0}, \mathbf{D})$ ,  $i = 1, \dots, 6$ , são os erros e os efeitos aleatórios, respectivamente, e são não correlacionados. A matriz de dispersão dos efeitos aleatórios  $\mathbf{D}$  é não estruturada. Observe que o modelo é não linear nos efeitos fixos, mas linear nos efeitos aleatórios.

Quando consideramos  $\alpha_2 > \alpha_4$ , como mencionado anteriormente, as previsões do modelo (2.3) são modificadas, caso os parâmetros  $\alpha_2$  e  $\alpha_4$  sejam trocados. Ao assegurarmos esta identificabilidade do modelo (2.3), o primeiro termo exponencial determina a fase inicial de eliminação da droga e o segundo termo determina, principalmente, a fase final de eliminação da droga (PINHEIRO; BATES, 2000).

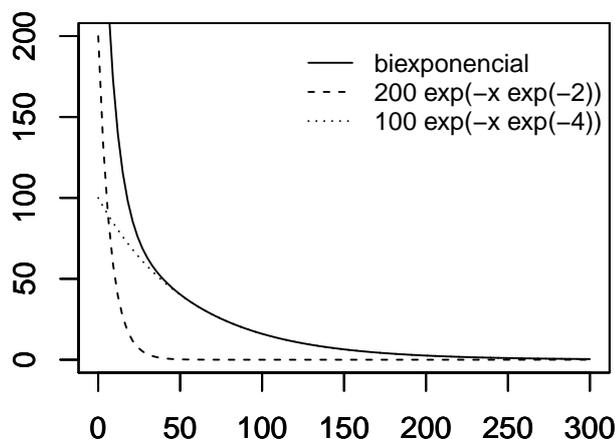


Figura 8 – A função biexponencial como combinação linear de duas exponenciais.

A Figura 8 mostra o gráfico da função biexponencial e os gráficos das funções exponenciais que a constituem, que são  $200e^{-xe^{-2}}$  e  $100e^{-xe^{-4}}$ .

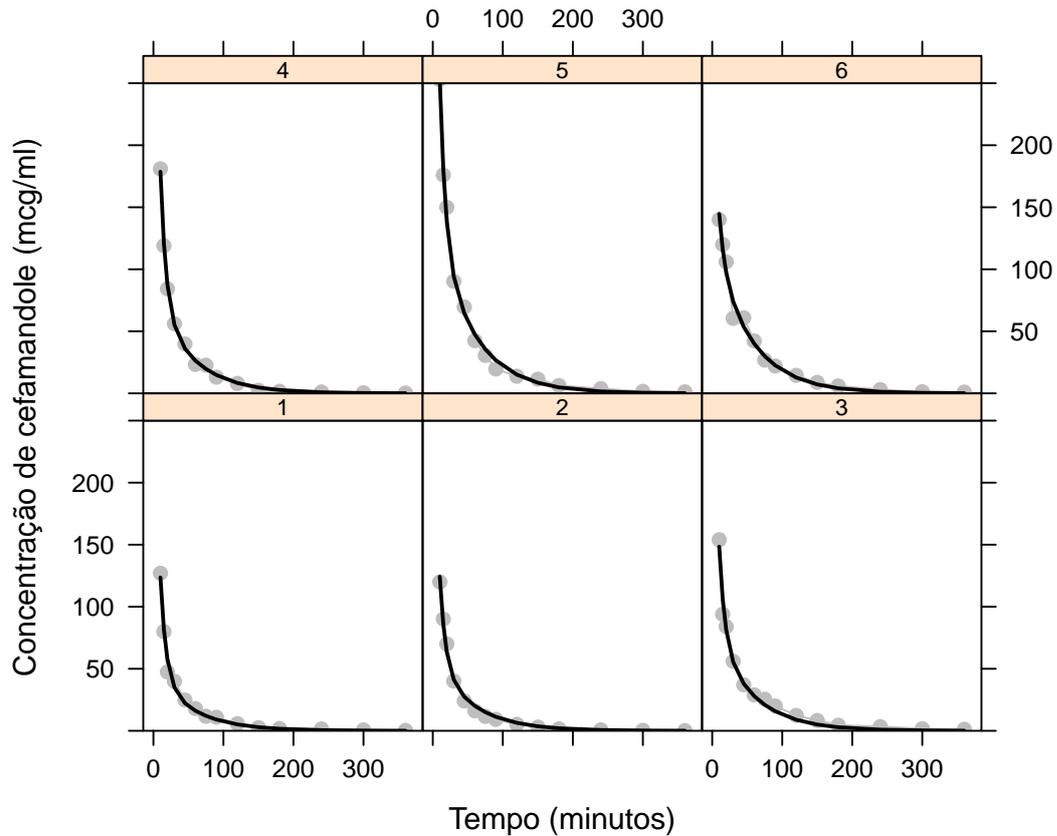


Figura 9 – Perfis ajustados sob a distribuição normal por indivíduo dos dados de cefamandol.

A Figura 9 mostra gráficos com valores observados, em cinza, e ajustados, em preto, dos dados de cefamandol para cada indivíduo.

Os histogramas apresentados na Figura 10 se referem aos valores preditos da variável resposta por medição em cada um dos 14 tempos, com os ajustes em que as componentes aleatórias seguem distribuições normais multivariadas, conforme modelo (2.3). Podemos verificar que há tendência a assimetria à esquerda.

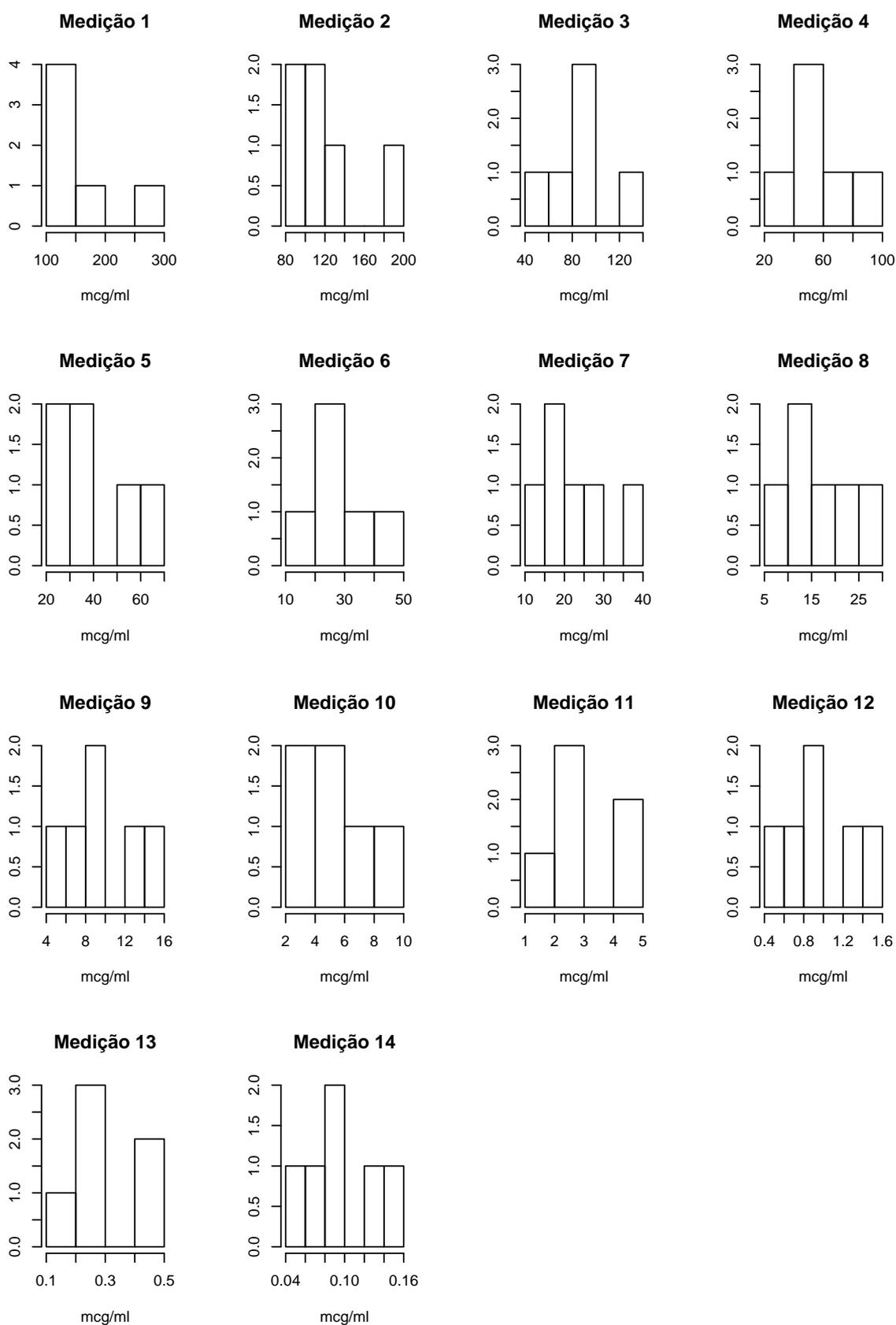


Figura 10 – Histogramas dos valores preditos da variável resposta por medição dos dados de cefamandol.

## 2.3 Dados do antiasmático *theophylline*

Com o objetivo de estudar a cinética do agente antiasmático *theophylline*, autores como Pinheiro e Bates (PINHEIRO; BATES, 2000) analisaram dados de um experimento em que a droga foi administrada em 12 pacientes, e as concentrações da substância no sangue (em  $mg/l$ ) foram coletadas em 11 tempos (em horas) em cada indivíduo nas 25 horas após a administração da dose  $D_s$  (em  $mg/kg$ ) da droga.

Para esse tipo de problema, que envolve a absorção e a eliminação de uma substância no organismo, é usual modelar a média da concentração de *theophylline* no sangue  $\mathbf{y}_i = (y_1, \dots, y_{11})^\top$  em função do tempo  $\mathbf{t}_i = (t_1, \dots, t_{11})^\top$  e da dose  $\mathbf{D}_s = (D_{s1}, \dots, D_{s11})^\top$  utilizando um modelo de compartimentos de primeira ordem com a relação não linear

$$f(\boldsymbol{\alpha}, t_{ij}) = \frac{D_{s_{ij}} e^{lke+lka-lcl}}{e^{lka} - e^{lke}} (e^{-e^{lke}t_{ij}} - e^{-e^{lka}t_{ij}}), \quad i = 1, \dots, 12, \quad j = 1, \dots, 11.$$

em que  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top = (lke, lka, lcl)^\top$  é o vetor de parâmetros.

A função não linear apresentada acima também apresenta problemas de não identificabilidade em relação aos parâmetros  $lke$  e  $lka$ , assim consideramos  $lka > lke$ .

Pinheiro e Bates (2000) (PINHEIRO; BATES, 2000) denominam este modelo não linear como modelo de um compartimento de primeira ordem com as seguintes interpretações dos parâmetros

- $lke$ : valor médio do logaritmo da constante da taxa de eliminação da substância ( $\alpha_1$ ),
- $lka$ : valor médio do logaritmo da constante da taxa de absorção da substância ( $\alpha_2$ ) e
- $lcl$ : valor médio do logaritmo da depuração plasmática ( $\alpha_3$ ).

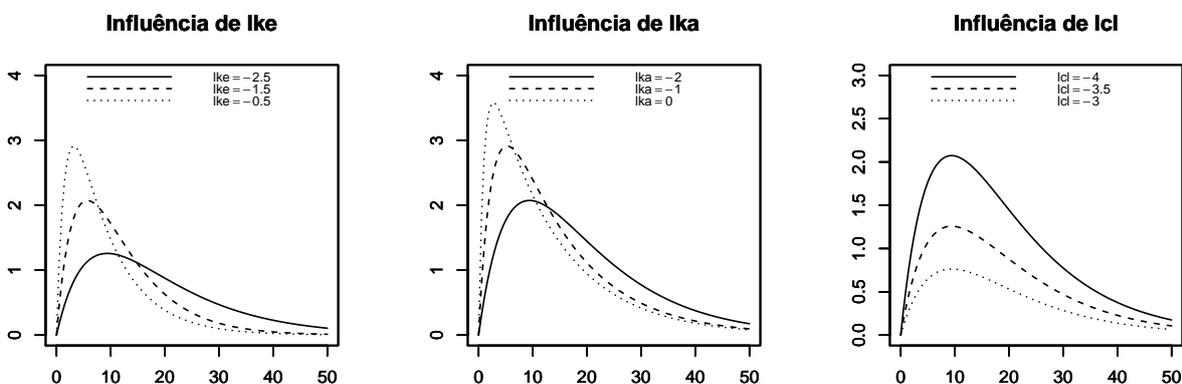


Figura 11 – Interpretação do parâmetro  $lke$ ,  $lka$  e  $lcl$  em um modelo de compartimentos.

Em cada gráfico da Figura 11, os demais parâmetros permaneceram constantes.

Willemsen et al. (WILLEMSSEN *et al.*, 2017) destacaram algumas características da curva quando há alteração nos parâmetros e que podemos constatar nos gráficos da Figura 11. Por exemplo, quanto maior é o valor de  $lke$ , mais acentuado é a diminuição em pontos posteriores e menor a inclinação inicial da curva, que está em concordância com o aumento da taxa de eliminação da substância. Com o aumento no valor de  $lka$  torna a inclinação inicial mais íngreme, pois a taxa de absorção aumenta e com a redução do valor de  $lcl$ , a curva é esticada para cima, ou seja, como a depuração plasmática é a medida da capacidade do organismo em eliminar um fármaco, se o  $lcl$  é pequeno então a absorção é maior.

Os dados estão apresentados na Figura 12 também podem ser encontrados no biblioteca *nlme* (PINHEIRO *et al.*, 2018) com o nome *Theoph* (*datasets*).

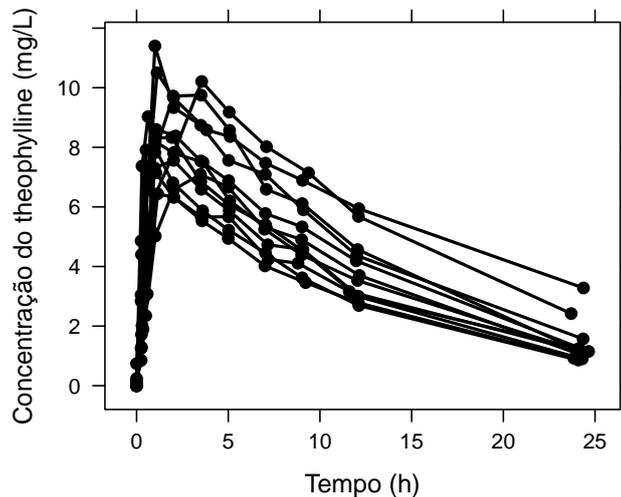


Figura 12 – Concentração de *theophylline* contra o tempo após a administração.

As medidas observadas ao longo do tempo na Figura 12, são diferentes para cada indivíduo. Ajustando os dados com um modelo não linear considerando efeitos aleatórios associados aos parâmetros  $lka$  e  $lcl$  e a distribuições normal para a variável resposta, teremos o seguinte modelo

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 12, \quad (2.4)$$

em que  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) = (f(\boldsymbol{\alpha}, t_{i1}), \dots, f(\boldsymbol{\alpha}, t_{i11}))^\top$ ,  $\mathbf{Z}_i$ , de dimensão  $11 \times 2$ , é uma matriz de constantes conhecidas, e  $\boldsymbol{\varepsilon}_i \stackrel{\text{ind.}}{\sim} N_{11}(\mathbf{0}, \sigma^2 \mathbf{I}_{11})$  e  $\mathbf{b}_i \stackrel{\text{i.i.d.}}{\sim} N_2(\mathbf{0}, \mathbf{D})$ ,  $i = 1, \dots, 12$ , são os erros e os efeitos aleatórios, respectivamente, e são não correlacionados, com  $\mathbf{D}$  não estruturada. Observe que o modelo apresentado é não linear nos efeitos fixos, mas linear nos efeitos aleatórios.

A Figura 13 mostra os valores observados, em cinza, e ajustados, em preto, para cada indivíduo com os dados de *theophylline*.

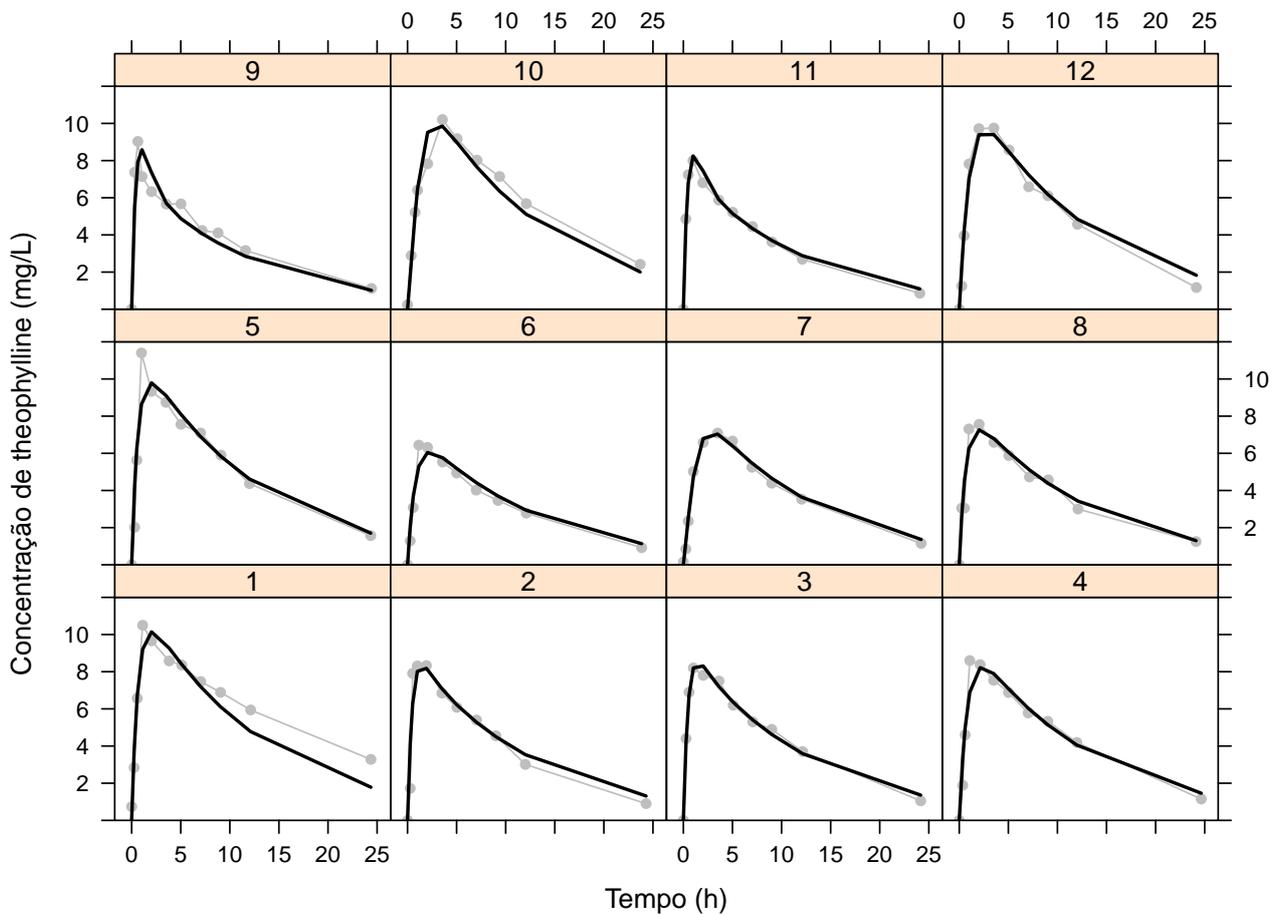
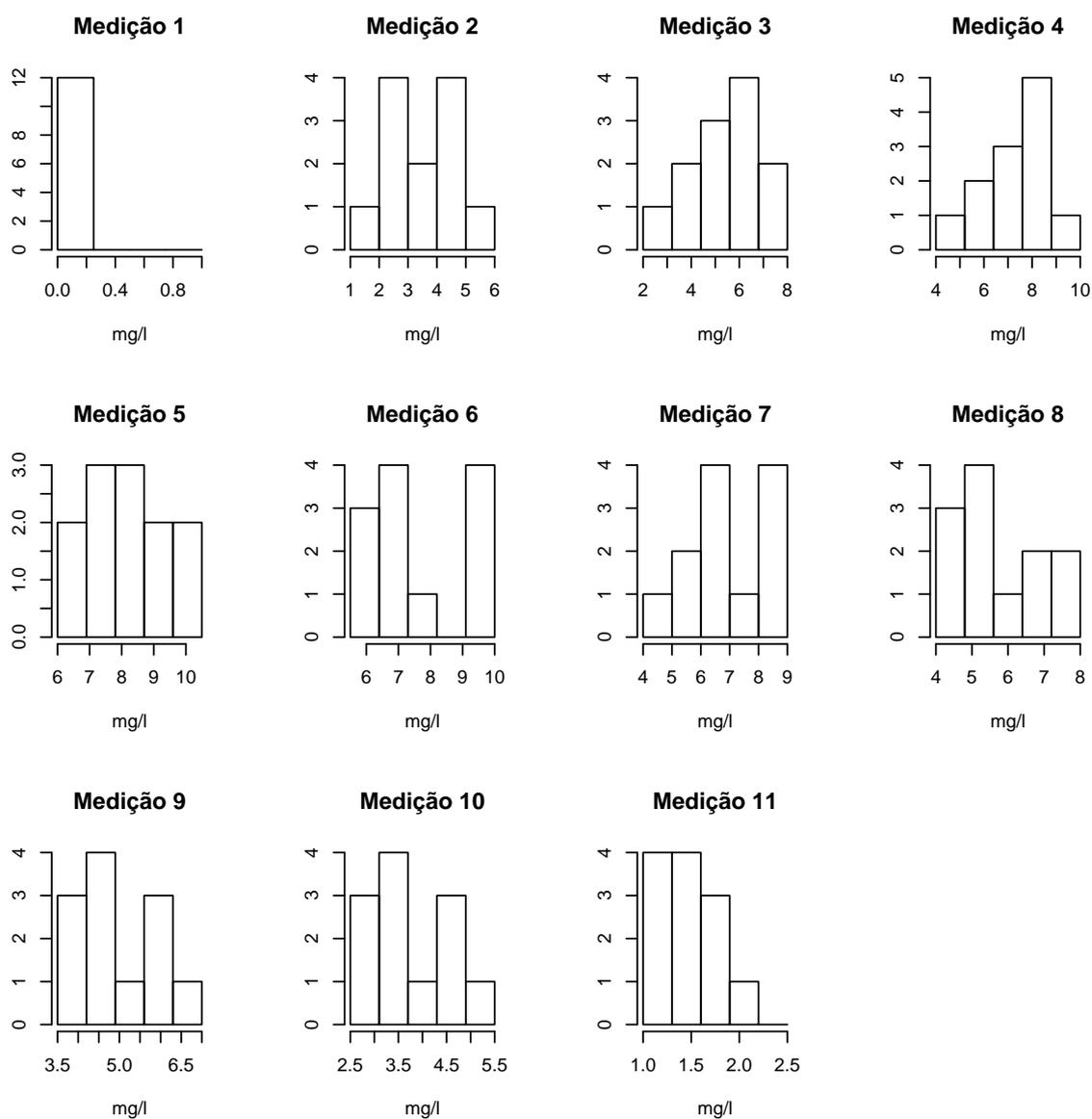


Figura 13 – Perfis ajustados sob a distribuição normal por indivíduo dos dados de *theophylline*.

Os histogramas apresentados na Figura 14 se referem aos valores preditos da variável resposta por medição em cada um dos 11 tempos, com os ajustes em que as componentes aleatórias seguem distribuições normais multivariadas, conforme modelo (2.4). Verificamos que a distribuição do efeito "indivíduo" tem uma tendência a seguir uma distribuição assimétrica.

Figura 14 – Histogramas dos valores preditos da variável resposta por medição dos dados de *theophylline*.



---

## FAMÍLIA DE DISTRIBUIÇÕES MISTURAS DE ESCALA SKEW-NORMAL

---

Distribuições derivadas de misturas de escala da distribuição normal são comuns na literatura e bastante utilizadas em procedimentos estatísticos. Mistura de escala da distribuição normal (ANDREWS; MALLOWS, 1974) fornece um grupo de distribuições com caudas pesadas, tais como a Student-t, slash, normal contaminada e exponencial potência, que são frequentemente utilizadas para a inferência robusta de dados simétricos. Porém, para analisar conjuntos de dados assimétricos essas distribuições tornam-se inadequadas, surgindo a necessidade de se obter distribuições que levem em consideração a assimetria. Nesse sentido, pesquisadores introduziram famílias de distribuições paramétricas flexíveis que acomodem desvios de normalidade e, assim, amenizem a necessidade de transformação de dados.

Famílias de distribuições assimétricas são bastante úteis para analisar dados com observações destoantes ou atípicas levando-se em consideração a assimetria, devido a flexibilidade para acomodar, inclusive, dados simétricos com caudas leves e pesadas, com uma variedade de distribuições tais como a skew-normal, skew-Student-t, skew-slash, skew-normal contaminada, podendo ser útil para a modelagem de dados, análise estatística e estudos de métodos robustos, além da facilidade de geração de dados artificiais e da estimação de parâmetros pelo método de máxima verossimilhança via algoritmos EM genuínos.

A família de distribuições SMSN apresentada por Branco e Dey (BRANCO; DEY, 2001) e Zeller et al. (ZELLER; LACHOS; VILCA-LABRA, 2011) é uma extensão da classe de distribuições SMN que combina assimetria com distribuições com caudas pesadas. De acordo com Arellano-Valle e Genton (ARELLANO-VALLE; GENTON, 2005) e Zeller (ZELLER, 2009) existem algumas definições da distribuição skew-normal e a definição apresentada a seguir será considerada na construção da classe SMSN.

Um vetor aleatório tem distribuição skew-normal multivariada,  $\mathbf{W} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ ,  $\mathbf{w} \in$

$\mathbb{R}^p$ , em que o vetor de locação é  $\boldsymbol{\mu} \in \mathbb{R}^p$ , matriz escala  $\boldsymbol{\Sigma}$  positiva definida de dimensão  $p \times p$  e vetor de assimetria  $\boldsymbol{\lambda} \in \mathbb{R}^p$ , tem função densidade de probabilidade (fdp) dada por

$$g(\mathbf{w}) = 2\phi_p(\mathbf{w}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(\boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{w} - \boldsymbol{\mu})), \quad (3.1)$$

em que  $\phi_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma})$  é a fdp da distribuição normal  $p$ -variada,  $\Phi(\cdot)$  é a função de distribuição acumulada (fda) da distribuição normal padrão univariada e  $\boldsymbol{\Sigma}^{-1/2}$  é tal que  $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$ . A fdp apresentada em (3.1) é semelhante ao modelo apresentado por Azzalini e Valle (AZZALINI; VALLE, 1996), o que difere é uma reparametrização. Quando  $\boldsymbol{\mu} = \mathbf{0}$  e  $\boldsymbol{\Sigma} = \mathbf{I}_p$ , denotamos por  $\mathbf{W} \sim SN_p(\boldsymbol{\lambda})$  a distribuição skew-normal multivariada padrão com a seguinte fdp

$$g(\mathbf{w}) = 2\phi_p(\mathbf{w})\Phi(\boldsymbol{\lambda}^\top \mathbf{w}). \quad (3.2)$$

Na Figura 15 são apresentadas densidades da distribuição skew-normal padrão univariada para diferentes valores do parâmetro  $\lambda$ .

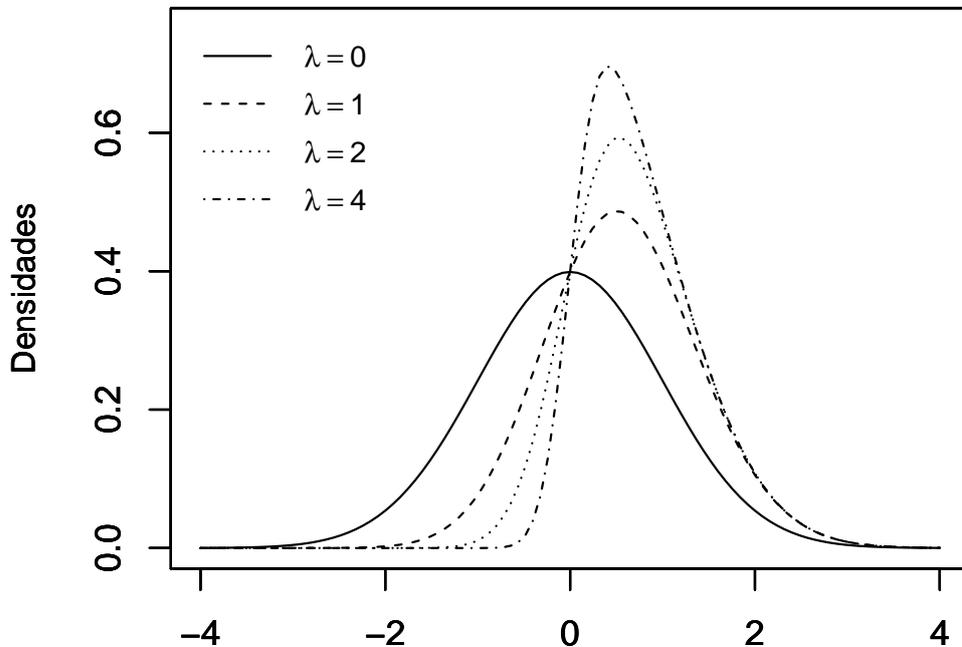


Figura 15 – Densidades da distribuição skew-normal padrão univariada com diferentes valores para o parâmetro de assimetria.

A representação estocástica de um vetor aleatório skew-normal padronizado,  $\mathbf{W} \sim SN_p(\boldsymbol{\lambda})$ , é dado por

$$\mathbf{W} \stackrel{d}{=} \boldsymbol{\delta}|T_0| + (\mathbf{I}_p - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1, \quad (3.3)$$

em que  $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}$ ,  $|T_0|$  representa o valor absoluto de  $T_0$ , com  $T_0 \sim N(0,1)$  e  $\mathbf{T}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p)$  são independentes.

Se  $\mathbf{X} \sim SN_p(\boldsymbol{\lambda})$  e considerando a transformação linear  $\mathbf{W} = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{X}$ , em que  $\boldsymbol{\Sigma}$  é positiva definida, então

$$\mathbf{W} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}). \quad (3.4)$$

Podemos construir a família SMSN de distribuições fazendo  $\mathbf{Z} = \mathbf{W} - \boldsymbol{\mu} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ , em que  $\mathbf{W} \sim SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ , e assim  $a\mathbf{Z} \sim SN_p(\mathbf{0}, a^2\boldsymbol{\Sigma}, \boldsymbol{\lambda})$ , com  $a > 0$  e considerando a construção da classe de distribuições mistura de escala normal/independente estudada por autores como Andrews e Mallows (ANDREWS; MALLOWS, 1974) e Lange e Sinsheimer (LANGE; SINSHEIMER, 1993). Um vetor aleatório com distribuição pertencente à classe SMSN é construído por meio da mistura de escala de uma variável aleatória skew-normal e uma variável aleatória positiva, e possui a forma

$$\mathbf{Y} = \boldsymbol{\mu} + \kappa^{1/2}(U)\mathbf{Z}, \quad (3.5)$$

em que  $\mathbf{Z} \sim SN_p(\mathbf{0}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ ,  $\kappa(\cdot)$  é uma função de pesos,  $U$  é uma variável aleatória positiva com fda  $H(u; \mathbf{v})$  e fdp  $h(u; \mathbf{v})$ , independente de  $\mathbf{Z}$ , indexada por um vetor de parâmetros  $\mathbf{v}$ .

Definimos a seguir, de acordo com Zeller et al. (ZELLER; LACHOS; VILCA-LABRA, 2011), a fdp de um vetor aleatório pertence à classe de distribuições SMSN.

Seja o vetor aleatório  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ , então sua fdp, integrando  $u$ , é dada por

$$\begin{aligned} g(\mathbf{y}) &= 2 \int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})\Phi(\kappa^{-1/2}(u)A)dH(u; \mathbf{v}) \\ &= 2 \int_0^\infty \frac{\kappa^{-p/2}(u)}{(2\pi)^{p/2}} |\boldsymbol{\Sigma}|^{-1/2} e^{-\kappa^{-1}(u)\frac{d}{2}} \Phi(\kappa^{-1/2}(u)A)dH(u; \mathbf{v}), \quad \mathbf{y} \in \mathbb{R}^p, \end{aligned} \quad (3.6)$$

em que  $d = (\mathbf{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\mu})$  é a distância de Mahalanobis e  $A = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$ .

Zeller (ZELLER, 2009) destaca algumas propriedades da classe SMSN de distribuições, tais como

- (i) permite combinar assimetria com caudas pesadas,
- (ii) admite uma representação estocástica que permite uma fácil implementação do algoritmo do tipo EM e estudos de suas propriedades,
- (iii) permite acomodações de observações atípicas e a obtenção de procedimentos estatísticos robustos combinando assimetria com caudas pesadas e
- (iv) permite estender modelos desenvolvidos sob normalidade ou sob outras distribuições simétricas.

Um vetor aleatório  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$  tem vetor de parâmetros de assimetria nulo, ou seja,  $\boldsymbol{\lambda} = \mathbf{0}$ , então  $\mathbf{Y} \sim SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$ , com função densidade de probabilidade dada por

$$\begin{aligned} g(\mathbf{y}) &= \int_0^\infty \phi_p(\mathbf{y}|\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma})dH(u; \mathbf{v}) \\ &= \int_0^\infty \frac{\kappa^{-p/2}(u)}{(2\pi)^{p/2}} |\boldsymbol{\Sigma}|^{-1/2} e^{-\kappa^{-1}(u)\frac{d}{2}} dH(u; \mathbf{v}). \end{aligned} \quad (3.7)$$

Além disso, se a função de pesos  $\kappa(u) = 1/u$ , a família SMSN de distribuições se reduz à classe de distribuições SNI discutida em Lachos et al. (LACHOS; GHOSH; ARELLANO-VALLE, 2010) e em Cabral et al. (CABRAL; LACHOS; PRATES, 2012).

A família de distribuições SMSN introduzida por Branco e Dey (BRANCO; DEY, 2001) é derivada da representação estocástica de uma variável aleatória skew-normal, como apresentada em Azzalini e Capitanio (AZZALINI; CAPITANIO, 1999). Ferreira et al. (FERREIRA; BOLFARINE; LACHOS, 2011) desenvolveram uma classe de distribuições *skew-scale mixtures of normal* (SSMN), cuja construção é baseada na fdp da família SMN de distribuições.

A seguir são apresentados alguns casos triviais da família de distribuições SMSN

- quando  $\boldsymbol{\mu} = \mathbf{0}$  e  $\boldsymbol{\Sigma} = \mathbf{I}_p$  é uma matriz identidade, teremos a distribuição  $SMSN_p(\boldsymbol{\lambda}; H)$  (SMSN padrão),
- quando  $\boldsymbol{\lambda} = \mathbf{0}$ , teremos a distribuição  $SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$  (classe SMN),
- quando  $\kappa(u) = 1$ , teremos a distribuição  $SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$  (skew-normal),
- quando  $\boldsymbol{\lambda} = \mathbf{0}$  e  $\kappa(u) = 1$ , teremos a distribuição  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  (normal),
- quando  $\boldsymbol{\mu} = \boldsymbol{\lambda} = \mathbf{0}$ ,  $\kappa(u) = 1$  e  $\boldsymbol{\Sigma} = \mathbf{I}_p$  é uma matriz identidade, teremos a distribuição  $N_p(\mathbf{0}, \mathbf{I}_p)$  (normal padrão).

Na família SMSN, como apresentada em (3.5), se a variável aleatória  $U$  é conhecida, então a distribuição condicional de  $\mathbf{Y}$  dado  $U = u$ , segue uma distribuição  $SN_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}, \boldsymbol{\lambda})$ , como verificamos a seguir.

Seja  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ , então

$$\mathbf{Y}|U = u \stackrel{\text{ind.}}{\sim} SN_p(\boldsymbol{\mu}, \kappa(u)\boldsymbol{\Sigma}, \boldsymbol{\lambda}). \quad (3.8)$$

A representação estocástica de um vetor aleatório com fdp definida em (3.6), de acordo com Zeller et al. (ZELLER; LACHOS; VILCA-LABRA, 2011), é dada por

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \kappa^{1/2}(u)\boldsymbol{\Sigma}^{1/2}\{\boldsymbol{\delta}|T_0| + (\mathbf{I}_n - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{T}_1\}, \quad (3.9)$$

com  $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}$ , em que  $|T_0|$  representa o valor absoluto de  $T_0$  e as variáveis aleatórias  $U \sim H(\cdot; \mathbf{v})$ ,  $T_0 \sim N(0, 1)$  e  $\mathbf{T}_1 \sim N_p(\mathbf{0}, \mathbf{I}_p)$  são todas independentes.

Alternativamente, teremos a representação estocástica

$$\mathbf{Y} \stackrel{d}{=} \boldsymbol{\mu} + \kappa^{1/2}(u)\{\boldsymbol{\Delta}t + \boldsymbol{\Gamma}^{1/2}\mathbf{T}_1\}, \quad (3.10)$$

em que  $\boldsymbol{\Delta} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\delta}$ , com  $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$ ,  $t = |T_0|$  e  $\boldsymbol{\Gamma} = \boldsymbol{\Sigma} - \boldsymbol{\Delta}\boldsymbol{\Delta}^\top$ .

Esta representação estocástica facilita a implementação do algoritmo tipo EM na estimação dos parâmetros do modelo pelo método de máxima verossimilhança.

### 3.1 Momentos

A função geradora de momentos (fgm) de um vetor aleatório com distribuição pertencente à classe SMSN de distribuições (BASSO, 2009), é apresentada a seguir.

Seja  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ , então sua fgm é dada por

$$M_{\mathbf{Y}}(\mathbf{s}) = E[e^{\mathbf{s}^\top\mathbf{Y}}] = \int_0^\infty 2e^{\mathbf{s}^\top\boldsymbol{\mu} + \frac{1}{2}\kappa(u)\mathbf{s}^\top\boldsymbol{\Sigma}\mathbf{s}}\Phi(\kappa^{1/2}(u)\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^{1/2}\mathbf{s})dH(u; \mathbf{v}), \quad \mathbf{s} \in \mathbb{R}^p. \quad (3.11)$$

Com a expressão (3.11) podemos derivar, segundo Basso (BASSO, 2009) e Zeller (ZELLER, 2009), o vetor de médias, a matriz de covariâncias e o coeficiente de curtose multidimensional  $\gamma_2(\mathbf{Y})$  de um vetor aleatório com distribuição pertencente a família SMSN.

Seja  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ , então

$$a) \text{ se } E[\kappa^{1/2}(U)] < \infty, \text{ então } E(\mathbf{Y}) = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}} E[\kappa^{1/2}(U)] \boldsymbol{\Delta}, \quad (3.12)$$

$$b) \text{ se } E[\kappa(U)] < \infty, \text{ então } Var(\mathbf{Y}) = E[\kappa(U)] \boldsymbol{\Sigma} - \frac{2}{\pi} E^2[\kappa^{1/2}(U)] \boldsymbol{\Delta} \boldsymbol{\Delta}^\top \text{ e} \quad (3.13)$$

$$c) \text{ se } E[\kappa^2(U)] < \infty, \text{ então } \gamma_2(\mathbf{Y}) = k_1 a_{1y} - 4k_2 a_{2y} + a_{3y} - p(p+2), \quad (3.14)$$

onde  $k_1 = \frac{E[\kappa^2(U)]}{E^2[\kappa(U)]}$ ,  $k_2 = \frac{E[\kappa^{3/2}(U)]}{E^2[\kappa(U)]}$ ,  $a_{1y} = p(p+2) + 2(p+2) \boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y + 3(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y)^2$ ,  $a_{2y} = \left(p + \frac{2}{E[\kappa^{1/2}(U)]}\right) \boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y + \left(1 + \frac{2}{E[\kappa^{1/2}(U)]} - \frac{\pi E[\kappa(U)]}{2E^2[\kappa^{1/2}(U)]}\right) (\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y)^2$  e  $a_{3y} = 2(p+2) \boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y + 3(\boldsymbol{\mu}_y^\top \boldsymbol{\Sigma}_y^{-1} \boldsymbol{\mu}_y)^2$ , com  $\boldsymbol{\mu}_y = E[\mathbf{Y} - \boldsymbol{\mu}] = \sqrt{\frac{2}{\pi}} E[\kappa^{1/2}(U)] \boldsymbol{\Delta}$ .

Em Lachos et al. (LACHOS; GHOSH; ARELLANO-VALLE, 2010) e Zeller et al. (ZELLER; LACHOS; VILCA-LABRA, 2011) são apresentados os momentos condicionais de funções  $U$  dado  $\mathbf{y}$ , denotados por  $u_r$  e  $\tau_r$ , como apresentado a seguir, que também são essenciais para implementação do algoritmo tipo EM.

Se  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$  e  $U \sim H$  é a variável aleatória de mistura de escala, então

$$u_r = E[\kappa^{-r}(U)|\mathbf{y}] = \frac{2g_0(\mathbf{y})}{g(\mathbf{y})} E[\kappa^{-r}(U_y) \Phi(\kappa^{-1/2}(U_y)A)] \text{ e} \quad (3.15)$$

$$\tau_r = E[\kappa^{-r/2}(U) W_\Phi(\kappa^{-1/2}(U)A)|\mathbf{y}] = \frac{2g_0(\mathbf{y})}{g(\mathbf{y})} E[\kappa^{-r/2}(U_y) \phi_1(\kappa^{-1/2}(U_y)A)], \quad (3.16)$$

em que  $g_0$  é a fdp de  $\mathbf{Y}_0 \sim SMN(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$ ,  $U_y \stackrel{d}{=} (U|\mathbf{Y}_0 = \mathbf{y})$ ,  $W_\Phi(x) = \phi_1(x)/\Phi(x)$ ,  $x \in \mathbb{R}$  e  $A = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$ .

## 3.2 Distribuição marginal e independência

Os resultados a seguir mostram que um vetor aleatório com distribuição na família SMSN é invariante quanto a transformações lineares, assim, a distribuição marginal desse vetor também pertence à família SMSN (ZELLER, 2009; BASSO, 2009). Seja  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ , então para qualquer vetor fixo  $\mathbf{b} \in \mathbb{R}^m$  e qualquer matriz  $\mathbf{B} \in \mathbb{R}^{m \times p}$  de posto completo, teremos

$$\mathbf{V} = \mathbf{b} + \mathbf{B}\mathbf{Y} \sim SMSN_m(\mathbf{b} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top, \boldsymbol{\lambda}^*; H),$$

onde  $\boldsymbol{\lambda}^* = \frac{\boldsymbol{\delta}^*}{\sqrt{1 - \boldsymbol{\delta}^{*\top} \boldsymbol{\delta}^*}}$ , com  $\boldsymbol{\delta}^* = (\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^\top)^{-1/2} \mathbf{B}\boldsymbol{\Sigma}^{1/2} \boldsymbol{\delta}$ . Além disso, se  $m = p$  e a matriz  $\mathbf{B}$  é não singular, então  $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}$ .

Como resultado adicional temos que se  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$  está particionado como  $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$  com dimensões  $p_1$  e  $p_2$ , respectivamente, tal que  $p = p_1 + p_2$ . Particionando  $\boldsymbol{\Sigma}$ ,  $\boldsymbol{\mu}$  e  $\boldsymbol{\lambda}$  de acordo com a partição correspondente de  $\mathbf{Y} = (\mathbf{Y}_1^\top, \mathbf{Y}_2^\top)^\top$ , teremos

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} \quad \text{e} \quad \boldsymbol{\lambda} = \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{pmatrix}. \quad (3.17)$$

Assim, a fdp marginal de  $\mathbf{Y}_1$  é  $SMSN_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{11}^{1/2} \tilde{\mathbf{v}}; H)$ , com  $\tilde{\mathbf{v}} = \frac{\mathbf{v}_1 + \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12} \mathbf{v}_2}{\sqrt{1 + \mathbf{v}_2^\top \boldsymbol{\Sigma}_{22.1} \mathbf{v}_2}}$ , onde  $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}$  e  $\mathbf{v} = \boldsymbol{\Sigma}^{-1/2} \boldsymbol{\lambda} = (\mathbf{v}_1^\top, \mathbf{v}_2^\top)^\top$ .

O resultado a seguir é bastante útil em aplicações com modelos lineares e não lineares com efeitos mistos. Por exemplo, quando um modelo linear ou não linear depende de um vetor não observável de efeitos aleatórios e de um vetor de erros aleatórios, onde os efeitos aleatórios seguem uma distribuição na família SMSN e os erros seguem uma distribuição na família SMN.

Utilizando as notações definidas nos resultados anteriores, se  $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21} = \mathbf{0}$  e  $\boldsymbol{\lambda}_2 = \mathbf{0}$ , então dado  $U = u$ ,  $\mathbf{Y}_1$  e  $\mathbf{Y}_2$  são independentes. Adicionalmente, se  $\boldsymbol{\mu}_2 = \mathbf{0}$ ,  $\mathbf{Y}_1$  e  $\mathbf{Y}_2$  são não correlacionados.

Os momentos condicionais de  $\mathbf{Y}_2$  dado  $\mathbf{Y}_1 = \mathbf{y}_1$  para a distribuição skew-normal foram desenvolvidos por Lin e Lee (LIN; LEE, 2008). A seguir serão apresentados extensões desses resultados para a família SMSN de distribuições.

Considerando a partição definida em (3.17), se  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ , então

a) o primeiro momento de  $\mathbf{Y}_2$  dado  $\mathbf{Y}_1 = \mathbf{y}_1$  e  $U = u$  é obtido por

$$E[\mathbf{Y}_2 | \mathbf{y}_1, u] = \boldsymbol{\mu}_{2.1} + \kappa^{1/2}(u) W_\Phi(\kappa^{-1/2}(u) \tilde{\mathbf{v}}^\top (\mathbf{y}_1 - \boldsymbol{\mu}_1)) \boldsymbol{\rho},$$

onde  $\boldsymbol{\mu}_{2.1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{y}_1 - \boldsymbol{\mu}_1)$  e  $\boldsymbol{\rho} = \frac{\boldsymbol{\Sigma}_{22.1} \mathbf{v}_2}{\sqrt{1 + \mathbf{v}_2^\top \boldsymbol{\Sigma}_{22.1} \mathbf{v}_2}}$ .

b) o segundo momento de  $\mathbf{Y}_2$  dado  $\mathbf{Y}_1 = \mathbf{y}_1$  e  $U = u$  é obtido por

$$\begin{aligned} E[\mathbf{Y}_2 \mathbf{Y}_2^\top | \mathbf{y}_1, u] &= \boldsymbol{\mu}_{2.1} \boldsymbol{\mu}_{2.1}^\top + \kappa(u) \boldsymbol{\Sigma}_{22.1} \\ &+ \kappa^{1/2}(u) W_\Phi(\kappa^{-1/2}(u) \tilde{\mathbf{v}}^\top (\mathbf{y}_1 - \boldsymbol{\mu}_1)) (\boldsymbol{\mu}_{2.1} \boldsymbol{\rho}^\top + \boldsymbol{\rho} \boldsymbol{\mu}_{2.1}^\top - \boldsymbol{\rho} \boldsymbol{\rho}^\top). \end{aligned}$$

O resultado a seguir é útil em aplicações em que os dados não são completamente observados.

Se  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ , então o primeiro momento de  $\mathbf{Y}_2$  condicionado a  $\mathbf{Y}_1 = \mathbf{y}_1$  é obtido por

$$E[\mathbf{Y}_2 | \mathbf{y}_1] = \boldsymbol{\mu}_{2.1} + \frac{\boldsymbol{\Sigma}_{22.1} \mathbf{v}_2}{\sqrt{1 + \mathbf{v}_2^\top \boldsymbol{\Sigma}_{22.1} \mathbf{v}_2}} \boldsymbol{\tau}_{-1}, \quad (3.18)$$

em que  $\tau_{-1}$  é definido em (3.16).

### 3.3 Casos particulares

A seguir apresentaremos alguns casos especiais de distribuições na classe SMSN derivados de (3.6), cujas propriedades foram discutidas em Lachos et al. (LACHOS; GHOSH; ARELLANO-VALLE, 2010), e seus respectivos momentos condicionais  $u_r$  e  $\tau_r$ . Em todos os casos consideramos  $\kappa(u) = 1/u$  e  $A = \boldsymbol{\lambda}^\top \boldsymbol{\Sigma}^{-1/2}(\mathbf{y} - \boldsymbol{\mu})$

#### • Distribuição skew-normal multivariada

A distribuição skew-normal multivariada (AZZALINI; CAPITANIO, 1999), denotada por  $SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ , é obtida quando  $\kappa(u) = 1/u$  e  $U = 1$  com a seguinte fdp

$$g(\mathbf{y}) = 2\phi_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\Phi(A), \mathbf{y} \in \mathbb{R}^p. \quad (3.19)$$

O vetor de médias e a matriz de covariâncias da distribuição skew-normal são dados por

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\pi}}\boldsymbol{\Delta} \quad e \quad (3.20)$$

$$Var[\mathbf{Y}] = \boldsymbol{\Sigma} - \frac{2}{\pi}\boldsymbol{\Delta}\boldsymbol{\Delta}^\top. \quad (3.21)$$

As expressões dos momentos condicionais para a distribuição SN são

$$\begin{aligned} u_r &= 1 \quad e \\ \tau_r &= \frac{\phi(A)}{\Phi(A)}. \end{aligned}$$

Propriedades da distribuição skew-normal foram estudadas e estão apresentadas em Azzalini e Capitanio (AZZALINI; CAPITANIO, 1999), e Arellano-Valle e Genton (ARELLANO-VALLE; GENTON, 2005).

#### • Distribuição skew-Student-t multivariada

A distribuição ST (ZELLER; LACHOS; VILCA-LABRA, 2011; BRANCO; DEY, 2001) com  $\nu$  graus de liberdade representada por  $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ , é obtida quando  $\kappa(u) = 1/u$  e  $U \sim \text{Gamma}(\frac{\nu}{2}, \frac{\nu}{2})$ ,  $u > 0$ ,  $\nu > 0$ . A fdp de  $\mathbf{Y}$  é dada por

$$g(\mathbf{y}) = 2t_p(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)T\left(\sqrt{\frac{p+\nu}{d+\nu}}A; \nu+p\right), \mathbf{y} \in \mathbb{R}^p \quad (3.22)$$

em que  $t_p(\cdot|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$  e  $T(\cdot; \nu)$  são respectivamente fdp  $p$ -variada e a fda univariada da distribuição Student-t. Um caso particular da distribuição skew-Student-t é a distribuição skew-Cauchy, quando  $\nu = 1$  e também, se  $\nu \uparrow \infty$ , temos a distribuição skew-normal como caso limite. Algumas aplicações da distribuição skew-t em estimações robustas podem ser encontradas em Azzalini e Genton (AZZALINI; GENTON, 2008). Neste caso, o vetor de médias e a matriz de covariâncias são dados por

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{\nu}{\pi}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \boldsymbol{\Delta}, \quad \nu > 1, \quad e \quad (3.23)$$

$$Var[\mathbf{Y}] = \frac{\nu}{\nu-2} \boldsymbol{\Sigma} - \frac{\nu}{\pi} \left( \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \right)^2 \boldsymbol{\Delta} \boldsymbol{\Delta}^\top, \quad \nu > 2. \quad (3.24)$$

Pelo fato de  $U_{\mathbf{y}} \stackrel{d}{=} U | \mathbf{Y}_0 = \mathbf{y} \sim \text{Gamma}(\frac{\nu+p}{2}, \frac{\nu+d}{2})$  as expressões dos momentos condicionais para a distribuição ST são

$$u_r = \frac{g_0(\mathbf{y})}{g(\mathbf{y})} \frac{2^{r+1} \Gamma(\frac{\nu+p+2r}{2}) (\nu+d)^{-r}}{\Gamma(\frac{\nu+p}{2})} T\left(\sqrt{\frac{\nu+p+2r}{\nu+d}} A; \nu+p+2r\right) \quad e$$

$$\tau_r = \frac{g_0(\mathbf{y})}{g(\mathbf{y})} \frac{2^{\frac{r+1}{2}} \Gamma(\frac{\nu+p+r}{2})}{\pi^{1/2} \Gamma(\frac{\nu+p}{2})} \frac{(\nu+d)^{\frac{\nu+p}{2}}}{(\nu+d+A^2)^{\frac{\nu+p+r}{2}}},$$

onde  $\Gamma(x)$ ,  $x > 0$ , é a função gama.

#### • Distribuição skew-slash multivariada

Outro membro da família SMSN de distribuições é a skew-slash multivariada (ZELLER; LACHOS; VILCA-LABRA, 2011) que representamos por  $SSL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu)$ , obtida quando  $\kappa(u) = 1/u$  e  $U \sim \text{Beta}(\nu, 1)$ ,  $\nu > 0$  com fdp de  $\mathbf{Y}$  da forma

$$g(\mathbf{y}) = 2\nu \int_0^1 u^{\nu-1} \phi_p(\mathbf{y}|\boldsymbol{\mu}, u^{-1}\boldsymbol{\Sigma}) \Phi(u^{1/2}A) du, \quad \mathbf{y} \in \mathbb{R}^p. \quad (3.25)$$

A distribuição SSL se reduz a distribuição SN quando  $\nu \uparrow \infty$  e possui as seguintes expressões para o vetor de médias e matriz de covariâncias

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\nu}} \frac{2\nu}{2\nu-1} \boldsymbol{\Delta}, \quad \nu > 1/2, \quad e \quad (3.26)$$

$$Var[\mathbf{Y}] = \frac{\nu}{\nu-1} \boldsymbol{\Sigma} - \frac{2}{\pi} \left( \frac{2\nu}{2\nu-1} \right)^2 \boldsymbol{\Delta} \boldsymbol{\Delta}^\top, \quad \nu > 1. \quad (3.27)$$

Os momentos condicionais para a distribuição skew-slash seguem do fato que  $U_{\mathbf{y}} \sim \text{Gamma}(\frac{2\nu+p+2r}{2}, \frac{d}{2}) \mathbb{I}_{(0,1)}$ , cujas expressões são

$$u_r = \frac{g_0(\mathbf{y})}{g(\mathbf{y})} \frac{2\Gamma(\frac{2\nu+p+2r}{2})}{\Gamma(\frac{2\nu+p}{2})} \left(\frac{2}{d}\right)^r \frac{P_1(\frac{2\nu+p+2r}{2}, \frac{d}{2})}{P_1(\frac{2\nu+p}{2}, \frac{d}{2})} E[\Phi(S^{1/2}A)] \quad e$$

$$\tau_r = \frac{g_0(\mathbf{y})}{g(\mathbf{y})} \frac{2^{\frac{r+1}{2}} \Gamma(\frac{2\nu+p+r}{2})}{\pi^{1/2} \Gamma(\frac{2\nu+p}{2})} \frac{d^{\frac{2\nu+p}{2}}}{(d+A^2)^{\frac{2\nu+p+r}{2}}} \frac{P_1(\frac{2\nu+p+r}{2}, \frac{d+A^2}{2})}{P_1(\frac{2\nu+p}{2}, \frac{d}{2})},$$

onde  $P_1(a, b)$  denota a fda da distribuição  $Gamma(a, b)$  avaliada em 1 e  $S \sim Gamma(\frac{2\nu+p+2r}{2}, \frac{d}{2}) \mathbb{I}_{(0,1)}$ . O valor de  $E[\Phi(S^{1/2}A)]$  pode ser calculado como  $E[\Phi(S_0^{1/2}A)|0 < S_0 < 1]$  com  $S_0 \sim Gamma(\frac{2\nu+p+2r}{2}, \frac{d}{2})$  por meio de uma integral de Monte Carlo, gerando observações independentes  $S_j, j = 1, \dots, n$ , de  $S$  e aproximar a esperança por  $\sum_{j=1}^n \Phi(S_j^{1/2}A)/n$ . Observações da distribuição gamma truncada podem ser geradas usando a biblioteca *Runuran* (LEYDOLD; HÖRMANN, 2017).

#### • Distribuição skew-normal contaminada multivariada

A distribuição skew-normal contaminada multivariada (ZELLER; LACHOS; VILCALABRA, 2011) denotada por  $SNC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu_1, \nu_2)$ , obtida quando  $\kappa(u) = 1/u$  e  $U$  é uma variável aleatória discreta, dado  $\mathbf{v} = (\nu_1, \nu_2)^\top$ , com fdp

$$h(u, \mathbf{v}) = \nu_1 \mathbb{I}_{(u=\nu_2)} + (1 - \nu_1) \mathbb{I}_{(u=1)}, \quad 0 < \nu_1 < 1, \quad 0 < \nu_2 \leq 1. \quad (3.28)$$

Segue que

$$g(\mathbf{y}) = 2 \left\{ \nu_1 \phi_p(\mathbf{y}; \boldsymbol{\mu}, \nu_2^{-1} \boldsymbol{\Sigma}) \Phi(\nu_2^{1/2} A) + (1 - \nu_1) \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi(A) \right\}, \quad \mathbf{y} \in \mathbb{R}^p. \quad (3.29)$$

A distribuição SNC se reduz a distribuição SN quando  $\nu_2 = 1$ . O vetor de médias e a matriz de covariâncias tem a forma

$$E[\mathbf{Y}] = \boldsymbol{\mu} + \sqrt{\frac{2}{\nu_1}} \left( \frac{\nu_1}{\nu_2^{1/2}} + 1 - \nu_1 \right) \boldsymbol{\Delta}, \quad e \quad (3.30)$$

$$Var[\mathbf{Y}] = \left( \frac{\nu_1}{\nu_2} + 1 - \nu_1 \right) \boldsymbol{\Sigma} - \frac{2}{\pi} \left( \frac{\nu_1}{\nu_2^{1/2}} + 1 - \nu_1 \right)^2 \boldsymbol{\Delta} \boldsymbol{\Delta}^\top. \quad (3.31)$$

Os momentos condicionais quando  $U_y$  é uma variável aleatória discreta com função densidade de probabilidade condicional  $h_0(u|\mathbf{y}) = (1/g_0(\mathbf{y})) \{ \nu_1 \phi_p(\mathbf{y}; \boldsymbol{\mu}, \nu_2^{-1} \boldsymbol{\Sigma}) \mathbb{I}_{(u=\nu_2)} + (1 - \nu_1) \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathbb{I}_{(u=1)} \}$  tem as seguintes expressões

$$u_r = \frac{2}{g(\mathbf{y})} \left\{ v_1 v_2^r \phi_p(\mathbf{y}; \boldsymbol{\mu}, v_2^{-1} \boldsymbol{\Sigma}) \Phi(v_2^{1/2} A) + (1 - v_1) \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \Phi(A) \right\} \quad \text{e}$$

$$\tau_r = \frac{2}{g(\mathbf{y})} \left\{ v_1 v_2^r \phi_p(\mathbf{y}; \boldsymbol{\mu}, v_2^{-1} \boldsymbol{\Sigma}) \phi_1(v_2^{1/2} A) + (1 - v_1) \phi_p(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \phi_1(A) \right\}.$$

No Apêndice B são apresentadas as expressões dos elementos da matriz de informação de Fisher observada para as distribuições skew-normal, skew-t, skew-slash e skew-normal contaminada.

Alguns resultados importantes sobre a distância de Mahalanobis (ZELLER, 2009) são apresentados a seguir.

Considere que  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ , então a forma quadrática  $d = (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$  tem a mesma distribuição de  $d_0 = (\mathbf{Y}_0 - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y}_0 - \boldsymbol{\mu})$ , com  $\mathbf{Y}_0 \sim SMN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}; H)$  e

$$E[d^m] = \frac{2^m \Gamma(m + \frac{p}{2})}{\Gamma(\frac{p}{2})} E[\kappa^m(U)]. \quad (3.32)$$

Na Tabela 3 são apresentados alguns resultados relacionados a  $d = (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu})$ , com  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ .

Tabela 3 – Resultados sobre a distância de Mahalanobis por distribuição pertencente à classe SMSN.

Distribuição	resultado
skew-normal	$d \sim \chi_p^2$
skew-t	$d \sim pF(p, \nu)$
skew-slash	$P(d \leq t) = P(\chi_p^2 \leq t) - \frac{2^{\nu} \Gamma(\nu + \frac{p}{2})}{t^{\nu} \Gamma(\frac{p}{2})} P(\chi_{2\nu+p}^2 \leq t)$
skew-normal contaminada	$P(d \leq t) = v_1 P(\chi_p^2 \leq v_2 t) + (1 - v_1) P(\chi_p^2 \leq t)$

Os resultados apresentados na Tabela 3 são importantes para avaliar modelos no ponto de vista de análise de diagnóstico, por exemplo, na construção de envelopes simulados e com gráficos das distâncias de Mahalanobis para identificação de *outliers* (ZELLER, 2009; PINHEIRO; LIU; WU, 2001).

A Figura 16 mostra densidades univariadas construídas a partir de 5000 observações simuladas com o *software R* (R Core Team, 2018). Em todas as distribuições consideramos  $\mu = 0$ ,  $\sigma^2 = 4$  e  $\lambda = 5$ . Na distribuição skew-t ( $\nu = 4$ ), skew-slash ( $\nu = 0,95$ ) e na distribuição skew-normal contaminada ( $\mathbf{v} = (0,85;0,25)^\top$ ).

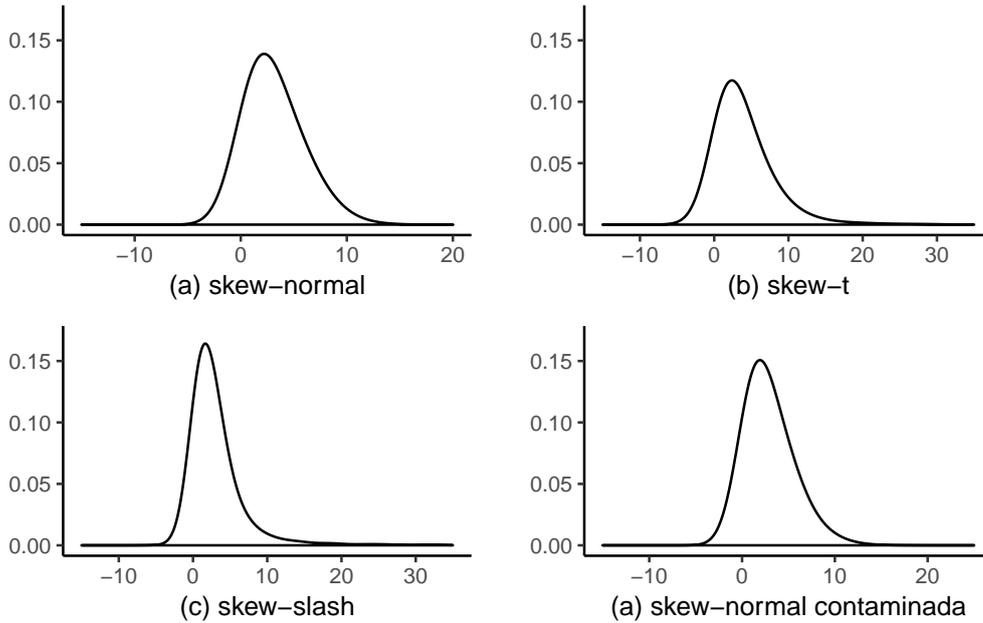


Figura 16 – Densidades de distribuições univariadas pertencentes a família SMSN.

Os gráficos nas Figuras 17-19 mostram contornos das distribuições bivariadas  $SN_2(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda})$ ,  $ST_2(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda}, 4)$  e  $SNC_2(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{\lambda}; \mathbf{v})$ ,  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$  e  $\mathbf{v} = (0,3;0,5)^\top$ , com

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \sigma \\ \sigma & 1 \end{pmatrix}.$$

Os gráficos foram construídos com 1000 observações simuladas por meio do *software R* (R Core Team, 2018) e das bibliotecas *sn* (AZZALINI, 2018), *mixsmsn* (PRATES; CABRAL; LACHOS, 2013) e *ggplot2* (WICKHAM, 2016). As bibliotecas *sn* e *mixsmsn* também foram usadas na construção dos gráficos da Figura 16.

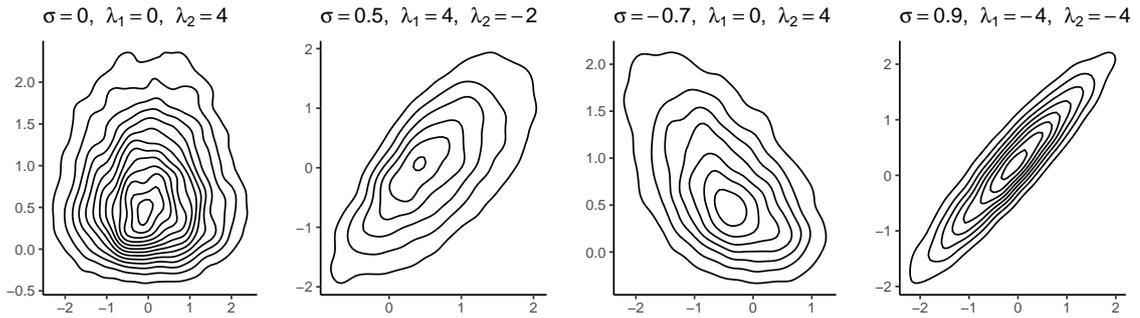


Figura 17 – Contornos da distribuição skew-normal bivariada.

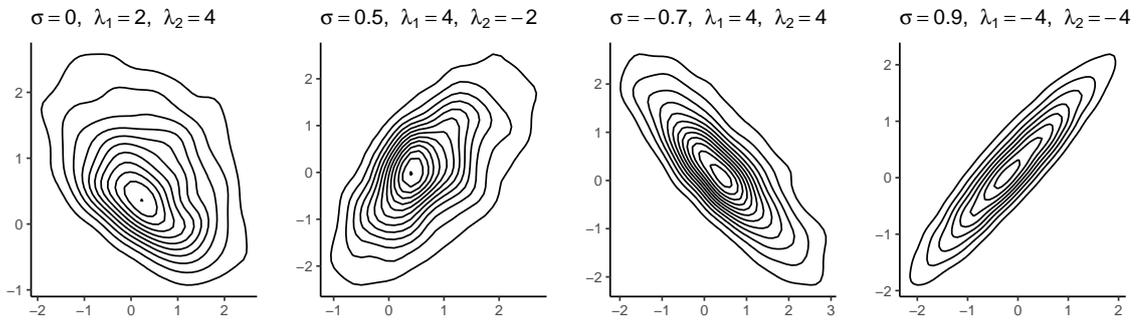


Figura 18 – Contornos da distribuição skew-t bivariada.

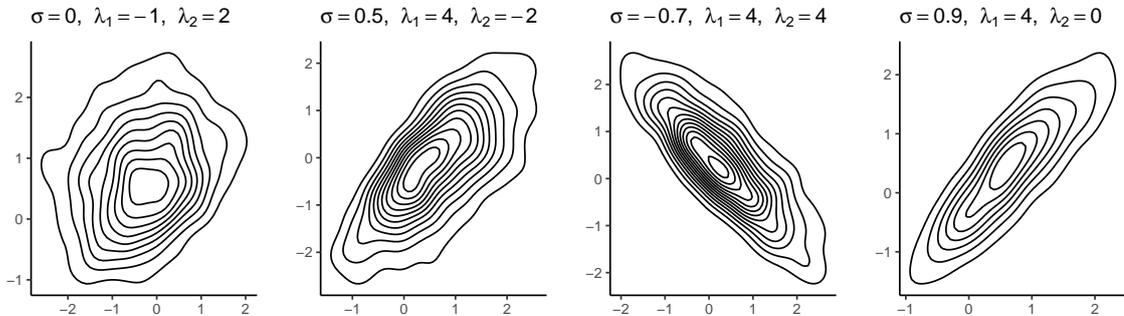


Figura 19 – Contornos da distribuição skew-normal contaminada bivariada.

Apresentamos a seguir alguns resultados relacionados com as propriedades da distribuição normal truncada (JOHNSON; KOTZ; BALAKRISHNAN, 1994; LACHOS, 2004) que são bastante úteis para implementação do algoritmo do tipo EM.

Suponha que  $\mathbf{Y}|T = t \sim N_n(\boldsymbol{\mu} + \Delta t, \boldsymbol{\Psi})$  e  $T$  com distribuição *half*-normal (HN) univariada, caso particular da distribuição normal truncada,  $T \sim HN(0, 1)$ . Seja  $\boldsymbol{\Sigma} = \boldsymbol{\Psi} + \Delta\Delta^\top$ , então a distribuição conjunta de  $(\mathbf{Y}^\top, T)^\top$  pode ser escrita como

$$f_{\mathbf{Y}, T}(\mathbf{y}, t) = 2\phi_n(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})\phi(t|\eta, \tau^2)\mathbb{I}_{(t>0)}, \quad (3.33)$$

em que

$$\eta = \frac{\mathbf{\Delta}^\top \mathbf{\Psi}^{-1}(\mathbf{y} - \boldsymbol{\mu})}{1 + \mathbf{\Delta}^\top \mathbf{\Psi}^{-1} \mathbf{\Delta}} \quad \text{e} \quad (3.34)$$

$$\tau^2 = \frac{1}{1 + \mathbf{\Delta}^\top \mathbf{\Psi}^{-1} \mathbf{\Delta}}. \quad (3.35)$$

Seja  $X \sim N(\eta, \tau^2)$ , então para  $a \in \mathbb{R}$ , temos que

$$E(X|X > a) = \eta + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi\left(\frac{a-\eta}{\tau}\right)} \tau \quad \text{e} \quad (3.36)$$

$$E(X^2|X > a) = \eta^2 + \tau^2 + \frac{\phi\left(\frac{a-\eta}{\tau}\right)}{1 - \Phi\left(\frac{a-\eta}{\tau}\right)} (\eta + a) \tau. \quad (3.37)$$

Seja uma variável aleatória  $T \sim HN(0, 1)$ , então

$$E(T^k|\mathbf{y}) = E(X^k|X > a), \quad (3.38)$$

onde  $X \sim N(\eta, \tau^2)$ , com  $\eta$  e  $\tau^2$  como apresentado em (3.34) e (3.35). Assim,

$$E(T|\mathbf{y}) = \eta + \frac{\phi\left(\frac{\eta}{\tau}\right)}{\Phi\left(\frac{\eta}{\tau}\right)} \tau \quad \text{e} \quad (3.39)$$

$$E(T^2|\mathbf{y}) = \eta^2 + \tau^2 + \frac{\phi\left(\frac{\eta}{\tau}\right)}{\Phi\left(\frac{\eta}{\tau}\right)} \eta \tau. \quad (3.40)$$

### 3.4 Representação hierárquica

Nesta Seção apresentamos uma representação hierárquica que é muito utilizada em modelos sob a família SMSN de distribuições (ZELLER; LACHOS; VILCA-LABRA, 2011). Por meio desta representação podemos tratar o problema de estimação dos parâmetros pelo método de máxima verossimilhança via algoritmo do tipo EM, considerando a abordagem com dados incompletos.

Considere a amostra  $\mathbf{Y}_i \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}; H)$ ,  $i = 1, \dots, n$ , com o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$ , de modo que  $\boldsymbol{\Sigma}$  é positiva definida e depende de parâmetros desconhecidos reduzidos a  $\boldsymbol{\beta}$ , então o modelo hierárquico para cada vetor aleatório SMSN dessa amostra é dado por

$$\mathbf{Y}_i | T = t_i, U = u_i \stackrel{\text{ind.}}{\sim} N_{mi}(\boldsymbol{\mu} + \mathbf{\Delta} t_i, \kappa(u_i) \mathbf{\Psi}), \quad (3.41)$$

$$T_i | U = u_i \stackrel{\text{iid.}}{\sim} HN(0, \kappa(u_i)) \quad \text{e} \quad (3.42)$$

$$U_i \stackrel{\text{iid.}}{\sim} H(u_i; \mathbf{v}), \quad i = 1, \dots, n, \quad (3.43)$$

em que  $HN$  denota a distribuição *half*-normal e a fda  $H$  está indexada pelo vetor  $\mathbf{v}$ , com  $\mathbf{\Delta} = \mathbf{\Sigma}^{1/2}\boldsymbol{\delta}$ ,  $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1+\boldsymbol{\lambda}^\top\boldsymbol{\lambda}}}$ ,  $\boldsymbol{\Psi} = \mathbf{\Sigma} - \mathbf{\Delta}\mathbf{\Delta}^\top$ .

Seja o vetor de respostas observadas  $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$  e considerando  $\mathbf{u} = (u_1, \dots, u_n)^\top$  e  $\mathbf{t} = (t_1, \dots, t_n)^\top$ , onde  $\mathbf{u}$  e  $\mathbf{t}$  são variáveis não observadas, sob a representação hierárquica (3.41)-(3.43), a função de log-verossimilhança completa associada a  $\mathbf{y}_c^\top = (\mathbf{y}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$  é dada por

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{y}_c) &= C + \frac{1}{2} \sum_{i=1}^n \left\{ -\log |\boldsymbol{\Psi}_i| - \kappa^{-1}(u_i)(\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{\Delta}t_i)^\top \boldsymbol{\Psi}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu} - \mathbf{\Delta}t_i) + \log(h(u_i; \mathbf{v})) \right\} \\ &= C + \frac{1}{2} \sum_{i=1}^n \left\{ -\log |\boldsymbol{\Psi}_i| - \kappa^{-1}(u_i)(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Psi}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) \right. \\ &\quad \left. + 2\kappa^{-1}(u_i)t_i(\mathbf{y}_i - \boldsymbol{\mu})^\top \boldsymbol{\Psi}_i^{-1}\mathbf{\Delta} - \kappa^{-1}(u_i)t_i^2\mathbf{\Delta}^\top \boldsymbol{\Psi}_i^{-1}\mathbf{\Delta} + \log(h(u_i; \mathbf{v})) \right\}, \end{aligned} \quad (3.44)$$

onde  $C$  é uma constante que não depende do vetor de parâmetros  $\boldsymbol{\theta}$ .

### 3.5 Algoritmo EM em modelos SMSN

Com a representação hierárquica apresentada na seção anterior, podemos obter as estimativas de máxima verossimilhança dos modelos na classe SMSN via o algoritmo EM, considerando a abordagem com dados aumentados. Apresentamos alguns resultados que serão utilizados posteriormente.

Considerando  $\boldsymbol{\theta}^{(k)} = (\boldsymbol{\mu}^{(k)\top}, \boldsymbol{\beta}^{(k)\top}, \boldsymbol{\lambda}^{(k)\top})^\top$  a estimativa de  $\boldsymbol{\theta}$  na  $k$ -ésima iteração do algoritmo EM, a esperança condicional da função de log-verossimilhança completa é obtida por

$$\begin{aligned} Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) &= E[\ell(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}] \\ &\propto \frac{1}{2} \sum_{i=1}^n \left\{ -\log |\widehat{\boldsymbol{\Psi}}_i^{(k)}| - u^{(k)}(\mathbf{y}_i - \boldsymbol{\mu}^{(k)})^\top \widehat{\boldsymbol{\Psi}}_i^{(k)-1}(\mathbf{y}_i - \boldsymbol{\mu}^{(k)}) \right. \\ &\quad \left. + 2\widehat{u}t_i^{(k)}(\mathbf{y}_i - \boldsymbol{\mu}^{(k)})^\top \widehat{\boldsymbol{\Psi}}_i^{(k)-1}\widehat{\boldsymbol{\Delta}}^{(k)} - \widehat{u}t_i^2{}^{(k)}\widehat{\boldsymbol{\Delta}}^{(k)\top}\widehat{\boldsymbol{\Psi}}_i^{(k)-1}\widehat{\boldsymbol{\Delta}}^{(k)} \right\}. \end{aligned} \quad (3.45)$$

em que  $\widehat{u}_i^{(k)} = E[\kappa^{-1}(u_i)|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}]$ ,  $\widehat{u}t_i^{(k)} = E[\kappa^{-1}(u_i)t_i|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}]$  e  $\widehat{u}t_i^2{}^{(k)} = E[\kappa^{-1}(u_i)t_i^2|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}]$ .

Com os resultados (3.39)-(3.40) relacionados aos momentos da distribuição normal truncada e usando propriedades de esperança condicional (BASSO, 2009), obtemos as expressões

$$\begin{aligned}
ut_i &= E_{U_i|y_i} \{ E_{T_i|u_i, y_i} [\kappa^{-1}(U_i)T_i] \} \\
&= E_{U_i|y_i} \left\{ \kappa^{-1}(U_i) \left[ \mu_{T_i} + W_{\Phi} \left( \frac{\kappa^{-1}(U_i)\mu_{T_i}}{M_{T_i}} \right) \kappa^{1/2}(U_i)M_{T_i} \right] \right\} \\
&= \mu_{T_i} E_{U_i|y_i} \{ \kappa^{-1}(U_i) \} + M_{T_i} E_{U_i|y_i} \left\{ W_{\Phi} \left( \frac{\kappa^{-1}(U_i)\mu_{T_i}}{M_{T_i}} \right) \kappa^{-1/2}(U_i) \right\} \\
&= u_i \mu_{T_i} + M_{T_i} \tau_i, \quad i = 1, \dots, n,
\end{aligned}$$

em que  $M_{T_i}^2 = 1/(1 + \Delta\Psi^{-1}\Delta^{\top})$  e  $\mu_{T_i} = M_{T_i}^2 \Delta^{\top} \Psi^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)$ .

Analogamente temos que

$$ut_i^2 = u_i \mu_{T_i}^2 + M_{T_i}^2 + M_{T_i} \mu_{T_i} \tau_i.$$

Por meio destas expressões podemos derivar os estimadores dos elementos do vetor de parâmetros  $\boldsymbol{\theta}$ .

Uma extensão da família SMSN é a família de distribuições misturas de escala skew-normal estendida (SSMSN), útil para inferência Bayesiana, apresentada por Zeller (ZELLER, 2009), juntamente com um modelo de regressão linear com assimetria, denominado SSMSN-RM.

## MODELO NÃO LINEAR MISTO COM MISTURAS DE ESCALA SKEW-NORMAL

A suposição de normalidade para as componentes aleatórias em modelos estatísticos é bastante comum, porém, como já discutimos anteriormente, esta suposição pode ocasionar a falta de precisão nas estimativas dos parâmetros do modelo. Assim, para minimizar a sensibilidade no processo de estimação dos parâmetros devido a observações atípicas ou assimetria, desenvolvemos modelos não lineares com efeitos mistos em que as componentes aleatórias seguem uma distribuições pertencentes à classe SMSN, que são alternativas para o uso da distribuição normal e outras distribuições simétricas. O modelo SMSN-NLME é bastante flexível, pois, além da possibilidade de ser ajustado sob diversas distribuições, considera ajustes levando-se em conta simetria ou assimetria, efeitos mistos ou somente efeitos fixos. O modelo proposto tem essa característica devido a família SMSN de distribuições, que proporciona versatilidade aos modelos estatísticos. Na representação hierárquica do modelo SMSN-NLME, o vetor aleatório de respostas é condicionado ao efeito aleatório, como iremos verificar em (4.5)-(4.8).

Um modelo não linear assimétrico com efeitos mistos, não linear nos efeitos fixos mas linear nos efeitos aleatórios, assumindo que os erros e os efeitos aleatórios sigam distribuições pertencentes à família SMSN, com o vetor de resposta observadas  $\mathbf{y}_i$ , de dimensão  $m_i \times 1$ ,  $i = 1, \dots, n$ , conforme em Vonesh e Carter (VONESH; CARTER, 1992), e que nós consideramos uma extensão do trabalho de Russo et al. (RUSSO; PAULA; AOKI, 2009), pode ser assim ajustado

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (4.1)$$

em que  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top$  é o vetor de parâmetros desconhecidos,  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)$  é uma função não linear conhecida do vetor de respostas,  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im_i})^\top$  é um vetor de variáveis explanatórias de dimensão  $m_i \times 1$ ,  $\mathbf{Z}_i$  de dimensão  $m_i \times q$ , com  $1 \leq q \leq p$ , é uma matriz de constantes

conhecidas, os efeitos aleatórios  $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^\top$  e os erros  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^\top$  são não correlacionados com  $\mathbf{b}_i \stackrel{\text{i.i.d.}}{\sim} SMSN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}; H)$  e  $\boldsymbol{\varepsilon}_i \stackrel{\text{ind.}}{\sim} SMN_{m_i}(\mathbf{0}, \mathbf{R}_i; H)$ ,  $i = 1, \dots, n$ . As matrizes  $\mathbf{D} = \mathbf{D}(\boldsymbol{\beta})$  e  $\mathbf{R}_i = \mathbf{R}_i(\gamma)$  são matrizes de dispersão positivas definidas e dependem de parâmetros desconhecidos reduzidos a  $\boldsymbol{\beta}$  e  $\gamma$ , respectivamente, em que  $\mathbf{D}$  é não estruturada.

Como podemos notar, nós assumimos que os erros seguem uma distribuição simétrica e os efeitos aleatórios seguem uma distribuição assimétrica, para garantirmos que a resposta  $\mathbf{y}_i$  pertença a família SMSN de distribuições. Com o mesmo argumento, poderíamos considerar que os erros têm distribuição assimétrica e os efeitos aleatórios têm distribuição simétrica, porém, com tal suposição, não teríamos os mesmos resultados, assim vamos estudar esta abordagem posteriormente.

Baseado no modelo (4.1) com o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \gamma, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$ , consideramos o seguinte vetor aleatório

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{pmatrix} \sim SMSN_{q+m_i} \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_i \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda} \\ \mathbf{0} \end{pmatrix}; H \right), \quad i = 1, \dots, n, \quad (4.2)$$

em que  $\boldsymbol{\lambda}$  é o vetor de parâmetros de assimetria do efeito aleatório e  $H = H(\cdot; \mathbf{v})$  é a função de distribuição acumulada que determina a distribuição de probabilidade que o modelo assume, indexado ao vetor de parâmetros  $\mathbf{v}$ .

A distribuição marginal de  $\mathbf{y}_i$  (4.1) associado ao vetor de parâmetros de assimetria  $\boldsymbol{\lambda}_{\mathbf{y}_i}$ , na qual baseamos a inferência sobre o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \gamma, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$ , segue uma distribuição pertence à família SMSN de distribuições com fdp dada por

$$g(\mathbf{y}_i, \boldsymbol{\theta}) = 2 \int_0^\infty \phi_{m_i} \{ \mathbf{y}_i | \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i), u_i^{-1} \boldsymbol{\Sigma}_i \} \Phi \{ u_i^{1/2} \boldsymbol{\lambda}_{\mathbf{y}_i}^\top \boldsymbol{\Sigma}_i^{-1/2} (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)) \} dH(u; \mathbf{v}), \quad (4.3)$$

em que  $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \mathbf{R}_i$ , a função de pesos  $\kappa(u_i) = 1/u_i$  e

$$\boldsymbol{\lambda}_{\mathbf{y}_i} = \frac{\boldsymbol{\Sigma}_i^{-1/2} \mathbf{Z}_i \mathbf{D} \boldsymbol{\zeta}}{\sqrt{1 + \boldsymbol{\zeta}^\top \boldsymbol{\Lambda}_i \boldsymbol{\zeta}}}, \quad \text{com } \boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top \mathbf{R}_i^{-1} \mathbf{Z}_i)^{-1} \quad \text{e } \boldsymbol{\zeta} = \mathbf{D}^{-1/2} \boldsymbol{\lambda},$$

em que a matriz  $\mathbf{D}^{-1/2}$  é tal que  $\mathbf{D}^{-1/2} \mathbf{D}^{-1/2} = \mathbf{D}^{-1}$ .

Este resultado facilita a implementação de inferências com rotinas de otimização em *softwares* estatísticos, como por exemplo o algoritmo do tipo EM que implementamos no *software* R (R Core Team, 2018).

Nos ajustes realizados com o modelo SMSN-NLME, consideramos várias formas para a composição dos elementos da matriz de constantes  $\mathbf{Z}_i$ , como, por exemplo, todos os elementos iguais a um ou assumindo funções das variáveis explanatórias  $\mathbf{x}_i$ , como mostrado a seguir.

- $\mathbf{Z}_i = \mathbf{1} = (1, \dots, 1)^\top$ ,
- $\mathbf{Z}_i = (\mathbf{1}, \mathbf{x}_i)$ , com  $\mathbf{x}_i = (x_{i1}, \dots, x_{im_i})^\top$ ,
- $\mathbf{Z}_i = (\mathbf{1}, \mathbf{x}_i, \mathbf{x}_i^2)$ , com  $\mathbf{x}_i^2 = (x_{i1}^2, \dots, x_{im_i}^2)^\top$  e
- Modelos PNMEM (apresentados a seguir).

Porém, por apresentar menores valores para os critérios AIC (4.33) e BIC (4.34), consideramos a matriz de constantes  $\mathbf{Z}_i$ ,  $i = 1, \dots, n$ , conforme discutido em Vonesh e Carter (VONESH; CARTER, 1992) em *modelos mistos pseudo não lineares* (PNMEM), definida como

$$\mathbf{Z}_i = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_1}, \dots, \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_p} \right] \Big|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}}, \quad i = 1, \dots, n \quad (4.4)$$

com  $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_p)^\top$  representando, por exemplo, a estimativa de mínimos quadrados ordinários de  $\boldsymbol{\alpha}$ , e analogamente foram definidos para os casos mais simples considerando a inclusão de um ou mais efeitos aleatórios.

Assim, obtemos melhores ajustes com a matriz de constantes  $\mathbf{Z}_i$ , sendo a primeira derivada da série de Taylor na expansão de  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)$ , omitindo a segunda derivada e derivadas de ordens superiores.

## 4.1 Estimação por máxima verossimilhança

Maximizar a função de log-verossimilhança derivada do modelo (4.1) para encontrar suas estimativas não é uma tarefa simples, devido a presença de integrais em (4.3). Portanto, para a estimação dos parâmetros no modelo SMSN-NLME vamos utilizar uma solução numérica via algoritmo do tipo EM, que é uma combinação de algoritmos derivados do algoritmo EM e suas extensões e algoritmo de Newton-Raphson. O modelo (4.1), conforme (3.41)-(3.43) e com  $\kappa(u_i) = 1/u_i$ , pode ser representado hierarquicamente como

$$\mathbf{y}_i | \mathbf{b}_i, U_i = u_i \stackrel{\text{ind.}}{\sim} N_{m_i}(\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \mathbf{Z}_i \mathbf{b}_i, u_i^{-1} \mathbf{R}_i), \quad (4.5)$$

$$\mathbf{b}_i | T_i = t_i, U_i = u_i \stackrel{\text{ind.}}{\sim} N_q(\boldsymbol{\Delta} t_i, u_i^{-1} \boldsymbol{\Gamma}), \quad (4.6)$$

$$T_i | U_i = u_i \stackrel{\text{ind.}}{\sim} HN(0, u_i^{-1}) \quad \text{e} \quad (4.7)$$

$$U_i \stackrel{\text{i.i.d.}}{\sim} h(u_i; \mathbf{v}), \quad i = 1, \dots, n, \quad (4.8)$$

em que  $HN$  denota a distribuição *half*-normal univariada,  $\boldsymbol{\Delta} = \mathbf{D}^{1/2} \boldsymbol{\delta}$ ,  $\boldsymbol{\delta} = \frac{\boldsymbol{\lambda}}{\sqrt{1 + \boldsymbol{\lambda}^\top \boldsymbol{\lambda}}}$  e  $\boldsymbol{\Gamma} = \mathbf{D} - \boldsymbol{\Delta} \boldsymbol{\Delta}^\top$ .

Seja  $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$ ,  $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top)^\top$ ,  $\mathbf{u} = (u_1, \dots, u_n)^\top$  e  $\mathbf{t} = (t_1, \dots, t_n)^\top$ , em que  $\mathbf{b}$ ,  $\mathbf{u}$  e  $\mathbf{t}$  são variáveis não observadas e, sob a representação hierárquica (4.5)-(4.8), a função de log-verossimilhança completa associada a  $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{b}^\top, \mathbf{u}^\top, \mathbf{t}^\top)^\top$  é dada por

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{y}_c) &\propto \frac{1}{2} \sum_{i=1}^n \left\{ -\log |\mathbf{R}_i| - u_i(\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) - \mathbf{Z}_i \mathbf{b}_i)^\top \mathbf{R}_i^{-1} (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) - \mathbf{Z}_i \mathbf{b}_i) \right\} + \\ &\frac{1}{2} \sum_{i=1}^n \left\{ -\log |\boldsymbol{\Gamma}| - u_i(\mathbf{b}_i - \boldsymbol{\Delta} \mathbf{t}_i)^\top \boldsymbol{\Gamma}^{-1} (\mathbf{b}_i - \boldsymbol{\Delta} \mathbf{t}_i) \right\}. \end{aligned} \quad (4.9)$$

Considerando  $\widehat{\boldsymbol{\theta}}^{(k)} = (\widehat{\boldsymbol{\alpha}}^{(k)\top}, \widehat{\boldsymbol{\gamma}}^{(k)}, \widehat{\boldsymbol{\beta}}^{(k)\top}, \widehat{\boldsymbol{\lambda}}^{(k)\top})^\top$  a estimativa de  $\boldsymbol{\theta}$  na  $k$ -ésima iteração do algoritmo do tipo EM, a esperança condicional da função de log-verossimilhança dos dados completos tem a forma

$$Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)}) = E[\ell(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)}] \propto \sum_{i=1}^n Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}}^{(k)}) + \sum_{i=1}^n Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}}^{(k)}), \quad (4.10)$$

onde  $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \boldsymbol{\gamma})^\top$  e  $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$ . Além disso,

$$\begin{aligned} Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2} \log |\widehat{\mathbf{R}}_i^{(k)}| - \frac{\widehat{u}_i^{(k)}}{2} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i))^\top \widehat{\mathbf{R}}_i^{(k)-1} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i)) + \\ &(\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i))^\top \widehat{\mathbf{R}}_i^{(k)-1} \mathbf{Z}_i \widehat{(\mathbf{ub})}_i^{(k)} - \frac{1}{2} \text{tr} \left( \widehat{\mathbf{R}}_i^{(k)-1} \mathbf{Z}_i \widehat{(\mathbf{ubb}^\top)}_i^{(k)} \mathbf{Z}_i^\top \right) \text{ e } \end{aligned} \quad (4.11)$$

$$\begin{aligned} Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2} \log |\widehat{\boldsymbol{\Gamma}}^{(k)}| - \frac{1}{2} \text{tr} \left( \widehat{\boldsymbol{\Gamma}}^{(k)-1} \widehat{(\mathbf{ubb}^\top)}_i^{(k)} \right) + \widehat{(\mathbf{utb})}_i^{(k)\top} \widehat{\boldsymbol{\Gamma}}^{(k)-1} \widehat{\boldsymbol{\Delta}}^{(k)} - \\ &\frac{\widehat{(ut^2)}_i^{(k)}}{2} \widehat{\boldsymbol{\Delta}}^{(k)\top} \widehat{\boldsymbol{\Gamma}}^{(k)-1} \widehat{\boldsymbol{\Delta}}^{(k)}, \end{aligned} \quad (4.12)$$

onde  $\text{tr}(\mathbf{M})$  indica o traço da matriz  $\mathbf{M}$ .

Para o cálculo destas funções são necessárias as quantidades  $\widehat{u}_i^{(k)} = E(U_i|\mathbf{y}, \widehat{\boldsymbol{\theta}}^{(k)})$ ,  $\widehat{\mathbf{ub}}_i^{(k)} = E(U_i \mathbf{b}_i | \mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(k)})$ ,  $\widehat{(\mathbf{ubb}^\top)}_i^{(k)} = E(U_i \mathbf{b}_i \mathbf{b}_i^\top | \mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(k)})$ ,  $\widehat{ut}_i^{(k)} = E(U_i T_i | \mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(k)})$ ,  $\widehat{ut^2}_i^{(k)} = E(U_i T_i^2 | \mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(k)})$  e  $\widehat{\mathbf{utb}}_i^{(k)} = E(U_i T_i \mathbf{b}_i | \mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(k)})$ , cujas expressões são apresentados a seguir.

$$\widehat{ut}_i^{(k)} = \widehat{u}_i^{(k)} \widehat{\mu}_i^{(k)} + \widehat{M}_i^{(k)} \widehat{\tau}_i^{(k)}, \quad (4.13)$$

$$\widehat{ut^2}_i^{(k)} = \widehat{u}_i^{(k)} \widehat{\mu}_i^{(k)2} + \widehat{M}_i^{(k)2} + \widehat{M}_i^{(k)} \widehat{\mu}_i^{(k)} \widehat{\tau}_i^{(k)}, \quad (4.14)$$

$$\widehat{\mathbf{ub}}_i^{(k)} = \widehat{u}_i^{(k)} \widehat{\mathbf{G}}_i^{(k)} + \widehat{\mathbf{s}}_i^{(k)} \widehat{(ut)}_i^{(k)}, \quad (4.15)$$

$$\widehat{\mathbf{utb}}_i^{(k)} = \widehat{\mathbf{G}}_i^{(k)} \widehat{(ut)}_i^{(k)} + \widehat{\mathbf{s}}_i^{(k)} \widehat{(ut^2)}_i^{(k)} \text{ e } \quad (4.16)$$

$$\widehat{(\mathbf{ubb}^\top)}_i^{(k)} = \widehat{\mathbf{B}}_i^{(k)} + \widehat{u}_i^{(k)} \widehat{\mathbf{G}}_i^{(k)} \widehat{\mathbf{G}}_i^{(k)\top} + \widehat{(ut)}_i^{(k)} (\widehat{\mathbf{G}}_i^{(k)} \widehat{\mathbf{s}}_i^{(k)\top} + \widehat{\mathbf{s}}_i^{(k)} \widehat{\mathbf{G}}_i^{(k)\top}) + \widehat{(ut^2)}_i^{(k)} \widehat{\mathbf{s}}_i^{(k)} \widehat{\mathbf{s}}_i^{(k)\top}, \quad (4.17)$$

em que

$$\begin{aligned}\widehat{\boldsymbol{\mu}}_i^{(k)} &= \widehat{M}_i^{(k)} 2 \widehat{\boldsymbol{\Delta}}^{(k)\top} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Omega}}_i^{(k)-1} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i)), \\ \widehat{M}_i^{(k)} &= (1 + \widehat{\boldsymbol{\Delta}}^{(k)\top} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Omega}}_i^{(k)-1} \mathbf{Z}_i \widehat{\boldsymbol{\Delta}}^{(k)})^{-1/2}, \\ \widehat{\mathbf{B}}_i^{(k)} &= (\widehat{\boldsymbol{\Gamma}}^{(k)-1} + \mathbf{Z}_i^\top \widehat{\mathbf{R}}_i^{(k)-1} \mathbf{Z}_i)^{-1}, \\ \widehat{\mathbf{G}}_i^{(k)} &= \widehat{\mathbf{B}}_i^{(k)} \mathbf{Z}_i^\top \widehat{\mathbf{R}}_i^{(k)-1} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i)), \\ \widehat{\mathbf{s}}_i^{(k)} &= (\mathbf{I}_q - \widehat{\mathbf{B}}_i^{(k)} \mathbf{Z}_i^\top \widehat{\mathbf{R}}_i^{(k)-1} \mathbf{Z}_i) \widehat{\boldsymbol{\Delta}}^{(k)} \mathbf{e} \\ \widehat{\boldsymbol{\Omega}}_i^{(k)} &= \widehat{\mathbf{R}}_i^{(k)} + \mathbf{Z}_i \widehat{\boldsymbol{\Gamma}}^{(k)} \mathbf{Z}_i^\top.\end{aligned}$$

Os estimadores de  $\widehat{u}_i^{(k)} = \widehat{u}_{1i}^{(k)} = E(U_i | \mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(k)})$  e  $\widehat{\tau}_i^{(k)} = \widehat{\tau}_{1i}^{(k)} = E(U_i^{1/2} W_{\Phi}(U_i^{1/2} \widehat{A}_i^{(k)}) | \mathbf{y}_i, \widehat{\boldsymbol{\theta}}^{(k)})$  estão definidos em (3.15) e (3.16), e apresentados para cada distribuição pertencente a família SMSN, com  $\widehat{A}_i^{(k)} = \boldsymbol{\lambda}_{\mathbf{y}_i}^{(k)\top} \boldsymbol{\Sigma}_i^{(k)-1/2} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i))$ .

Esse processo iterativo de esperança e maximização função de log-verossimilhança dos dados completos se repete até que a convergência seja alcançada, por exemplo, até que  $\max \|\widehat{\boldsymbol{\theta}}^{(k+1)} - \widehat{\boldsymbol{\theta}}^{(k)}\|$  seja suficientemente pequeno. Em cada iteração  $k$  do algoritmo do tipo EM, o valor das estimativas dos parâmetros da função não linear  $\widehat{\boldsymbol{\alpha}}^{(k)}$  são obtidas por meio do algoritmo Newton-Raphson. Inicialmente optamos pelo algoritmo escore de Fisher, entretanto, em virtude de  $E(\mathbf{y}_i) \neq \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)$ , quando  $\boldsymbol{\lambda} \neq \mathbf{0}$ , conforme (3.12), o algoritmo escore de Fisher mostrou-se menos eficiente. Assim, obtemos as estimativas  $\widehat{\boldsymbol{\alpha}}^{(k)}$  pelo algoritmo de Newton-Raphson em  $v$  iterações, por meio da função escore e da matriz de informação de Fisher observada. Os passos para a implementação do algoritmo Newton-Raphson apresentados a seguir têm por objetivo facilitar sua implementação. Nas expressões a seguir consideramos  $\widehat{\mathbf{R}}_i = \widehat{\boldsymbol{\sigma}}^2 \mathbf{I}_{m_i}$ , ou seja,  $\widehat{\boldsymbol{\gamma}} = \widehat{\boldsymbol{\sigma}}^2$ . A função escore de  $\boldsymbol{\alpha}$  na  $v$ -ésima iteração do algoritmo de Newton-Raphson é dada por

$$\mathbf{U}_{\boldsymbol{\alpha}}^{(v+1)} = \frac{1}{\widehat{\boldsymbol{\sigma}}^2} \sum_{i=1}^n \left\{ \widehat{\mathbf{J}}_i^{(v)\top} \left( \widehat{u}_i^{(k)} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x}_i)) - \mathbf{Z}_i (\widehat{\mathbf{ub}})_i^{(k)} \right) \right\},$$

em que  $\widehat{\mathbf{J}}_i^{(v)}$  é uma matriz de elementos da forma  $\frac{\partial \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x}_i)}{\partial \widehat{\alpha}_r^{(v)}}$ ,  $r = 1, \dots, p$ , é uma matriz de dimensão  $m_i \times p$ .

Para obtermos a expressão da matriz de informação de Fisher observada, calculamos a matriz  $\widehat{\mathbf{L}}_i^{(v)}$  cujos elementos são  $\frac{\partial^2 \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x}_i)}{\partial \widehat{\alpha}_r^{(v)} \partial \widehat{\alpha}_s^{(v)}}$ ,  $r, s = 1, \dots, p$ , de dimensão  $m_i \times p^2$ , em seguida calculamos o vetor  $\widetilde{\mathbf{L}}_i^{(v)}$  de dimensão  $p^2 \times 1$  com a expressão

$$\widetilde{\mathbf{L}}_i^{(v)} = \widehat{\mathbf{L}}_i^{(v)\top} \{ \mathbf{Z}_i (\widehat{\mathbf{ub}})_i^{(k)} - \widehat{u}_i^{(k)} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x}_i)) \}.$$

A matriz de informação de Fisher observada é dada pela expressão

$$\mathbf{K}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(v+1)} = \frac{1}{\widehat{\boldsymbol{\sigma}}^2} \sum_{i=1}^n \left\{ \widehat{u}_i^{(k)} \widehat{\mathbf{J}}_i^{(v)\top} \widehat{\mathbf{J}}_i^{(v)} + \text{vec}^{-1}(\widetilde{\mathbf{L}}_i^{(v)}) \right\}.$$

O estimador do vetor de parâmetros  $\boldsymbol{\alpha}$  no algoritmo de Newton-Raphson na  $(v+1)$ -ésima iteração é obtido por

$$\widehat{\boldsymbol{\alpha}}^{(v+1)} = \widehat{\boldsymbol{\alpha}}^{(v)} + \left[ \mathbf{K}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(v+1)} \right]^{-1} \mathbf{U}_{\boldsymbol{\alpha}}^{(v+1)}.$$

Após a convergência do algoritmo Newton-Raphson a estimativa  $\boldsymbol{\alpha}^{(v+1)}$  é computada no algoritmo tipo EM na  $k$ -ésima iteração. No passo M do algoritmo do tipo EM são feitas maximizações condicionais e maximização da função de log-verossimilhança marginal restrita. Sob a distribuição SSL são calculadas integrais de Monte Carlo. Portanto, utilizamos uma combinação entre os algoritmos MCEM (WEI; TANNER, 1990) e ECME (LIU; RUBIN, 1994) que são extensões rápidas dos algoritmos EM (DEMPSTER; LAIRD; RUBIN, 1977) e ECM (MENG; RUBIN, 1993), cujos passos estão descritos a seguir.

**Passo E:** Calcular o valores de  $\widehat{u}_i^{(k)}$ ,  $\widehat{ut}_i^{(k)}$ ,  $\widehat{ut^2}_i^{(k)}$ ,  $\widehat{ub}_i^{(k)}$ ,  $\widehat{utb}_i^{(k)}$  e  $\widehat{ubb^\top}_i^{(k)}$  com as expressões (4.13)-(4.17).

Maximizando a função  $Q(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(k)})$  em relação a  $\boldsymbol{\theta}$ , obtemos as estimativas de  $\widehat{\boldsymbol{\theta}}^{(k+1)}$  cujas expressões são apresentadas no Passo CM.

**Passo CM:** Atualizar  $\widehat{\boldsymbol{\alpha}}^{(k)}$ ,  $\widehat{\boldsymbol{\Delta}}^{(k)}$ ,  $\widehat{\boldsymbol{\Gamma}}^{(k)}$  e  $\widehat{\sigma^2}^{(k)}$  com as expressões

$$\widehat{\boldsymbol{\alpha}}^{(k+1)} = \operatorname{argmin}_{\boldsymbol{\alpha}} \left\{ \frac{\widehat{u}_i^{(k)}}{2} \mathbf{r}_i^{(k)\top} \mathbf{r}_i^{(k)} - \mathbf{r}_i^{(k)\top} \mathbf{Z}_i \widehat{(\mathbf{ub})}_i^{(k)} \right\}, \quad (4.18)$$

$$\widehat{\boldsymbol{\Delta}}^{(k+1)} = \frac{\sum_{i=1}^n \widehat{(\mathbf{utb})}_i^{(k)}}{\sum_{i=1}^n \widehat{(ut^2)}_i^{(k)}}, \quad (4.19)$$

$$\widehat{\boldsymbol{\Gamma}}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left\{ \widehat{(\mathbf{ubb^\top})}_i^{(k)} - \widehat{(\mathbf{utb})}_i^{(k)} \widehat{\boldsymbol{\Delta}}_i^{(k+1)\top} - \widehat{\boldsymbol{\Delta}}_i^{(k+1)} \widehat{(\mathbf{utb})}_i^{(k)\top} + \widehat{(ut^2)}_i^{(k)} \widehat{\boldsymbol{\Delta}}_i^{(k+1)} \widehat{\boldsymbol{\Delta}}_i^{(k+1)\top} \right\} \mathbf{e} \quad (4.20)$$

$$\widehat{\sigma^2}^{(k+1)} = \frac{1}{N} \sum_{i=1}^n \left\{ \widehat{u}_i^{(k)} \widehat{\mathbf{r}}_i^{(k+1)\top} \widehat{\mathbf{r}}_i^{(k+1)} - \widehat{\mathbf{r}}_i^{(k+1)\top} \mathbf{Z}_i \widehat{(\mathbf{ub})}_i^{(k)} - \widehat{(\mathbf{ub})}_i^{(k)\top} \mathbf{Z}_i^\top \widehat{\mathbf{r}}_i^{(k+1)} + \operatorname{tr} \left( \mathbf{Z}_i \widehat{(\mathbf{ubb^\top})}_i^{(k)} \mathbf{Z}_i^\top \right) \right\}, \quad (4.21)$$

em que  $N = \sum_{i=1}^n m_i$  e  $\widehat{\mathbf{r}}_i^{(k)} = \mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i)$ .

**Passo CML:** Atualizar o valor do vetor de parmetros  $\mathbf{v}^{(k)}$  considerando a funo de log-verossimilhana (4.3) na k-sima iterao do algoritmo do tipo EM por meio da expresso

$$\widehat{\mathbf{v}}^{(k+1)} = \operatorname{argmax}_{\mathbf{v}} \{g(\mathbf{y}; \boldsymbol{\theta}^{(k+1)}, \mathbf{v})\}. \quad (4.22)$$

A matriz de disperso e o vetor de parmetros de assimetria dos efeitos aleatrios so estimados com as expresses

$$\widehat{\mathbf{D}}^{(k+1)} = \widehat{\mathbf{\Gamma}}^{(k+1)} + \widehat{\mathbf{\Delta}}^{(k+1)} \widehat{\mathbf{\Delta}}^{(k+1)\top} \mathbf{e} \quad (4.23)$$

$$\widehat{\boldsymbol{\lambda}}^{(k+1)} = \frac{\widehat{\mathbf{D}}^{(k+1)-1/2} \widehat{\mathbf{\Delta}}^{(k+1)}}{(1 - \widehat{\mathbf{\Delta}}^{(k+1)\top} \widehat{\mathbf{D}}^{(k+1)-1} \widehat{\mathbf{\Delta}}^{(k+1)})^{1/2}}. \quad (4.24)$$

Sob certas condies de regularidade (ZACKS, 1971), a matriz do covarincias dos estimadores de mxima verossimilhana pode ser estimada por meio da inversa da matriz de informao de Fisher observada apresentada a seguir.

$$\mathbf{L}(\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n \left\{ \mathbf{U}_i(\widehat{\boldsymbol{\theta}}) \mathbf{U}_i(\widehat{\boldsymbol{\theta}})^\top \right\} \quad \text{com} \quad (4.25)$$

$$\mathbf{U}_i(\widehat{\boldsymbol{\theta}}) = \left[ \frac{\partial \log g(\mathbf{y}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}, \quad (4.26)$$

em que  $g(\mathbf{y}_i, \boldsymbol{\theta})$   a fdp da distribuio pertencente  classe SMSN.

#### 4.1.1 Predio dos efeitos aleatrios e da varivel resposta

Para predio dos efeitos aleatrios  $\mathbf{b}_i$  em modelos lineares com efeitos mistos sob a classe SNI, Lachos et al. (2010) (LACHOS; GHOSH; ARELLANO-VALLE, 2010) desenvolveu um estimador emprico de Bayes, e que ns estendemos para o modelo SMSN-NLME, obtido pela expresso

$$\widehat{\mathbf{b}}_i = E[\mathbf{b}_i | \mathbf{y}_i, \widehat{\boldsymbol{\theta}}] = \widehat{\boldsymbol{\mu}}_{\mathbf{b}_i} + \frac{\widehat{\tau}_{-1i}}{\sqrt{1 + \widehat{\boldsymbol{\zeta}}^\top \widehat{\boldsymbol{\Lambda}}_i \widehat{\boldsymbol{\zeta}}}} \widehat{\boldsymbol{\Lambda}}_i \widehat{\boldsymbol{\zeta}}, \quad (4.27)$$

em que  $\widehat{\boldsymbol{\mu}}_{\mathbf{b}_i} = \widehat{\mathbf{D}} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Sigma}}^{-1} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}, \mathbf{x}_i))$  e  $\widehat{\tau}_{-1i} = E(\kappa^{1/2} (U_i) W_\Phi(\kappa^{-1/2} (U_i) \widehat{A}_i) | \mathbf{y}_i, \widehat{\boldsymbol{\theta}})$ , conforme (3.16).

Assim, a predição de  $\mathbf{y}_i$  para o  $i$ -ésimo perfil assume a forma

$$\begin{aligned}\hat{\mathbf{y}}_i &= \mathbf{f}(\hat{\boldsymbol{\alpha}}, \mathbf{x}_i) + \mathbf{Z}_i \hat{\mathbf{b}}_i \\ &= \mathbf{f}(\hat{\boldsymbol{\alpha}}, \mathbf{x}_i) + \mathbf{Z}_i \widehat{\mathbf{D}} \mathbf{Z}_i^\top \widehat{\boldsymbol{\Sigma}}_i^{-1} (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}, \mathbf{x}_i)) + \mathbf{Z}_i \frac{\hat{\tau}_{-1i}}{\sqrt{1 + \hat{\boldsymbol{\zeta}}^\top \widehat{\boldsymbol{\Lambda}}_i \hat{\boldsymbol{\zeta}}}} \widehat{\boldsymbol{\Lambda}}_i \hat{\boldsymbol{\zeta}} \\ &= \widehat{\sigma^2} \widehat{\boldsymbol{\Sigma}}_i^{-1} \mathbf{f}(\hat{\boldsymbol{\alpha}}, \mathbf{x}_i) + (\mathbf{I}_{m_i} - \widehat{\sigma^2} \widehat{\boldsymbol{\Sigma}}_i^{-1}) \mathbf{y}_i + \mathbf{Z}_i \frac{\hat{\tau}_{-1i}}{\sqrt{1 + \hat{\boldsymbol{\zeta}}^\top \widehat{\boldsymbol{\Lambda}}_i \hat{\boldsymbol{\zeta}}}} \widehat{\boldsymbol{\Lambda}}_i \hat{\boldsymbol{\zeta}}.\end{aligned}\quad (4.28)$$

Observamos que o valor predito de  $\mathbf{y}_i$  é a soma de uma parcela relacionada com a função estimada, outra com seus valores observados e outra com a assimetria.

Na predição de uma observação futura da variável resposta que denotamos por  $\mathbf{y}_i^+$  de dimensão  $a \times 1$ , com  $\mathbf{x}_i^+$  e  $\mathbf{z}_i^+$  matrizes de regressores de predição, de dimensões  $a \times 1$  e  $a \times q$ , respectivamente, assumimos que

$$\begin{pmatrix} \mathbf{y}_i \\ \mathbf{y}_i^+ \end{pmatrix} \sim SMSN_{m_i+a}(\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i^*), \boldsymbol{\Sigma}_i^*, \boldsymbol{\lambda}_i^*; H), \quad (4.29)$$

onde  $\mathbf{x}_i^* = (\mathbf{x}_i^\top, \mathbf{x}_i^{+\top})^\top$ ,  $\mathbf{z}_i^* = (\mathbf{z}_i^\top, \mathbf{z}_i^{+\top})^\top$ ,  $\boldsymbol{\Sigma}_i^* = \mathbf{R}_i^* + \mathbf{Z}_i^* \mathbf{D} \mathbf{Z}_i^{*\top}$  e

$$\boldsymbol{\lambda}_i^* = \frac{\boldsymbol{\Sigma}_i^{*-1/2} \mathbf{Z}_i^* \mathbf{D} \boldsymbol{\zeta}}{\sqrt{1 + \boldsymbol{\zeta}^\top \boldsymbol{\Lambda}_i^* \boldsymbol{\zeta}}}, \text{ com } \boldsymbol{\Lambda}_i^* = (\mathbf{D}^{-1} + \mathbf{Z}_i^{*\top} \mathbf{R}_i^{*-1} \mathbf{Z}_i^*)^{-1} \text{ e } \boldsymbol{\zeta} = \mathbf{D}^{-1/2} \boldsymbol{\lambda}.$$

Portanto, o vetor aleatório  $\mathbf{y}_i$  segue a distribuição  $SMSN_{m_i}(\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i), \boldsymbol{\Sigma}_i, \boldsymbol{\Sigma}_i^{1/2} \tilde{\mathbf{V}}_i; H)$ , em que

$$\tilde{\mathbf{V}}_i = \frac{\mathbf{V}_i^{(1)} + \boldsymbol{\Sigma}_{i11}^{*-1} \boldsymbol{\Sigma}_{i12}^* \mathbf{V}_i^{(2)}}{\sqrt{1 + \mathbf{V}_i^{(2)\top} \boldsymbol{\Sigma}_{i22.1}^* \mathbf{V}_i^{(2)}}}, \quad (4.30)$$

$$\mathbf{V}_i = (\mathbf{V}_i^{(1)\top}, \mathbf{V}_i^{(2)\top})^\top = \boldsymbol{\Sigma}_i^{*-1/2} \boldsymbol{\lambda}_i^*, \quad (4.31)$$

onde  $\boldsymbol{\Sigma}_{i11}^* = \boldsymbol{\Sigma}_i$ ,  $\boldsymbol{\Sigma}_{i22.1}^* = \boldsymbol{\Sigma}_{i22}^* - \boldsymbol{\Sigma}_{i21}^* \boldsymbol{\Sigma}_{i11}^{*-1} \boldsymbol{\Sigma}_{i12}^*$  e  $\boldsymbol{\Sigma}_{i12}^* = \boldsymbol{\Sigma}_{i21}^{*\top}$ .

Assim, substituindo  $\boldsymbol{\theta}$  por  $\hat{\boldsymbol{\theta}}$ , o preditor da observação futura  $\hat{\mathbf{y}}_i^+ = \hat{\mathbf{y}}_i^+(\hat{\boldsymbol{\theta}})$  com menor erro quadrático médio é a esperança condicional de  $\mathbf{y}_i^+$  dado  $\mathbf{y}_i$  (LACHOS; GHOSH; ARELLANO-VALLE, 2010), ou seja

$$\hat{\mathbf{y}}_i^+(\hat{\boldsymbol{\theta}}) = E(\mathbf{y}_i^+ | \mathbf{y}_i, \hat{\boldsymbol{\theta}}) = \hat{\boldsymbol{\mu}}_{\mathbf{y}_i^+} + \frac{\widehat{\boldsymbol{\Sigma}}_{i22.1}^* \widehat{\mathbf{V}}_i^{(2)} \hat{\tau}_{-1i}}{\sqrt{1 + \widehat{\mathbf{V}}_i^{(2)\top} \widehat{\boldsymbol{\Sigma}}_{i22.1}^* \widehat{\mathbf{V}}_i^{(2)}}}, \quad (4.32)$$

onde  $\hat{\boldsymbol{\mu}}_{\mathbf{y}_i^+} = \mathbf{f}(\hat{\boldsymbol{\alpha}}, \mathbf{x}_i^+) + \widehat{\boldsymbol{\Sigma}}_{i21}^* \widehat{\boldsymbol{\Sigma}}_{i11}^{*-1} (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}, \mathbf{x}_i))$ .

## 4.2 Estatística da razão de verossimilhanças

Usamos a estatística da razão de verossimilhanças e a verossimilhança perfilada para realizar testes de hipóteses e construir intervalos de confiança para os parâmetros de assimetria. Testamos a hipótese nula  $H_0 : \boldsymbol{\lambda} = \mathbf{0}$  contra a hipótese alternativa  $H_a : \boldsymbol{\lambda} \neq \mathbf{0}$ , e construímos intervalos de confiança aproximados de 95%, como sugerem Montenegro et al. (MONTENEGRO; LACHOS; BOLFARINE, 2010) e Zeller (ZELLER, 2009).

### 4.2.1 Teste de hipóteses

Por meio do teste da razão de verossimilhanças (TRV) avaliamos o máximo das funções log-verossimilhanças dos modelos restrito ou simétrico e irrestrito ou assimétrico. Sob a hipótese nula, a estatística  $2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_0)\}$ , em que  $\boldsymbol{\theta}_0$  é o vetor de parâmetros do modelo ajustado sob  $H_0$ , possui distribuição assintótica qui-quadrado com  $q$  graus de liberdade ( $\chi_q^2$ ) (PAWITAN, 2001). Nas aplicações a seguir consideramos o nível de significância de 5%, assim rejeitamos a hipótese nula se  $2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\widehat{\boldsymbol{\theta}}_0)\} > \chi_{0,95;q}^2$ .

### 4.2.2 Intervalo de confiança aproximado

Usando também a estatística da razão de verossimilhanças, podemos construir um intervalo de confiança aproximado de 95% para um elemento do vetor de parâmetros de assimetria  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_q)^\top$ , por exemplo para  $\lambda_1$ , considerando que  $2\{\ell(\widehat{\boldsymbol{\theta}}) - \ell(\lambda_1)\} < \chi_{0,95;1}^2$ , em que  $\ell(\lambda_1)$  é a log-verossimilhança perfilada.

Realizamos algumas simulações e verificamos que a distribuição das estimativas dos elementos do vetor de parâmetros  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top$  se aderem à distribuição normal, o que não ocorreu com os elementos do vetor de parâmetros de assimetria. Assim, construímos também intervalos de confiança de 95% para  $\alpha_l$ ,  $l = 1, \dots, p$ , considerando o intervalo de confiança usual de Wald, por exemplo para  $\alpha_1$ , dado pela expressão  $[\widehat{\alpha}_1 - 1,96\widehat{\text{ep}}(\widehat{\alpha}_1); \widehat{\alpha}_1 + 1,96\widehat{\text{ep}}(\widehat{\alpha}_1)]$ , com  $\widehat{\text{ep}}(\widehat{\alpha}_1)$  sendo a estimativa do erro padrão de  $\widehat{\alpha}_1$ , obtida por meio da matriz de informação de Fisher observada. Estes intervalos de confiança são discutidos no estudo de simulações na Seção 4.4.

## 4.3 Aplicações

Realizamos algumas aplicações ajustando o modelo SMSN-NLME com os dados de crescimento de laranjeiras, e concentrações sanguíneas de cefamandol e *theophylline* descritos no Capítulo 2. As análises de resíduos estão apresentadas no Capítulo 5. Como critérios de seleção de modelos, além de testes de hipóteses com a estatística da razão de verossimilhanças, adotamos o critério de informação de Akaike (AIC) e o critério de informação Bayesiana (BIC), cujas expressões são dadas por

$$\text{AIC} = -2\ell(\hat{\boldsymbol{\theta}}) + 2l \quad \text{e} \quad (4.33)$$

$$\text{BIC} = -2\ell(\hat{\boldsymbol{\theta}}) + l\log(N), \quad (4.34)$$

em que  $l$  representa o número de elementos do vetor de parâmetros  $\hat{\boldsymbol{\theta}}$  e  $N = \sum_{i=1}^n m_i$ .

Na estimação de  $\mathbf{v}$  nos ajustes sob as distribuições ST, SSL e SNC, fizemos um *grid* de valores e maximizamos a função de log-verossimilhança, de acordo com (4.22), com o auxílio da biblioteca *EMMIXskew* (WANG; NG; MCLACHLAN., 2018).

### 4.3.1 Dados de crescimento de laranjeiras

Consideramos o modelo (4.1) ajustado aos dados de crescimento de árvores de laranjeiras apresentados na Seção 2.1, em que  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top = (\text{Asym}, \text{xmid}, \text{scal})^\top$  é o vetor de parâmetros da função não linear,  $\mathbf{R}_i = \sigma^2 \mathbf{I}_7$  é a matriz de dispersão dos erros,  $\mathbf{y}_i = (y_{i1}, \dots, y_{i7})^\top$ ,  $i = 1, \dots, 5$ , são as medições dos comprimentos das circunferências das árvores (em *mm*). O vetor que representa as medições do tempo (em dias) é  $\mathbf{x}_i = \mathbf{t}_i = (t_{i1}, \dots, t_{i7})^\top$ , e

$$\mathbf{Z}_i = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i)}{\partial \text{Asym}}, \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i)}{\partial \text{scal}} \right] \Bigg|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}}$$

é a matriz de constantes conhecidas, de dimensão  $7 \times 2$ , conforme apresentada em (4.4), com efeitos aleatórios associados aos parâmetros *Asym* e *scal*.

Assim, o modelo não linear com efeitos mistos que relaciona as medições dos comprimentos das circunferências das árvores e o tempo pode ser representado por

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 5, \quad (4.35)$$

em que  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) = (f(\boldsymbol{\alpha}, t_{i1}), \dots, f(\boldsymbol{\alpha}, t_{i7}))^\top$  é uma função não linear de parâmetros  $\boldsymbol{\alpha}$ , tal que

$$f(\boldsymbol{\alpha}, t_{ij}) = \frac{\text{Asym}}{1 + \exp[-(t_{ij} - \text{xmid})/\text{scal}]}, \quad i = 1, \dots, 5, \quad j = 1, \dots, 7. \quad (4.36)$$

Com  $\boldsymbol{\beta} = (d_{11}, d_{12}, d_{22})^\top$ , tal que  $d_{11}$ ,  $d_{12}$  e  $d_{22}$  são elementos da matriz  $\mathbf{D}^{1/2}$ , e  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ , as estimativas de  $\boldsymbol{\theta}$  e respectivos erros padrões, obtidos por meio da matriz de informação de Fisher observada com o auxílio da biblioteca *numDeriv* (GILBERT; VARADHAN, 2016), estão apresentados na Tabela 4.

Tabela 4 – Estimativa dos parâmetros (Estim.) com erros padrões (E.P.) para os modelos N-SMSN, SN-NLME, ST-NLME, SSL-NLME e SNC-NLME ajustados aos dados de crescimento de laranjeiras.

$\theta$	N-NLME		SN-NLME		ST-NLME		SSL-NLME		SNC-NLME	
	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)
	-	-	-	-	$\nu = 8$		$\nu = 4$		$\mathbf{v} = (0, 8; 0, 3)^T$	
Asym	192,20	(7,11)	190,95	(7,13)	155,53	(7,95)	183,49	(7,21)	158,88	(8,60)
xmid	729,01	(37,71)	729,72	(38,06)	757,24	(53,60)	733,93	(40,07)	771,85	(58,49)
scal	353,59	(28,78)	353,83	(29,01)	382,25	(38,33)	355,37	(30,31)	407,66	(40,19)
$\sigma^2$	64,40	(15,39)	64,42	(15,40)	56,67	(13,54)	52,23	(12,47)	19,44	(4,64)
$d_{11}$	877,89	(9,31)	880,94	(9,33)	1900,71	(1,51)	781,55	(8,70)	603,65	(0,09)
$d_{12}$	-97,32	(1,11)	-96,06	(1,10)	-843,17	(0,92)	-89,19	(1,07)	-588,28	(0,11)
$d_{22}$	11,81	(0,34)	11,49	(0,34)	374,98	(0,15)	11,18	(0,34)	573,91	(0,16)
$\lambda_1$	-	-	0,06	(0,45)	18,09	(5,76)	0,39	(0,46)	12,91	(4,16)
$\lambda_2$	-	-	-0,09	(0,44)	-7,51	(2,41)	-0,03	(0,44)	-11,86	(3,80)

Observa-se que o modelo ajustado sob a distribuição normal apresentou os menores erros padrões para os parâmetros Asym, xmid e scal. As componentes aleatórias apresentaram os menores erros padrões com a distribuição SNC. Os parâmetros de assimetria apresentaram menores erros padrões sob a distribuição SN. Na Tabela 5 encontram-se os valores dos critérios AIC e BIC.

Tabela 5 – Valores AIC e BIC para os modelos ajustados aos dados de crescimento de laranjeiras.

Critério	N-NLME	SN-NLME	ST-NLME	SSL-NLME	SNC-NLME
	-	-	$\nu = 8$	$\nu = 4$	$\mathbf{v} = (0, 8; 0, 3)^T$
AIC	277,49	277,50	275,96	278,02	273,38
BIC	291,48	291,49	289,95	292,01	287,37

Os valores dos critérios apresentados na Tabela 5 são bem próximos, principalmente sob as distribuições normal e skew-normal, porém o menor valor ocorreu com a distribuição skew-normal contaminada.

A Figura 20 mostra gráficos, gerados com o auxílio da biblioteca *lattice* (SARKAR, 2008), de perfis ajustados sob a distribuição SNC, em preto, e dados observados, em cinza, para cada árvore de laranjeira.

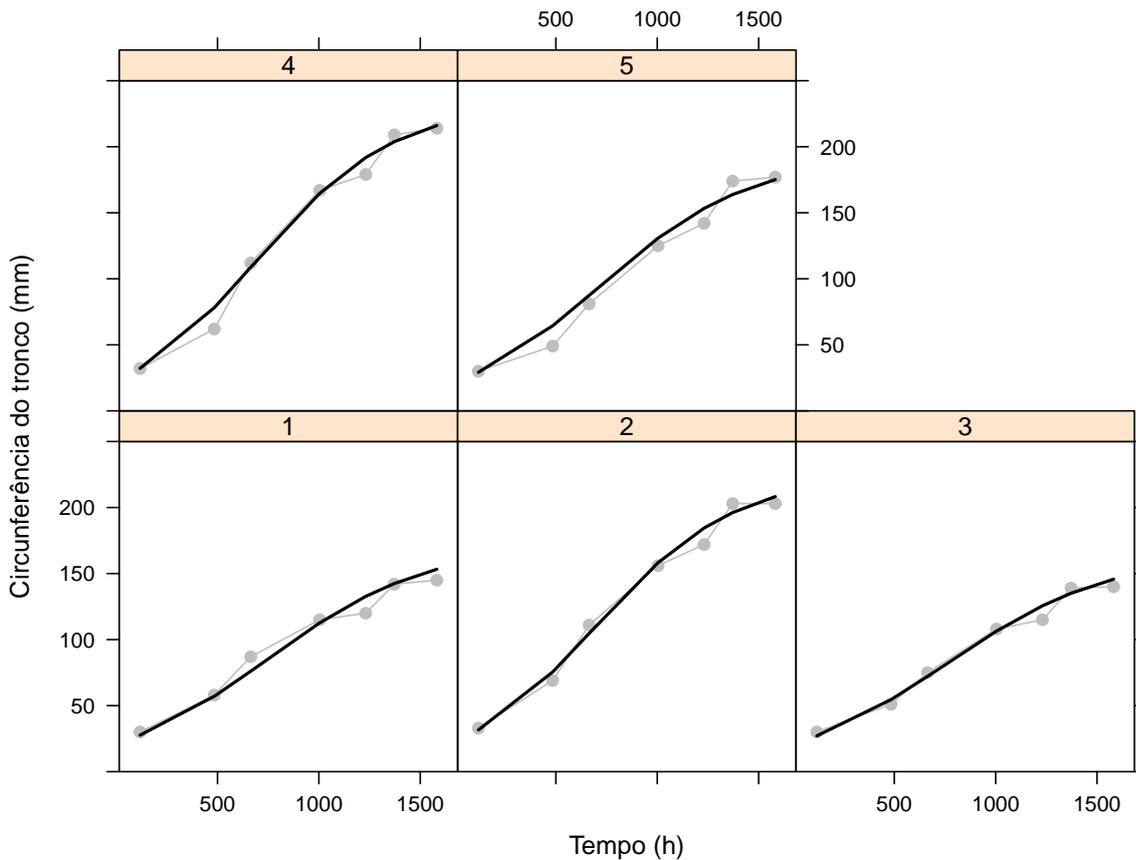


Figura 20 – Perfis ajustados sob o modelo skew-normal contaminada por árvore de laranjeira.

Apresentamos na Tabela 6 os resultados do teste da razão de verossimilhanças para testar se o modelo simétrico é suficiente nos ajustes dos dados.

Tabela 6 – Resultados do teste da razão de verossimilhanças para testar  $H_0 : \boldsymbol{\lambda} = \mathbf{0}$  aplicados aos dados de crescimento de laranjeiras.

Distribuições	$p$ -valor
normal / skew-normal	0,998
t / skew-t	0,167
slash / skew-slash	0,998
normal contaminada / skew-normal contaminada	0,056

Observamos que o modelo simétrico é mais adequado em todas as situações, considerando o nível de significância de 5%. A rejeição de  $H_0$  pode ser obtida se considerarmos o nível de significância maior que 5,6% e sob a distribuição skew-normal contaminada.

Para avaliar a adequabilidade da metodologia proposta, as medidas  $y_{i7}$ ,  $i = 1, \dots, 5$ , das circunferências das laranjeiras das árvores foram consideradas desconhecidas e, assim, previstas como observações futuras, de acordo com (4.32). O modelo SNC-NLME foi ajustado com apenas 6 medidas de cada indivíduo e as observações futuras  $\hat{y}_{i7}^+ = \hat{y}_{i7}^+(\hat{\boldsymbol{\theta}})$ ,  $i = 1, \dots, 5$ , foram

obtidas. Calculamos o valor absoluto do desvio relativo por meio da expressão  $|(y_{i7} - \widehat{y}_{i7}^+)/y_{i7}|$ ,  $i = 1, \dots, 5$ , entre os valores observados e preditos. Os resultados estão apresentados na Tabela 7 a seguir.

Tabela 7 – Valores absolutos do desvio relativo entre valores observados e preditos no modelo SNC-NMLE e dados de crescimento de laranjeiras.

Medida	Árvore				
	1	2	3	4	5
7	0,102	0,053	0,082	0,030	0,007

A árvore 1 apresentou o maior desvio relativo, enquanto a árvore 5 apresentou um valor menor. Observamos que o preditor de observações futuras se mostrou eficiente para esta aplicação, com baixos valores de desvios relativos.

### 4.3.2 Dados do antibiótico cefamandol

Os dados do antibiótico cefamandol, apresentados na Seção 2.2, foram ajustados com o modelo proposto, em que  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top$  e  $\mathbf{R}_i = \sigma^2 \mathbf{I}_{14}$ . O vetor de respostas  $\mathbf{y}_i = (y_{i1}, \dots, y_{i14})^\top$ ,  $i = 1, \dots, 6$ , são as medições da concentração da droga no sangue (em *mcg/ml*) e  $\mathbf{x}_i = \mathbf{t}_i = (t_{i1}, \dots, t_{i14})^\top$  é o vetor que representa as medições do tempo (em minutos) e

$$\mathbf{Z}_i = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i)}{\partial \alpha_2}, \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i)}{\partial \alpha_4} \right] \Bigg|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}}$$

é a matriz de constantes conhecidas, de dimensão  $14 \times 2$ , conforme apresentada em (4.4), com efeitos aleatórios associados aos parâmetros  $\alpha_2$  e  $\alpha_4$ . Essa escolha foi feita por apresentar os menores valores para os critérios AIC e BIC. Assim, o modelo não linear com efeitos mistos que relaciona a concentração de cefamandol e o tempo pode ser representado por

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 6, \quad (4.37)$$

em que  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) = (f(\boldsymbol{\alpha}, t_{i1}), \dots, f(\boldsymbol{\alpha}, t_{i14}))^\top$  é uma função não linear de parâmetros  $\boldsymbol{\alpha}$ , tal que

$$f(\boldsymbol{\alpha}, t_{ij}) = \alpha_1 e^{-t_{ij} e^{\alpha_2}} + \alpha_3 e^{-t_{ij} e^{\alpha_4}}, \quad i = 1, \dots, 6, \quad j = 1, \dots, 14, \quad (4.38)$$

com  $\alpha_2' = \log \alpha_2$ ,  $\alpha_4' = \log \alpha_4$ ,  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top$  e  $\alpha_2 > 0$ ,  $\alpha_4 > 0$ .

Com  $\boldsymbol{\beta} = (d_{11}, d_{12}, d_{22})^\top$ , tal que  $d_{11}$ ,  $d_{12}$  e  $d_{22}$  são elementos da matriz  $\mathbf{D}^{1/2}$ , e  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ , as estimativas de  $\boldsymbol{\theta}$  e os respectivos erros padrões, obtidos por meio da matriz de informação de Fisher observada com o auxílio da biblioteca *numDeriv* (GILBERT; VARADHAN, 2016), estão apresentados na Tabela 8.

Tabela 8 – Estimativa dos parâmetros (Estim.) com erros padrões (E.P.) para os modelos N-SMSN, SN-NLME, ST-NLME, SSL-NLME e SNC-NLME ajustados aos dados do cefamandol.

$\theta$	N-NLME		SN-NLME		ST-NLME		SSL-NLME		SNC-NLME	
	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)
	-	-	-	-	$\nu = 2$		$\nu = 1, 2$		$\mathbf{v} = (0, 34; 0, 2)^T$	
$\alpha_1$	300,78	(31,12)	300,78	(31,14)	583,96	(216,64)	369,14	(46,06)	388,46	(72,61)
$\alpha_2$	-2,06	(0,13)	-2,06	(0,13)	-1,42	(0,18)	-1,86	(0,12)	-1,72	(0,15)
$\alpha_3$	92,44	(11,50)	92,43	(11,49)	81,95	(7,35)	93,89	(12,11)	90,34	(11,44)
$\alpha_4$	-3,95	(0,10)	-3,95	(0,10)	-3,74	(0,08)	-3,66	(0,09)	-3,59	(0,10)
$\sigma^2$	27,72	(4,28)	27,72	(4,27)	189,25	(29,20)	9,67	(1,49)	12,60	(1,95)
$d_{11}$	0,13	(0,10)	0,13	(0,09)	2,32	(0,01)	0,04	(0,05)	0,06	(0,06)
$d_{12}$	0,07	(0,06)	0,07	(0,06)	1,26	(0,15)	0,02	(0,04)	0,05	(0,05)
$d_{22}$	0,08	(0,07)	0,07	(0,07)	0,93	(0,18)	0,04	(0,03)	0,07	(0,02)
$\lambda_1$	-	-	-0,02	(0,40)	-90,51	(26,68)	6,10	(6,05)	-4,387	(15,05)
$\lambda_2$	-	-	0,00	(0,40)	-27,74	(11,86)	-22,63	(6,67)	-76,78	(1,95)

O ajuste sob a distribuição skew-t apresentou o maior erro padrão para a maioria dos parâmetros. Os erros padrões dos parâmetros de assimetria são menores sob a distribuição SN. Na Tabela 9 encontram-se os valores dos critérios AIC e BIC.

Tabela 9 – Valores AIC e BIC para os modelos ajustados aos dados de cefamandol.

Critério	N-NLME	SN-NLME	ST-NLME	SSL-NLME	SNC-NLME
	-	-	$\nu = 2$	$\nu = 1, 2$	$\mathbf{v} = (0, 34; 0, 2)^T$
AIC	565,59	565,59	604,45	551,20	548,12
BIC	587,46	587,46	626,32	573,07	569,99

Os valores dos critérios AIC e BIC são bem próximos, com exceção do ajuste com a distribuição skew-t. O menor valor, porém, ocorreu com a distribuição skew-normal contaminada.

Os gráficos da Figura 21 mostram os perfis ajustados sob a distribuição SNC, em preto, e os dados observados, em cinza, para cada indivíduo relativo aos dados do antiasmático cefamandol.

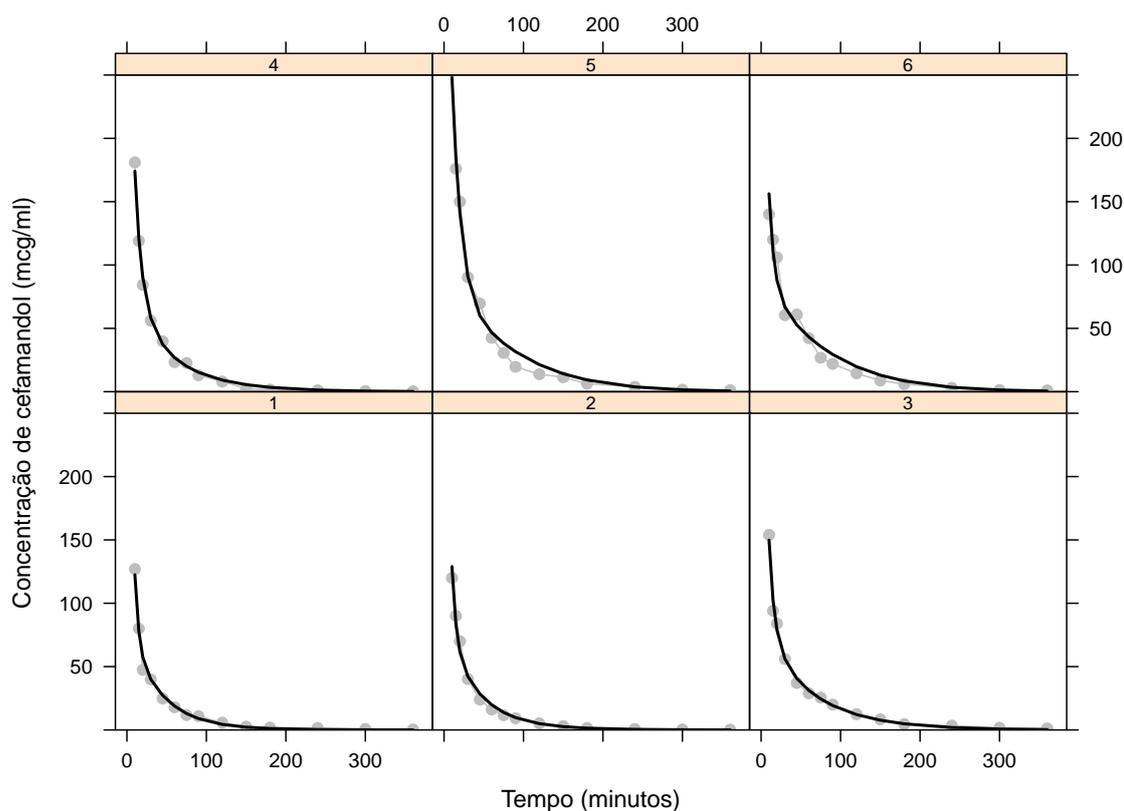


Figura 21 – Perfis ajustados sob o modelo skew-normal contaminada por indivíduo dos dados de cefamandol.

Os resultados do teste da razão de verossimilhanças estão apresentados na Tabela 10 a seguir, no qual estamos interessados em testar se o modelo simétrico é suficiente nos ajustes dos dados.

Tabela 10 – Resultados do teste da razão de verossimilhanças para testar  $H_0 : \boldsymbol{\lambda} = \mathbf{0}$  aplicados aos dados de cefamandol.

Distribuições	$p$ -valor
normal / skew-normal	0,99
t / skew-t	0,11
slash / skew-slash	0,18
normal contaminada / skew-normal contaminada	0,13

Nos testes de hipóteses baseados na estatística da razão de verossimilhanças podemos concluir que os modelos simétricos são os mais adequados sob todas as distribuições. Comparando os resultados dos ajustes sob os modelos normal e skew-normal na Tabela 8, podemos notar que há concordância com o  $p$ -valor obtido no teste da razão de verossimilhanças apresentado na Tabela 10.

### 4.3.3 Dados do antiasmático *theophylline*

Ajustamos os dados do antiasmático *theophylline*, apresentados na Seção 2.3, com o modelo proposto, em que  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top = (lke, lka, lcl)^\top$ ,  $\mathbf{R}_i = \sigma^2 \mathbf{I}_{11}$ . As entradas do vetor  $\mathbf{y}_i = (y_{i1}, \dots, y_{i11})^\top$ ,  $i = 1, \dots, 12$ , são as medições da concentração da droga (em mg/l),  $\mathbf{x}_i = \mathbf{t}_i = (t_{i1}, \dots, t_{i11})^\top$  é o vetor que representa as medições do tempo (em horas) e

$$\mathbf{Z}_i = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i)}{\partial lka}, \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i)}{\partial lcl} \right] \Big|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}}$$

é a matriz de constantes conhecidas, de dimensão  $11 \times 2$ , conforme apresentada em (4.4), com efeitos aleatórios associados aos parâmetros *lka* e *lcl*. Assim, o modelo não linear com efeitos mistos que relaciona a concentração de *theophylline* e o tempo pode ser representada por

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 12, \quad (4.39)$$

em que  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) = (f(\boldsymbol{\alpha}, t_{i1}), \dots, f(\boldsymbol{\alpha}, t_{i11}))^\top$  é uma função não linear de parâmetros  $\boldsymbol{\alpha}$ , tal que

$$f(\boldsymbol{\alpha}, t_{ij}) = \frac{Ds_{ij} e^{lke+lka-lcl}}{e^{lka} - e^{lke}} (e^{-e^{lke}t_{ij}} - e^{-e^{lka}t_{ij}}), \quad i = 1, \dots, 12, \quad j = 1, \dots, 11,$$

em que  $Ds_{ij}$  é a dose de *theophylline* administrada em cada paciente.

Com  $\boldsymbol{\beta} = (d_{11}, d_{12}, d_{22})^\top$ , tal que  $d_{11}$ ,  $d_{12}$  e  $d_{22}$  são elementos da matriz  $\mathbf{D}^{1/2}$ , e  $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ , as estimativas de  $\boldsymbol{\theta}$  e respectivos erros padrões, obtidos por meio da matriz de informação de Fisher observada com o auxílio da biblioteca *numDeriv* (GILBERT; VARADHAN, 2016), estão apresentados na Tabela 11.

Tabela 11 – Estimativa dos parâmetros (Estim.) com erros padrões (E.P.) para os modelos N-SMSN, SN-NLME, ST-NLME, SSL-NLME e SNC-NLME ajustados aos dados de *theophylline*.

	N-NLME		SN-NLME		ST-NLME		SSL-NLME		SNC-NLME	
	-		-		$v = 3.49$		$v = 0.95$		$\mathbf{v} = (0.57; 0.22)^\top$	
$\boldsymbol{\theta}$	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)	Estim.	(E.P.)
<i>lke</i>	-2,52	(0,05)	-2,52	(0,07)	-2,47	(0,05)	-2,47	(0,05)	-2,45	(0,05)
<i>lka</i>	0,45	(0,06)	0,36	(0,07)	0,36	(0,05)	0,38	(0,06)	0,38	(0,06)
<i>lcl</i>	-3,26	(0,04)	-3,02	(0,05)	-2,98	(0,03)	-2,98	(0,03)	-2,97	(0,03)
$\sigma^2$	0,56	(0,07)	0,56	(0,07)	0,30	(0,04)	0,15	(0,02)	0,18	(0,02)
$d_{11}$	0,41	(0,13)	0,43	(0,13)	0,36	(0,12)	0,15	(0,08)	0,19	(0,09)
$d_{12}$	0,00	(0,04)	-0,02	(0,05)	0,01	(0,04)	0,01	(0,03)	0,01	(0,03)
$d_{22}$	0,03	(0,03)	0,08	(0,02)	0,04	(0,01)	0,02	(0,01)	0,03	(0,01)
$\lambda_1$	-	-	6,72	(4,36)	3,57	(4,05)	1,87	(3,10)	2,95	(3,60)
$\lambda_2$	-	-	-21,89	(4,65)	-18,91	(3,95)	-13,78	(3,02)	-16,93	(3,53)

Observa-se que o modelo ajustado sob a distribuição ST apresentou os menores erros padrões para os parâmetros *lke*, *lka* e *lcl*. Nos demais parâmetros os menores erros padrões

ocorreram com a distribuição SSL. As estimativas dos erros padrões obtidas sob a distribuição SSL são próximas das estimativas obtidas com a distribuição SNC. Na Tabela 12 encontram-se os valores das estatísticas AIC e BIC.

Tabela 12 – Valores AIC e BIC para os modelos ajustados aos dados de *theophylline*.

Critério	N-NLME	SN-NLME	ST-NLME	SSL-NLME	SNC-NLME
	-	-	$\nu = 3,49$	$\nu = 0,95$	$\mathbf{v} = (0,57; 0,22)^T$
AIC	385,55	385,15	367,10	370,08	363,96
BIC	411,49	408,09	393,04	396,03	389,91

Os valores dos critérios de informação apresentados na Tabela 12 são próximos. Por meio destes critérios, podemos considerar que os ajustes são equivalentes, com destaque para o ajuste sob a distribuição SNC por apresentar menores valores para estas estatísticas.

Na Figura 22 apresentamos os gráficos dos ajustes por indivíduo com a distribuição SNC para esta aplicação.

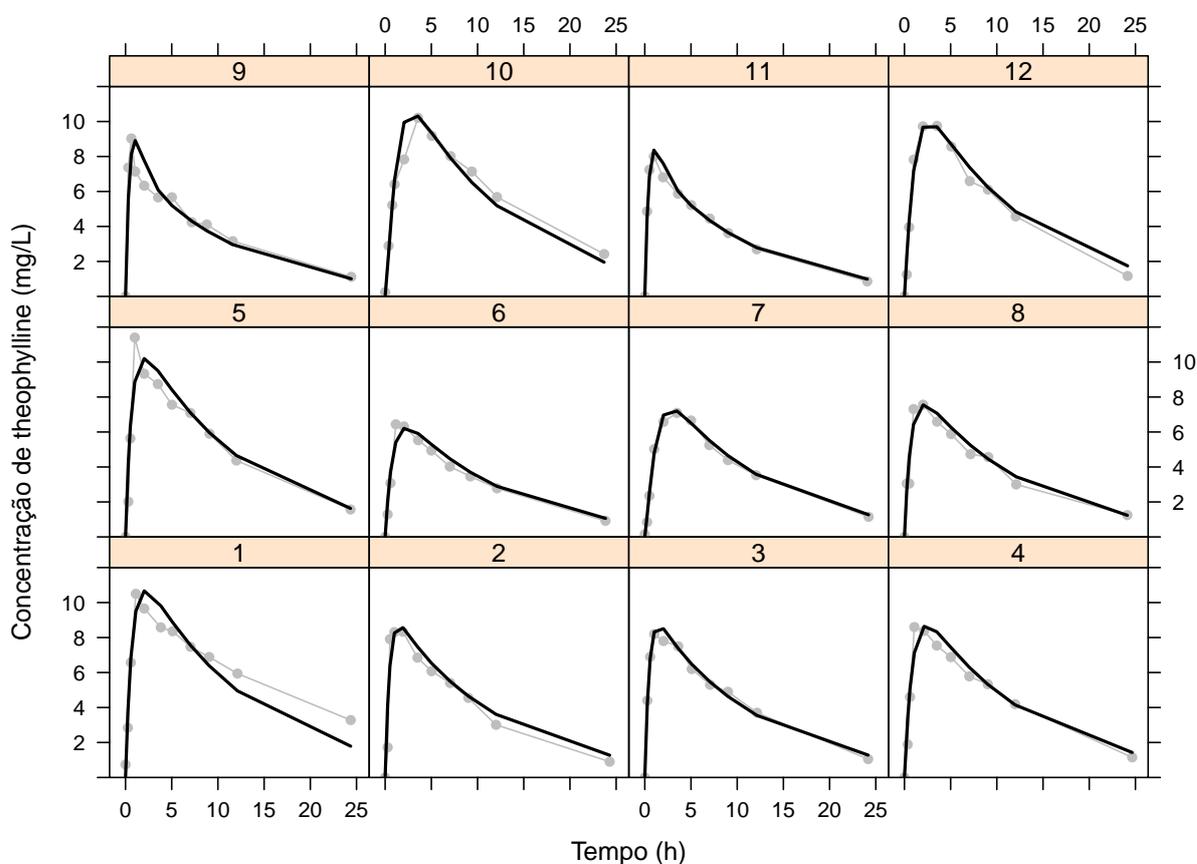


Figura 22 – Perfis ajustados sob o modelo skew-normal contaminada por indivíduo dos dados de *theophylline*.

Os gráficos da Figura 22, gerados com o auxílio da biblioteca *lattice* (SARKAR, 2008),

mostram os perfis ajustados sob a distribuição SNC, em preto, e os dados observados, em cinza, para cada indivíduo que o antiasmático *theophylline* foi administrado.

Apresentamos na Tabela 13 os resultados do TRV para testar se o modelo simétrico é suficiente.

Tabela 13 – Resultados do teste da razão de verossimilhanças para testar  $H_0 : \boldsymbol{\lambda} = \mathbf{0}$  aplicados aos dados de *theophylline*.

Distribuições	<i>p</i> -valor
normal / skew-normal	0,182
t / skew-t	0,044
slash / skew-slash	0,048
normal contaminada / skew-normal contaminada	0,019

Observamos que o modelo assimétrico é mais adequado do que o modelo simétrico para este conjunto de dados, considerando um nível de significância de 5%, sob as distribuições de caudas pesadas skew-t, skew-slash e skew-normal contaminada.

Com a estatística da razão de verossimilhanças, construímos intervalos de confiança de 95% para o parâmetro de assimetria  $\lambda_1$ , apresentados na Tabela 14.

Tabela 14 – Intervalos de confiança aproximados de 95% para o parâmetros de assimetria  $\lambda_1$  obtidos com a estatística de razão de verossimilhanças para os dados de *theophylline*.

Distribuições	I.C. ( $\lambda_1$ )
skew-normal	[1, 12; 13, 57]
skew-t	[0, 07; 7, 47]
skew-slash	[-0, 72; 5, 32]
skew-normal contaminada	[-0, 19; 6, 55]

Nesta aplicação, construímos também intervalos de confiança de 95% usual de Wald para o parâmetro  $\lambda_1$ , dado por  $\hat{\lambda}_1 \pm 1,96 \hat{\text{ep}}(\hat{\lambda}_1)$ , com o objetivo de comparar sua precisão em relação ao intervalo de confiança obtido com a estatística de razão de verossimilhanças e verificamos que a amplitude do intervalo de confiança do tipo Wald para  $\lambda_1$  é maior para todas as distribuições, por exemplo, para a distribuição skew-slash, em que o intervalo de Wald apresentou a menor amplitude, o intervalo de confiança é  $[-4,27; 7,15]$ , com amplitude maior do que o intervalo apresentado na Tabela 14. Isso indica que o intervalo de confiança obtido por meio da estatística de razão de verossimilhanças é mais apropriado. Observamos também que, para esta aplicação e sob o ponto de vista dos intervalos de confiança apresentados, as distribuições com caudas pesadas mostraram-se mais adequadas. Os resultados apresentados na Tabela 14 podem ser observados graficamente na Figura 23, em que a linha horizontal em cada gráfico refere-se ao quantil limite da distribuição qui-quadrado com 1 grau de liberdade, ou seja,  $\chi_{0,95;1}^2$  que vale, aproximadamente, 3,84.

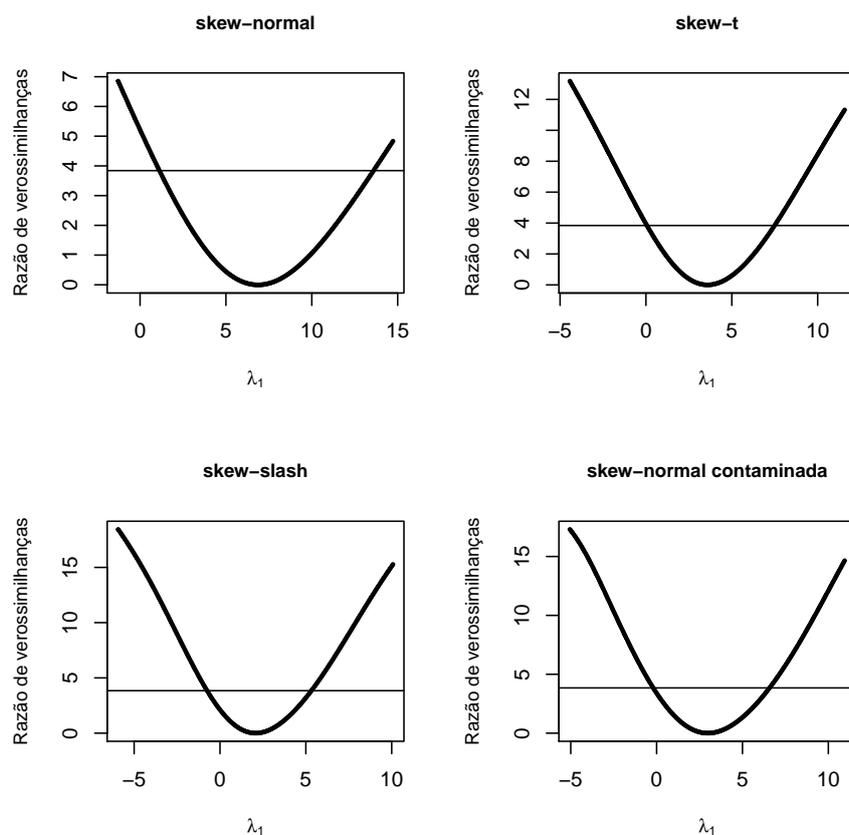


Figura 23 – Razões de verossimilhanças baseadas nas verossimilhanças perfiladas para os dados de *theophylline*.

Para avaliar a adequabilidade da metodologia proposta, as medidas  $\mathbf{y}'_i = (y_{i10}, y_{i11})^\top$ ,  $i = 1, \dots, 12$ , das concentrações de *theophylline* de todos os indivíduos foram consideradas desconhecidas e, assim, preditas como observações futuras, de acordo com (4.32). Assim, o modelo SNC-NLME foi ajustado com apenas 9 medidas de cada indivíduo e as observações futuras  $\hat{\mathbf{y}}_i^+(\hat{\boldsymbol{\theta}}) = (\hat{y}_{i10}^+, \hat{y}_{i11}^+)^\top$ ,  $i = 1, \dots, 12$ , foram obtidas. Os valores preditos  $\hat{\mathbf{y}}_i^+(\hat{\boldsymbol{\theta}})$ , em preto, foram comparados com os valores observados  $\mathbf{y}'_i$ , em cinza, e os resultados são mostrados nos gráficos da Figura 24.

Podemos observar na Figura 24 que existe boa concordância entre os valores preditos e observados, mas com a presença de alguns desvios, principalmente nas medidas dos indivíduos 1, 2, 8, 10 e 12. Assim, calculamos o valor absoluto do desvio relativo por meio da expressão  $|(y_{ij} - \hat{y}_{ij}^+)/y_{ij}|$ ,  $i = 1, \dots, 12$ ,  $j = 10, 11$ , entre os valores observados e preditos. Os resultados estão apresentados na Tabela 15.

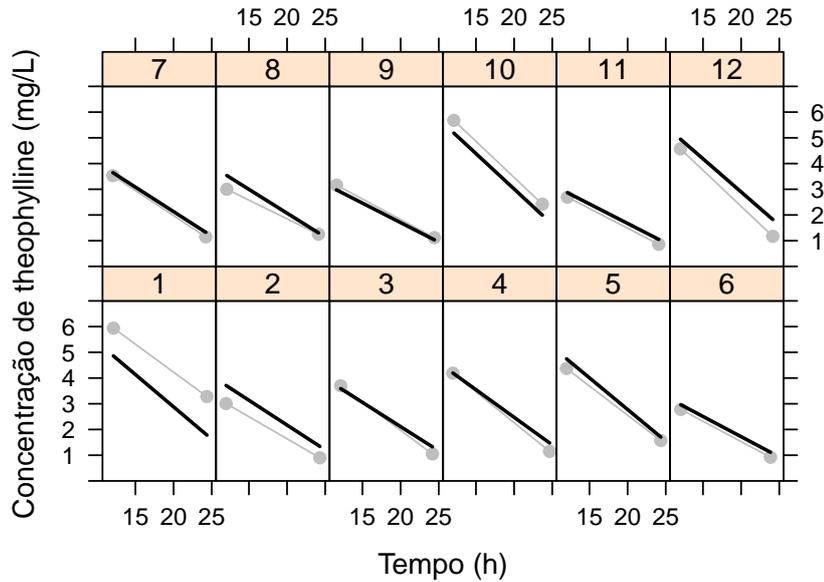


Figura 24 – Valores preditos e observados da concentração de *theophylline* em dois tempos.

Tabela 15 – Valores absolutos do desvio relativo entre valores observados e preditos no modelo SNC-NMLE e dados de *theophylline*.

Medida	Indivíduo											
	1	2	3	4	5	6	7	8	9	10	11	12
10	0,181	0,233	0,029	0,001	0,085	0,065	0,032	0,178	0,058	0,087	0,065	0,082
11	0,456	0,491	0,263	0,284	0,087	0,213	0,153	0,044	0,077	0,176	0,222	0,559

Observamos que o menor desvio relativo ocorreu com o indivíduo 4 e o maior com o indivíduo 12. Baseado na Tabela 15 e de maneira geral, consideramos que o preditor de observações futuras para esta aplicação foi eficiente.

## 4.4 Simulações

Nós realizamos uma série de simulações de Monte Carlo para examinar empiricamente a performance dos estimadores do modelo proposto. No estudo de simulação, temos o objetivo de investigar empiricamente propriedades assintóticas dos estimadores de  $\theta$ . Durante as simulações constatamos que as estimativas dos elementos de  $\alpha$  se aderem a distribuição normal com as condições que são satisfeitas para esses parâmetros dentro do espaço paramétrico, mas não no limite, o que não ocorreu com as estimativas dos parâmetros de dispersão e com os parâmetros de assimetria.

Em cada cenário, geramos 1000 amostras de Monte Carlo e estimamos os elementos do vetor de parâmetros  $\theta$ . Com estas estimativas nós calculamos a média, viés, erro quadrático médio (EQM) de cada elemento de  $\theta$  e a probabilidade de cobertura (PC) do intervalo de

confiança de 95% usual de Wald para cada elemento do vetor de parâmetros  $\boldsymbol{\alpha}$ . As expressões destas medidas empíricas são

$$\text{média} = \sum_{i=1}^{1000} \frac{\hat{\theta}_i}{1000}, \quad (4.40)$$

$$\text{viés} = \sum_{i=1}^{1000} \frac{\hat{\theta}_i - \theta}{1000}, \quad (4.41)$$

$$\text{EQM} = \sum_{i=1}^{1000} \frac{(\hat{\theta}_i - \theta)^2}{1000} \quad e \quad (4.42)$$

$$\text{I.C.}(\alpha_l) = [\hat{\alpha}_l - 1,96 \hat{\text{ep}}(\hat{\alpha}_l); \hat{\alpha}_l + 1,96 \hat{\text{ep}}(\hat{\alpha}_l)], \quad l = 1, \dots, p, \quad (4.43)$$

com  $\theta = \alpha_1, \dots, \alpha_p, \sigma^2, d, \lambda$ , e  $\hat{\text{ep}}(\hat{\alpha}_l), l = 1, \dots, p$ , é o erro padrão estimado de  $\hat{\alpha}_l$ , obtido por meio da matriz de informação de Fisher observada.

Optamos por não realizar simulações com ajustes sob a distribuição skew-slash pelo fato de ser custoso no ponto de vista computacional.

#### • Primeiro cenário

Por meio das bibliotecas *mixsmsn* (PRATES; CABRAL; LACHOS, 2013) e *mvtnorm* (GENZ; BRETZ, 2009), geramos dados artificiais sob as distribuições t multivariada para os erros e skew-Student-t univariada para os efeitos aleatórios, com a curva

$$y_{ij} = f(\boldsymbol{\alpha}, t_{ij}) + Z_j b_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (4.44)$$

em que  $\boldsymbol{\varepsilon}_i \sim T_{m_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i}, 3)$  e  $b_i \sim ST(0, d, \lambda, 3)$ . Os regressores  $t_{ij}$  não dependem do perfil  $i$ , assim  $t_{ij} = t_j$ , conseqüentemente,  $Z_{ij} = Z_j$ . Adotamos a seguinte função não linear

$$f(\boldsymbol{\alpha}, t_j) = \alpha_1 e^{-t_j e^{\alpha_2}} + \alpha_3 e^{-t_j e^{\alpha_4}}. \quad (4.45)$$

Este cenário de simulação foi baseado no conjunto de dados do antibiótico cefamandol, apresentado na Seção 2.2.

A matriz de constantes  $Z_j$  considerada para este estudo de simulação, conforme (4.4), tem a forma

$$Z_j = \left. \frac{\partial f(\boldsymbol{\alpha}, t_j)}{\partial \alpha_3} \right|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}},$$

desta forma o efeito aleatório está associado ao parâmetro  $\alpha_3$ .

Ajustamos os modelos SMSN-NLME com as distribuições skew-normal, skew-t e skew-normal contaminada. Os valores verdadeiros considerados para os parâmetros são  $\boldsymbol{\alpha} =$

$(2.8, 0.9, 0.6, -1.1)^\top$ ,  $\sigma^2 = 0.06$ ,  $d = 0.03$  e  $\lambda = 0$  e 2. Os valores dos regressores  $t_j$  foram mantidos constantes para  $\lambda = 0$  e 2. Os valores de  $\mathbf{v}$  são supostamente conhecidos. Os resultados das simulações estão apresentados nas Tabelas 16 e 17.

Tabela 16 – Média, viés, EQM e CP dos modelos ajustados sob as distribuições SN, ST e SNC por meio de 1000 amostras simuladas de Monte Carlo com  $n = 5$  e  $m_i = 7$ .

$\theta$	SN-NLME				ST-NLME				SNC-NLME			
	média	viés	EQM	PC(%)	média	viés	EQM	PC(%)	média	viés	EQM	PC(%)
$n = 5, m_i = 7$ e $\lambda = 0$												
$\alpha_1$	2,904	0,104	0,434	98,4	2,953	0,153	0,488	96,3	2,791	-0,009	0,401	97,6
$\alpha_2$	1,048	0,148	0,102	99,4	1,070	0,170	0,118	97,7	1,036	0,136	0,074	99,7
$\alpha_3$	0,781	0,181	0,155	99,4	0,814	0,214	0,196	95,1	0,840	0,240	0,174	98,5
$\alpha_4$	-1,002	0,098	0,243	97,8	-1,018	0,082	0,267	96,4	-0,938	0,162	0,219	97,4
$\sigma^2$	0,094	0,034	0,003	-	0,070	0,010	0,001	-	0,047	-0,013	0,001	-
$d$	0,086	0,056	0,025	-	0,069	0,039	0,015	-	0,045	0,015	0,005	-
$\lambda$	0,099	0,099	1,315	-	-0,214	-0,214	3,692	-	-0,158	-0,158	2,525	-
$n = 5, m_i = 7$ e $\lambda = 2$												
$\alpha_1$	2,933	0,133	0,444	98,3	2,967	0,167	0,474	96,3	2,831	0,031	0,355	97,8
$\alpha_2$	1,023	0,123	0,097	99,3	1,039	0,139	0,109	97,7	1,010	0,110	0,069	99,5
$\alpha_3$	0,807	0,207	0,172	99,2	0,838	0,238	0,203	95,1	0,864	0,264	0,177	97,5
$\alpha_4$	-1,035	0,065	0,228	97,6	-1,072	0,028	0,251	96,4	-1,011	0,089	0,178	97,9
$\sigma^2$	0,094	0,034	0,003	-	0,069	0,009	0,001	-	0,047	-0,013	0,001	-
$d$	0,059	0,029	0,010	-	0,049	0,019	0,007	-	0,030	0,000	0,002	-
$\lambda$	1,054	-0,946	2,113	-	0,993	-1,007	4,261	-	0,756	-1,244	3,956	-

Tabela 17 – Média, viés, EQM e CP dos modelos ajustados sob as distribuições SN, ST e SNC por meio de 1000 amostras simuladas de Monte Carlo com  $n = 6$  e  $m_i = 11$ .

$\theta$	SN-NLME				ST-NLME				SNC-NLME			
	média	viés	EQM	PC(%)	média	viés	EQM	PC(%)	média	viés	EQM	PC(%)
$n = 6, m_i = 11$ e $\lambda = 0$												
$\alpha_1$	2,951	0,151	0,381	98,3	2,981	0,181	0,421	97,1	2,900	0,100	0,343	97,3
$\alpha_2$	1,011	0,111	0,095	99,0	1,020	0,012	0,104	97,7	0,999	0,099	0,074	99,0
$\alpha_3$	0,705	0,105	0,117	98,1	0,741	0,141	0,145	94,1	0,735	0,135	0,123	97,3
$\alpha_4$	-1,126	-0,026	0,270	97,1	-1,090	0,010	0,256	95,1	-1,062	0,038	0,222	96,6
$\sigma^2$	0,107	0,047	0,031	-	0,088	0,028	0,002	-	0,059	-0,001	0,001	-
$d$	0,087	0,057	0,062	-	0,077	0,047	0,016	-	0,050	0,020	0,007	-
$\lambda$	0,042	0,042	2,325	-	-0,241	-0,241	4,011	-	-0,100	-0,100	2,971	-
$n = 6, m_i = 11$ e $\lambda = 2$												
$\alpha_1$	2,960	0,160	0,348	96,6	2,979	0,179	0,409	96,6	2,921	0,121	0,340	96,6
$\alpha_2$	1,001	0,101	0,098	97,2	0,996	0,096	0,105	97,2	0,975	0,075	0,078	98,0
$\alpha_3$	0,748	0,148	0,133	93,9	0,769	0,169	0,158	93,9	0,771	0,171	0,139	95,8
$\alpha_4$	-1,143	-0,043	0,247	95,3	-1,123	-0,023	0,256	95,3	-1,106	-0,006	0,204	97,2
$\sigma^2$	0,107	0,047	0,005	-	0,088	0,028	0,002	-	0,059	-0,001	0,001	-
$d$	0,072	0,042	0,049	-	0,057	0,027	0,012	-	0,037	0,007	0,005	-
$\lambda$	1,235	-0,765	3,735	-	1,048	-0,952	4,118	-	0,905	-1,095	3,760	-

### • Segundo cenário

Neste segundo cenário, também geramos, por meio das bibliotecas *mixsmsn* (PRATES; CABRAL; LACHOS, 2013) e *mvtnorm* (GENZ; BRETZ, 2009), dados artificiais sob as distri-

buições t multivariada para os erros e skew-Student-t univariada para os efeitos aleatórios, porém usamos outra curva como mostrado a seguir.

$$y_{ij} = f(\boldsymbol{\alpha}, t_{ij}) + Z_{ij}b_i + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (4.46)$$

em que  $\boldsymbol{\varepsilon}_i \sim T_{m_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i}, 3)$  e  $b_i \sim ST(0, d, \lambda, 3)$ . Adotamos a seguinte função não linear

$$f(\boldsymbol{\alpha}, t_{ij}) = \frac{Ds_{ij} e^{\alpha_1 + \alpha_2 - \alpha_3}}{e^{\alpha_2} - e^{\alpha_1}} (e^{-e^{\alpha_1} t_{ij}} - e^{-e^{\alpha_2} t_{ij}}), \quad i = 1, \dots, n, \quad j = 1, \dots, m_i.$$

Este cenário de simulação foi baseado no conjunto de dados do antiasmático *theophylline*, apresentado na Seção 2.3. A matriz de constantes  $Z_{ij}$  considerada para este estudo de simulação, conforme (4.4), tem a forma

$$Z_{ij} = \left. \frac{\partial f(\boldsymbol{\alpha}, t_{ij})}{\partial \alpha_3} \right|_{\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}},$$

assim, o efeito aleatório está associado ao parâmetro  $\alpha_3$ .

Ajustamos os modelos SMSN-NLME com as distribuições skew-normal, skew-t e skew-normal contaminada. Os valores verdadeiros considerados para os parâmetros são  $\boldsymbol{\alpha} = (-2, 5, 0, 4, -3, 1)^\top$ ,  $\sigma^2 = 1$ ,  $d = 0,5$  e  $\lambda = 0$  e 2. Os valores de  $\mathbf{v}$  são supostamente conhecidos. Os valores dos regressores  $t_{ij}$  e de  $Ds_{ij}$  foram mantidos constantes para  $\lambda = 0$  e 2. Os resultados das simulações estão apresentados nas Tabelas 19 e 18.

Tabela 18 – Média, viés, EQM e PC dos modelos ajustados sob as distribuições SN, ST e SNC por meio de 1000 amostras simuladas de Monte Carlo com  $n = 9$  e  $m_i = 6$ .

$\theta$	SN-NLME				ST-NLME				SNC-NLME			
	média	viés	EQM	PC(%)	média	viés	EQM	PC(%)	média	viés	EQM	PC(%)
$n = 9, m_i = 6$ e $\lambda = 0$												
$\alpha_1$	-2,501	-0,001	0,410	96,1	-2,490	0,010	0,402	93,1	-2,488	0,012	0,054	88,4
$\alpha_2$	0,389	-0,011	0,071	93,9	0,390	-0,010	0,090	87,0	0,398	-0,002	0,106	84,4
$\alpha_3$	-3,084	0,016	0,014	99,4	-3,050	0,050	0,041	95,1	-3,059	0,041	0,048	90,3
$\sigma^2$	2,016	1,016	1,893	-	1,382	0,382	0,483	-	0,605	-0,395	0,265	-
$d$	1,292	0,792	2,489	-	0,796	0,296	0,593	-	0,421	-0,079	0,243	-
$\lambda$	-0,248	-0,248	7,363	-	-0,205	-0,205	2,411	-	-0,149	-0,149	5,821	-
$n = 9, m_i = 6$ e $\lambda = 2$												
$\alpha_1$	-2,482	0,018	0,035	95,9	-2,471	0,029	0,043	91,2	-2,486	0,014	0,074	87,1
$\alpha_2$	0,376	-0,024	0,079	92,7	0,386	-0,014	0,073	88,1	0,380	-0,020	0,137	83,7
$\alpha_3$	-3,064	0,036	0,019	94,6	-3,020	0,080	0,041	93,7	-3,006	0,094	0,117	79,2
$\sigma^2$	2,026	1,026	1,959	-	1,329	0,329	0,474	-	0,618	-0,382	0,251	-
$d$	1,173	0,673	1,945	-	0,679	0,179	0,375	-	0,355	-0,145	0,231	-
$\lambda$	4,485	2,485	19,301	-	3,881	1,881	9,614	-	4,130	2,130	14,386	-

Tabela 19 – Média, viés, EQM e PC dos modelos ajustados sob as distribuições SN, ST e SNC por meio de 1000 amostras simuladas de Monte Carlo com  $n = 12$  e  $m_i = 11$ .

$\theta$	SN-NLME				ST-NLME				SNC-NLME			
	média	viés	EQM	PC(%)	média	viés	EQM	PC(%)	média	viés	EQM	PC(%)
$v = 3, 0$												
$n = 12, m_i = 11$ e $\lambda = 0$												
$\alpha_1$	-2,527	-0,027	0,001	100	-2,509	-0,009	0,011	95,2	-2,504	-0,004	0,013	92,3
$\alpha_2$	0,565	0,165	0,027	100	0,409	0,009	0,017	92,4	0,402	0,002	0,020	90,2
$\alpha_3$	-3,110	-0,010	0,001	100	-3,117	-0,017	0,001	100	-3,110	-0,010	0,001	99,9
$\sigma^2$	1,487	0,487	0,237	-	1,662	0,662	0,813	-	0,620	-0,380	0,265	-
$d$	1,256	0,756	1,837	-	1,046	0,546	1,006	-	0,379	-0,121	0,222	-
$\lambda$	-0,027	-0,027	0,367	-	-0,011	-0,011	0,460	-	0,000	0,000	0,427	-
$n = 12, m_i = 11$ e $\lambda = 2$												
$\alpha_1$	-2,508	-0,008	0,011	94,1	-2,514	-0,014	0,012	95,3	-2,507	-0,007	0,013	93,3
$\alpha_2$	0,410	0,010	0,018	96,8	0,401	0,001	0,018	92,5	0,401	0,001	0,017	92,5
$\alpha_3$	-3,117	-0,017	0,001	100	-3,105	-0,005	0,002	99,6	-3,110	-0,010	0,002	99,0
$\sigma^2$	2,169	1,169	2,140	-	1,608	0,608	0,771	-	0,617	-0,383	0,276	-
$d$	1,193	0,693	1,340	-	0,744	0,244	0,413	-	0,339	-0,161	0,191	-
$\lambda$	3,205	1,205	6,960	-	3,410	1,410	6,724	-	3,306	1,306	7,506	-

Nas Tabelas 16-19 podemos observar que, na maioria dos casos, o viés e o EQM diminuem quando o tamanho da amostra aumenta, indicando uma convergência assintótica para os valores reais dos parâmetros à medida que o tamanho da amostra aumenta, com  $\lambda = 0$  e também com  $\lambda = 2$ , além de confirmar a flexibilidade do modelo SMSN-NLME para ajustar dados simétricos e assimétricos. O EQM de  $\lambda$  é menor sob a distribuição SN no primeiro cenários e sob a skew-t na maioria dos casos do segundo cenário. Os componentes aleatórios têm menor EQM sob a distribuição SNC em ambos cenários. Para todo  $\lambda$ , a PC referente aos elementos do vetor de parâmetros  $\alpha$  aproxima-se do valor nominal, confirmando que a distribuição assintótica de  $(\hat{\alpha}_l - \alpha_l)$ ,  $l = 1, 2, 3, 4$ , sob condições que são preenchidas para estes parâmetros dentro do espaço paramétrico, mas não no limite, é  $N(0, I(\alpha_l)^{-1})$ , onde  $I(\alpha_l)$ ,  $l = 1, 2, 3, 4$ , é um elemento da matriz de informação Fisher observada. Além disso, o parâmetro  $\alpha_2$  apresentou a maior PC em todas as distribuições no primeiro cenário e  $\alpha_3$  no segundo cenário. Em geral, nos dois cenários, os ajustes sob a distribuição SN apresentaram PCs maiores.

---

## ANÁLISE DE DIAGNÓSTICO

---

Técnicas de análise de diagnóstico têm o objetivo de avaliar a qualidade de um modelo. Em modelagem estatística, a verificação da coerência das suposições é de fundamental importância e recebe contribuições de diversos autores, tanto na detecção de observações atípicas e a verificação de seu impacto nos ajustes, quanto no desenvolvimento de medidas corretivas. Assim, vamos propor algumas técnicas de análise de diagnóstico no modelo não linear com efeitos mistos sob a classe SMSN de distribuições. Faremos uma breve discussão sobre alguns pontos a respeito de análise de resíduos e diagnóstico de influência, considerando técnicas de influência local e global nos ajustes com os conjuntos de dados apresentados no Capítulo 2. De maneira geral, verificamos que ao introduzir perturbação nos dados na direção de observações aberrantes, as estimativas das componentes aleatórias são mais sensíveis, assim como quando perturbamos essas componentes.

### 5.1 Análise de resíduos

A análise dos resíduos compreende a um conjunto de técnicas baseadas na leitura e interpretação dos resíduos, e são utilizadas para auxiliar na verificação da validade das suposições de um modelo de regressão e, conseqüentemente, analisar a aderência e a adequação da distribuição considerada na formulação do modelo. Existem várias técnicas de diagnóstico que propõem distâncias para analisar resíduos. A forma mais simples e imediata de se desenvolver um estudo para analisar resíduos está baseado nos resíduos ordinários, obtidos por meio do modelo marginal, por exemplo de (4.1), dados por

$$\hat{r}_{ij} = y_{ij} - f(\hat{\boldsymbol{\alpha}}, x_{ij}), \quad i = 1, \dots, n, \quad j = 1, \dots, m_i. \quad (5.1)$$

No caso de modelos mistos, pode-se utilizar os resíduos condicionados aos efeitos

aleatórios, definidos por

$$\widehat{rc}_{ij} = y_{ij} - f(\widehat{\boldsymbol{\alpha}}, x_{ij}) - \mathbf{Z}_{ij}^{\top} \widehat{\mathbf{b}}_i, \quad i = 1, \dots, n, \quad j = 1, \dots, m_i. \quad (5.2)$$

Outro resíduo que vamos analisar e que vem sendo bastante utilizado em trabalhos científicos é o *resíduo quantílico aleatorizado*, proposto por Dunn e Smith (DUNN; SMYTH, 1996), que convergem para a distribuição normal padrão se os parâmetros do modelo são estimados de forma consistente. O resíduo quantílico aleatorizado, considerando o modelo marginal, é dado pela expressão

$$\widehat{rq}_{ij} = \Phi^{-1} \left\{ G_Y(y_{ij} | f(\widehat{\boldsymbol{\alpha}}, x_{ij}), \widehat{\boldsymbol{\theta}}) \right\} \quad i = 1, \dots, n, \quad j = 1, \dots, m_i, \quad (5.3)$$

em que  $\Phi$  e  $G_Y$  são as fda da distribuição normal padrão e da distribuição considerada no ajuste, respectivamente.

Na análise de resíduos vamos considerar, entre outras medidas de comparação, a distância de Mahalanobis do modelo marginal dado o efeito aleatório, que é uma medida bastante conhecida na literatura, obtida com a expressão

$$d_i = (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}, \mathbf{x}_i) - \mathbf{Z}_{ij}^{\top} \widehat{\mathbf{b}}_i)^{\top} \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}, \mathbf{x}_i) - \mathbf{Z}_{ij}^{\top} \widehat{\mathbf{b}}_i), \quad i = 1, \dots, n. \quad (5.4)$$

## 5.2 Diagnóstico de influência

Para verificar se as estimativas do modelo proposto são resistentes a pequenas perturbações, vamos utilizar algumas técnicas de diagnóstico de influência utilizando medidas apropriadas. Essas técnicas são ferramentas que avaliam o grau de sensibilidade das estimativas dos parâmetros de um modelo de regressão quando alguma perturbação ocorre nos dados ou nas suposições. Se o modelo ajustado não apresentar uma boa descrição dos dados observados, uma perturbação pode causar efeitos desproporcionais nos seus resultados. Portanto, é importante o estudo sobre a robustez de um modelo estatístico, a fim de se obter boa concordância entre os dados observados e os dados estimados.

### 5.2.1 Influência local

Uma das formas mais utilizadas e mais simples para detectar influência de observações sobre um modelo ajustado por meio de perturbações, é baseada na exclusão de casos, que é conhecido como um método de diagnóstico de influência global e consiste em analisar um ou mais modelos ajustados, após a exclusão de um subconjunto de observações. O estudo de influência local, introduzido por Cook (COOK, 1986), é uma abordagem mais moderna, baseada na análise das curvaturas das seções normais de uma superfície na vizinhança de um determinado

ponto. O estudo de influência local consiste, portanto, em observar o comportamento de alguma medida particular de influência com pequenas perturbações nos dados ou no modelo. A seguir, apresentamos a forma de abordagem de influência local apresentada por Zhu e Lee (ZHU; LEE, 2001) que é fundamentada no trabalho de Cook (COOK, 1986).

Considerando o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \gamma, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$ , com  $\gamma = \sigma^2$ , e  $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$  a esperança condicional da função de log-verossimilhança dos dados completos, conforme (4.10), os elementos da matriz hessiana do modelo SMSN-NLME são obtidos pelas expressões

$$\ddot{Q}(\boldsymbol{\theta}) = \sum_{i=1}^n \ddot{Q}_i(\boldsymbol{\theta}) = \sum_{i=1}^n \ddot{Q}_{1i}(\boldsymbol{\theta}_1) + \sum_{i=1}^n \ddot{Q}_{2i}(\boldsymbol{\theta}_2), \quad (5.5)$$

em que

$$\ddot{Q}_{1i}(\boldsymbol{\theta}_1) = -\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top} \quad e \quad (5.6)$$

$$\ddot{Q}_{2i}(\boldsymbol{\theta}_2) = -\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top}, \quad (5.7)$$

com  $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \sigma^2)^\top$  e  $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}, \boldsymbol{\lambda}^\top)^\top$ .

Assim,

$$\ddot{Q}_i(\boldsymbol{\theta}) = -\frac{\partial^2 Q_i(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{Q}_{1i}(\boldsymbol{\theta}_1) & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_{2i}(\boldsymbol{\theta}_2) \end{pmatrix}. \quad (5.8)$$

Observamos que com o particionamento do vetor de parâmetros  $\boldsymbol{\theta}$ , a matriz  $\ddot{Q}(\boldsymbol{\theta})$  é bloco diagonal.

As expressões dos elementos da matriz hessiana do modelo SMSN-NLME são mostradas no Apêndice C.

Considerando o vetor de perturbação  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_g)^\top$  de dimensão  $g \times 1$ , restritas a um conjunto aberto  $\Omega \in \mathbb{R}^g$ , e seja  $\ell(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)$  a função de log-verossimilhança dos dados completos do modelo perturbado, assumindo a existência do vetor de não perturbação  $\boldsymbol{\omega}_0$ , tal que  $\ell(\boldsymbol{\theta}, \boldsymbol{\omega}_0|\mathbf{y}_c) = \ell(\boldsymbol{\theta}|\mathbf{y}_c)$  para todo  $\boldsymbol{\theta}$ . Denotamos por  $\hat{\boldsymbol{\theta}}(\boldsymbol{\omega})$  como a função que maximiza  $Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = E[\ell(\boldsymbol{\theta}, \boldsymbol{\omega}|\mathbf{y}_c)|\mathbf{y}, \hat{\boldsymbol{\theta}}]$ . Com o objetivo de verificar a existência de influência das perturbações nas estimativas de máxima verossimilhança de  $\hat{\boldsymbol{\theta}}$ , considera-se o afastamento da função de log-verossimilhança

$$f_Q(\boldsymbol{\omega}) = 2 \left[ Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) - Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) \right]. \quad (5.9)$$

Em influência local, estudamos o comportamento de  $f_Q(\boldsymbol{\omega})$  na vizinhança do vetor  $\boldsymbol{\omega}_0$ . No procedimento, é selecionada uma direção  $\mathbf{d}$ , tal que  $\|\mathbf{d}\| = 1$ , e então estudar o gráfico de  $f_Q(\boldsymbol{\omega}_0 + a\mathbf{d})$  contra  $a$ , com  $a \in \mathbb{R}$ . Como  $f_Q(\boldsymbol{\omega}_0) = 0$ , então  $f_Q(\boldsymbol{\omega}_0 + a\mathbf{d})$  tem um mínimo local em  $a = 0$ . Cada linha projetada pode ser caracterizada pela curvatura normal  $C_{\mathbf{d}}(\boldsymbol{\theta})$  em torno de

$a = 0$  e a sugestão é considerar a direção  $\mathbf{d}_{\text{máx}}$  que corresponde a máxima curvatura  $C_{\mathbf{d}_{\text{máx}}}(\boldsymbol{\theta})$ . O gráfico de  $\mathbf{d}_{\text{máx}}$  contra a ordem das observações pode revelar as observações que têm influência desproporcional sob pequenas perturbações em  $f_Q(\boldsymbol{\omega})$ . Cook (COOK, 1986) mostrou que a curvatura normal na direção unitária  $\mathbf{d}$  tem a forma

$$C_{\mathbf{d}}(\boldsymbol{\theta}) = 2|\mathbf{d}^\top \nabla_{\boldsymbol{\theta}}^\top [\ddot{Q}(\boldsymbol{\theta})]^{-1} \nabla_{\boldsymbol{\theta}} \mathbf{d}|, \quad (5.10)$$

em que a matriz  $\nabla_{\boldsymbol{\theta}}$  é apresentada em (5.15) e  $\ddot{Q}(\boldsymbol{\theta})$  é a matriz hessiana, conforme (5.5).

Obtemos  $C_{\mathbf{d}_{\text{máx}}}(\boldsymbol{\theta})$  como sendo o maior autovalor da matriz  $\nabla_{\boldsymbol{\theta}}^\top [\ddot{Q}(\boldsymbol{\theta})]^{-1} \nabla_{\boldsymbol{\theta}}$  e  $\mathbf{d}_{\text{máx}}$  é o autovetor correspondente.

Para facilitar a identificação de observações influentes, uma vez que a curvatura normal  $C_{\mathbf{d}}(\boldsymbol{\theta})$  pode assumir qualquer valor e não é invariante sob uma transformação uniforme de escala, Poon e Poon (POON; POON, 1999) definiram uma curvatura normal conformal  $B_{\mathbf{d}}(\boldsymbol{\theta})$ , em uma direção unitária  $\mathbf{d}$ , tal que  $0 \leq |B_{\mathbf{d}}(\boldsymbol{\theta})| \leq 1$ , que é invariante sob reparametrizações conformais, cuja expressão é

$$B_{\mathbf{d}}(\boldsymbol{\theta}) = \frac{-\mathbf{d}^\top \nabla_{\boldsymbol{\theta}}^\top [\ddot{Q}(\boldsymbol{\theta})]^{-1} \nabla_{\boldsymbol{\theta}} \mathbf{d}}{\sqrt{\text{tr}(\nabla_{\boldsymbol{\theta}}^\top [\ddot{Q}(\boldsymbol{\theta})]^{-1} \nabla_{\boldsymbol{\theta}})^2}}. \quad (5.11)$$

Desta forma, um autovetor  $\mathbf{d}_i$  é considerado  $q$ -influyente se sua medida de curvatura normal conformal  $B_{\mathbf{d}_i}$  é tal que

$$|B_{\mathbf{d}}(\boldsymbol{\theta})| \geq \frac{q}{\sqrt{g}}, \quad (5.12)$$

onde  $q \geq 0$  e  $g$  é a dimensão de  $\boldsymbol{\omega}$ .

Denominamos de vetor de perturbação básica  $\mathbf{u}_i$  de dimensão  $g \times 1$  com o valor 1 na  $i$ -ésima posição e zero nas demais posições (ESCOBAR; MEEKER, 1992), assim uma possibilidade para avaliar a influência da curvatura normal conformal é considerar o gráfico de índices de  $\mathbf{B}$ , definido por

$$\mathbf{B} = (B_1, \dots, B_g)^\top = (B_{\mathbf{u}_1}(\boldsymbol{\theta}), \dots, B_{\mathbf{u}_g}(\boldsymbol{\theta}))^\top,$$

tal que  $\sum_{i=1}^g B_i^2 = 1$ . Em um caso mais geral se  $\{\mathbf{u}_1, \dots, \mathbf{u}_g\}$  é um conjunto de autovetores ortonormais, então  $\sum_{i=1}^g B_{\mathbf{u}_i}^2 = 1$ .

Sejam os autovalores  $\eta_1, \dots, \eta_g$  ordenados em ordem decrescente, padronizados e em valor absoluto, cujos autovetores correspondentes são  $\mathbf{d}_1, \dots, \mathbf{d}_g$  da matriz  $\nabla_{\boldsymbol{\theta}}^\top [\ddot{Q}(\boldsymbol{\theta})]^{-1} \nabla_{\boldsymbol{\theta}}$ , ou seja, de forma que

$$\eta_1 > \dots > \eta_g \quad \text{e} \quad \sum_{i=1}^g \eta_i^2 = 1. \quad (5.13)$$

Segundo Poon e Poon (POON; POON, 1999), podemos calcular a contribuição agregada  $\mathbf{m}(q)$  de todos os  $k$  autovetores considerados  $q$ -influentes. O  $j$ -ésimo elemento do vetor  $\mathbf{m}(q)$  é obtido por meio da expressão

$$m(q)_j = \sqrt{\sum_{i=1}^k \eta_i a_{ij}^2}, \quad (5.14)$$

em que  $a_{ij}$  é a  $j$ -ésima entrada do autovetor  $\mathbf{d}_i$ ,  $i = 1, \dots, k$  e  $j = 1, \dots, g$ .

A escolha de  $q$  deve ser feita de forma apropriada para avaliar a contribuição agregada dos  $k$  autovetores de interesse. Em particular, para  $q = 0$  é possível avaliar a contribuição agregada de todos os autovetores e sempre existe  $q$  suficientemente grande para avaliar apenas a influência na direção do autovetor correspondente ao maior autovalor, ou seja, a influência de  $\mathbf{d}_{\text{máx}}$ , assim

$$\eta_{\text{máx}} = \eta_1 \geq \dots \geq \eta_k \geq \frac{q}{\sqrt{g}} > \eta_{k+1} \geq \dots \geq \eta_g \geq 0.$$

Quando avaliamos a influência da observação devido à contribuição agregada de todos os autovetores correspondentes aos autovalores não nulos, estamos avaliando  $B_i = B_{\mathbf{u}_i}$ , conforme discutido em Escobar e Meeker (ESCOBAR; MEEKER, 1992), assim, as entradas do vetor  $\mathbf{m}(0)$  indicariam a influência das observações devido à contribuição agregada de todos os autovetores. Portanto analisar  $\mathbf{m}(0)$  é equivalente a analisar  $\mathbf{B}$ .

Segundo Russo (RUSSO, 2010), uma discussão recorrente na literatura é a obtenção de uma regra geral para julgar a influência de uma observação específica a partir das medidas discutidas apresentadas neste Capítulo, mas alguns autores desenvolveram propostas para julgar as magnitudes dessas medidas, como veremos a seguir.

Seja  $\bar{B}$  a média de  $\mathbf{B}$  e, respectivamente,  $\bar{m}(0)$  e  $sm(0)$  a média e o desvio padrão de  $\{m(0)_j : j = 1, \dots, g\}$ , Poon e Poon (POON; POON, 1999) propuseram o uso de  $2\bar{B}$  e  $\sqrt{2}\bar{m}(0)$  como pontos de cortes para  $\mathbf{B}$  e  $\mathbf{m}(0)$ . Porém Lee e Xu (LEE; XU, 2004) propuseram  $\bar{m}(0) + c^*sm(0)$  como marca de referência, em que  $c^*$  é uma constante positiva arbitrária.

Consideramos a seguir alguns esquemas de perturbação no diagnóstico de influência local para comparar a sensibilidade das estimativas dos parâmetros e para identificar possíveis observações influentes nos ajustes com o modelo SMSN-NLME.

### 5.2.1.1 Esquemas de perturbação

Apresentamos a seguir a matriz  $\nabla_{\boldsymbol{\theta}}$ , para cada esquema de perturbação no modelo SMSN-NLME, cujos elementos têm a forma

$$\nabla_{\boldsymbol{\theta}} = \frac{\partial^2 Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\omega}^T} \Bigg|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}, \boldsymbol{\omega} = \boldsymbol{\omega}_0}, \quad (5.15)$$

em que  $\widehat{\boldsymbol{\theta}}$  a estimativa de máxima verossimilhança de  $\boldsymbol{\theta}$  e  $\boldsymbol{\omega}_0$  é o vetor de não perturbação.

### • Ponderação de casos

Nesse esquema de perturbação, multiplicamos o valor esperado do logaritmo da função de verossimilhança por uma pequena perturbação, como podemos verificar a seguir

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \boldsymbol{\omega}E[\ell(\boldsymbol{\theta}|\mathbf{y}_c)|\mathbf{y}, \widehat{\boldsymbol{\theta}}] = \sum_{i=1}^n \omega_i Q_{1i}(\boldsymbol{\theta}_1|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^n \omega_i Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}}), \quad (5.16)$$

em que o  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_g)^\top$  é o vetor de perturbação, com  $0 \leq \omega_i \leq 1$ ,  $i = 1, \dots, g$ . O vetor de não perturbação nesse esquema é dado por  $\boldsymbol{\omega}_0 = (1, \dots, 1)^\top$ .

Para este esquema de perturbação, encontramos as seguintes expressões para os elementos da matriz  $\nabla_{\boldsymbol{\theta}}$

$$\nabla_{\boldsymbol{\alpha}, i} = \frac{1}{\sigma^2} \left[ \mathbf{J}_i^\top \{ \widehat{u}_i \widehat{\mathbf{r}}_i - \mathbf{Z}_i(\widehat{u\mathbf{b}})_i \} \right], \quad (5.17)$$

$$\nabla_{\sigma^2, i} = -\frac{m_i}{2\sigma^2} + \frac{1}{2\sigma^4} \left[ \widehat{u}_i \widehat{\mathbf{r}}_i^\top \widehat{\mathbf{r}}_i - 2\widehat{\mathbf{r}}_i^\top \mathbf{Z}_i(\widehat{u\mathbf{b}})_i + \text{tr} \left( \mathbf{Z}_i(\widehat{u\mathbf{b}\mathbf{b}^\top})_i \mathbf{Z}_i^\top \right) \right], \quad (5.18)$$

$$\nabla_{\beta_r, i} = -\frac{1}{2} \text{tr} \left( \boldsymbol{\Gamma}^{-1} \dot{\boldsymbol{\Gamma}}_{\beta_r} \{ \mathbf{I}_q - \boldsymbol{\Gamma}^{-1} \widehat{\mathbf{N}}_i \} \right) + \boldsymbol{\delta}^\top \dot{\mathbf{F}}_r \boldsymbol{\Gamma}^{-1} \left[ (\widehat{u\mathbf{b}})_i - (\widehat{u^2})_i \boldsymbol{\Delta} \right] \quad \text{e} \quad (5.19)$$

$$\nabla_{\lambda_s, i} = -\frac{1}{2} \text{tr} \left( \boldsymbol{\Gamma}^{-1} \dot{\boldsymbol{\Gamma}}_{\lambda_s} \{ \mathbf{I}_q - \boldsymbol{\Gamma}^{-1} \widehat{\mathbf{N}}_i \} \right) + \dot{\boldsymbol{\delta}}_s^\top \mathbf{F} \boldsymbol{\Gamma}^{-1} \left[ (\widehat{u\mathbf{b}})_i - (\widehat{u^2})_i \boldsymbol{\Delta} \right], \quad (5.20)$$

onde a matriz  $\mathbf{F} = \mathbf{D}^{1/2}$ , ou seja,  $\mathbf{D} = \mathbf{F}^2$ ,  $\dot{\boldsymbol{\Gamma}}_{\beta_r} = \frac{\partial \boldsymbol{\Gamma}}{\partial \beta_r} = \dot{\mathbf{F}}_r \mathbf{F} + \mathbf{F} \dot{\mathbf{F}}_r - \dot{\mathbf{F}}_r \boldsymbol{\delta} \boldsymbol{\delta}^\top \mathbf{F} - \mathbf{F} \boldsymbol{\delta} \boldsymbol{\delta}^\top \dot{\mathbf{F}}_r$ ,  $\dot{\mathbf{F}}_r = \frac{\partial \mathbf{F}}{\partial \beta_r}$ ,  $r = 1, \dots, \dim(\boldsymbol{\beta})$ ,  $\dot{\boldsymbol{\Gamma}}_{\lambda_s} = \frac{\partial \boldsymbol{\Gamma}}{\partial \lambda_s} = -\mathbf{F} \left( \frac{\partial \boldsymbol{\delta} \boldsymbol{\delta}^\top}{\partial \lambda_s} \right) \mathbf{F}$ ,  $\dot{\boldsymbol{\delta}}_s = \frac{\partial \boldsymbol{\delta}}{\partial \lambda_s}$ ,  $s = 1, \dots, q$ ,  $\widehat{\mathbf{N}}_i = (\widehat{u\mathbf{b}\mathbf{b}^\top})_i - (\widehat{u\mathbf{b}})_i \boldsymbol{\Delta}^\top - \boldsymbol{\Delta} (\widehat{u\mathbf{b}})_i^\top + (\widehat{u^2})_i \boldsymbol{\Delta} \boldsymbol{\Delta}^\top$ ,  $\mathbf{I}_q$  é a matriz identidade  $q \times q$ .

### • Perturbação na variável resposta

Seja o vetor de respostas  $(\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$  e considerando a inclusão de uma perturbação aditiva a cada grupo, teremos  $\mathbf{y}_i(\boldsymbol{\omega}) = \mathbf{y}_i^\top + \boldsymbol{\omega}_i \mathbf{1}_{m_i}$ , em que  $\mathbf{1}_{m_i}$  é um vetor de uns de dimensão  $m_i \times 1$ ,  $i = 1, \dots, n$ . A função perturbada é obtida com a expressão

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n Q_{1i}(\boldsymbol{\theta}_1, \omega_i|\widehat{\boldsymbol{\theta}}) + \sum_{i=1}^n Q_{2i}(\boldsymbol{\theta}_2|\widehat{\boldsymbol{\theta}}), \quad (5.21)$$

onde podemos encontrar  $Q_{1i}(\boldsymbol{\theta}_1, \omega_i|\widehat{\boldsymbol{\theta}})$  substituindo em (4.11)  $\mathbf{y}_i$  por  $\mathbf{y}_i(\boldsymbol{\omega})$ . Sob esse esquema de perturbação com o vetor  $\boldsymbol{\omega}_0 = \mathbf{0} \in \mathbb{R}^g$ , de A matriz  $\nabla_{\boldsymbol{\theta}}$  tem os seguintes elementos

$$\nabla_{\boldsymbol{\alpha},i} = \frac{\widehat{u}_i}{\sigma^2} \mathbf{J}_i^\top \mathbf{1}_{m_1}, \quad (5.22)$$

$$\nabla_{\sigma^2,i} = \frac{1}{2\sigma^4} \left[ \widehat{u}_i \mathbf{r}_i - \mathbf{Z}_i(\widehat{u\mathbf{b}})_i \right]^\top \mathbf{1}_{m_i}, \quad (5.23)$$

$$\nabla_{\boldsymbol{\beta}_r,i} = \mathbf{0} \quad \text{e} \quad (5.24)$$

$$\nabla_{\boldsymbol{\lambda}_r,i} = \mathbf{0}. \quad (5.25)$$

$$(5.26)$$

### • Perturbação na matriz de dispersão dos efeitos aleatórios

A perturbação na matriz escala tem por objetivo identificar possíveis observações influentes ou padrões adicionais de heteroscedasticidade entre os grupos e pode ser introduzida considerando que o efeito aleatório

$$\mathbf{b}_i \sim SMSN_q(\mathbf{0}, \frac{\mathbf{D}}{\omega_i}, \boldsymbol{\lambda}; H). \quad (5.27)$$

Sob esse esquema de perturbação, em que  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_g) \in \mathbb{R}^g - \{0\}$ , os resultados a seguir são obtidos quando  $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = (1, \dots, 1)$ .

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega} | \widehat{\boldsymbol{\theta}}) = \sum_{i=1}^n Q_{1i}(\boldsymbol{\theta}_1 | \widehat{\boldsymbol{\theta}}) + \sum_{i=1}^n Q_{2i}(\boldsymbol{\theta}_2, \omega_i | \widehat{\boldsymbol{\theta}}). \quad (5.28)$$

A expressão de  $Q_{2i}(\boldsymbol{\theta}_2, \omega_i | \widehat{\boldsymbol{\theta}})$  é obtida substituindo em (4.12)  $\boldsymbol{\Delta}$  por  $\boldsymbol{\Delta}(\boldsymbol{\omega}) = \omega^{-1/2} \boldsymbol{\Delta}$  e  $\boldsymbol{\Gamma}$  por  $\boldsymbol{\Gamma}(\boldsymbol{\omega}) = \omega^{-1} \boldsymbol{\Gamma}$ . A matriz  $\nabla_{\boldsymbol{\theta}}$  tem elementos dados por

$$\nabla_{\boldsymbol{\alpha},i} = \mathbf{0}, \quad (5.29)$$

$$\nabla_{\sigma^2,i} = \mathbf{0}, \quad (5.30)$$

$$\nabla_{\boldsymbol{\beta}_r,i} = \frac{1}{2} \text{tr} \left( \boldsymbol{\Gamma}^{-1} \dot{\boldsymbol{\Gamma}}_{\boldsymbol{\beta}_r} \boldsymbol{\Gamma}^{-1} \{ (\widehat{u\mathbf{b}\mathbf{b}^\top})_i - (\widehat{u\mathbf{b}})_i \boldsymbol{\Delta}^\top \} \right) + \frac{1}{2} \boldsymbol{\delta}^\top \dot{\mathbf{F}}_r \boldsymbol{\Gamma}^{-1} (\widehat{u\mathbf{b}})_i \quad \text{e} \quad (5.31)$$

$$\nabla_{\boldsymbol{\lambda}_r,i} = \frac{1}{2} \text{tr} \left( \boldsymbol{\Gamma}^{-1} \dot{\boldsymbol{\Gamma}}_{\boldsymbol{\lambda}_s} \boldsymbol{\Gamma}^{-1} \{ (\widehat{u\mathbf{b}\mathbf{b}^\top})_i - (\widehat{u\mathbf{b}})_i \boldsymbol{\Delta}^\top \} \right) + \frac{1}{2} \dot{\boldsymbol{\delta}}_s^\top \mathbf{F} \boldsymbol{\Gamma}^{-1} (\widehat{u\mathbf{b}})_i. \quad (5.32)$$

### • Perturbação na matriz de dispersão dos erros

Para identificar possíveis observações influentes na matriz escala dos erros consideramos o esquema de perturbação por meio do modelo

$$\boldsymbol{\varepsilon}_i \sim SMN_{m_i}(\mathbf{0}, \frac{\sigma^2}{\omega_i} \mathbf{I}_{m_i}; H). \quad (5.33)$$

Sob esse esquema de perturbação, em que  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_g) \in \mathbb{R}^g - \{0\}$ , os resultados a seguir são obtidos quando  $\boldsymbol{\omega} = \boldsymbol{\omega}_0 = (1, \dots, 1)$ . Temos que a função perturbada é dada por

$$Q(\boldsymbol{\theta}, \boldsymbol{\omega}|\hat{\boldsymbol{\theta}}) = \sum_{i=1}^n Q_{1i}(\boldsymbol{\theta}_1, \omega_i|\hat{\boldsymbol{\theta}}) + \sum_{i=1}^n Q_{2i}(\boldsymbol{\theta}_2|\hat{\boldsymbol{\theta}}), \quad (5.34)$$

onde a expressão de  $Q_{1i}(\boldsymbol{\theta}_1, \omega_i|\hat{\boldsymbol{\theta}})$  é obtida substituindo em (4.11)  $\sigma^2$  por  $\sigma^2(\boldsymbol{\omega}) = \boldsymbol{\omega}^{-1}\sigma^2$ . A matriz  $\nabla_{\boldsymbol{\theta}}$  tem elementos dados por

$$\nabla_{\boldsymbol{\alpha},i} = \frac{1}{\sigma^2} \left[ \mathbf{J}_i^\top \{ \hat{u}_i \mathbf{r}_i - \mathbf{Z}_i(\widehat{u\mathbf{b}})_i \} \right], \quad (5.35)$$

$$\nabla_{\sigma^2,i} = \frac{1}{2\sigma^4} \left[ \hat{u}_i \mathbf{r}_i^\top \mathbf{r}_i - 2\mathbf{r}_i^\top \mathbf{Z}_i(\widehat{u\mathbf{b}})_i + \text{tr} \left( \mathbf{Z}_i(\widehat{u\mathbf{b}\mathbf{b}^\top})_i \mathbf{Z}_i^\top \right) \right], \quad (5.36)$$

$$\nabla_{\boldsymbol{\beta}_r,i} = 0 \quad \text{e} \quad (5.37)$$

$$\nabla_{\boldsymbol{\lambda}_r,i} = 0. \quad (5.38)$$

Para ilustrar as metodologias apresentadas neste trabalho realizamos estudos de diagnóstico, por meio de análise de resíduos, influência local e global em três conjuntos de dados apresentados no Capítulo 2 com o objetivo de verificar a presença de observações destoantes e/ou influentes.

Na análise de resíduos, vamos analisar gráficos das distâncias de Mahalanobis, de acordo com (5.4), dos resíduos ordinários condicionados (5.2) e dos resíduos quantílicos aleatorizados que possuem distribuição normal padrão, como apresentado em (5.3), por meio do gráfico de resíduos e índices e envelopes simulados (ATKINSON, 1981), construídos com o auxílio das bibliotecas *mixsmsn* (PRATES; CABRAL; LACHOS, 2013), *sn* (AZZALINI, 2018) e *hnp* (MORAL; HINDE; DEMÉTRIO, 2017). Uma maneira detalhada para construir gráficos de envelopes simulados é apresentada no Apêndice C.

### 5.2.2 Dados de crescimento de laranjeiras

Os dados de crescimento de laranjeiras estão apresentados na Seção 2.1.

#### • Análise de resíduos

Observamos nos gráficos da Figura 25 que as medições referentes às árvores 4 e 5 possuem as maiores distâncias de Mahalanobis.

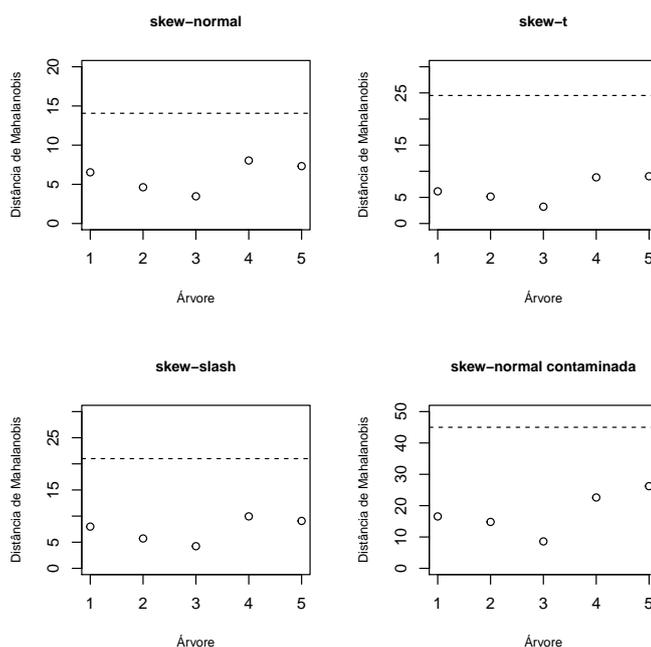


Figura 25 – Distâncias de Mahalanobis dos modelos ajustados aos dados de crescimento de laranjeiras.

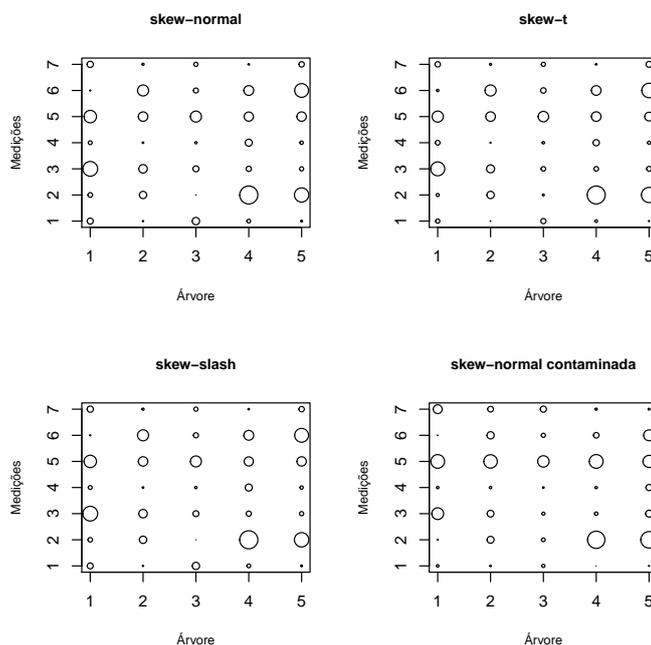


Figura 26 – Gráficos dos resíduos condicionados das medições por árvore dos dados de crescimento de laranjeiras.

Os gráficos da Figura 26 mostram resíduos condicionados de cada medição e por indivíduo, em que os diâmetros das circunferências são proporcionais aos valores absolutos dos resíduos, desta forma é possível verificar que as medições com maiores resíduos, em valores absolutos, são a segunda medição das árvores 4 e 5, a terceira medição da árvore 1, a quinta medição de todas as árvores e a sexta medição da árvore 5.

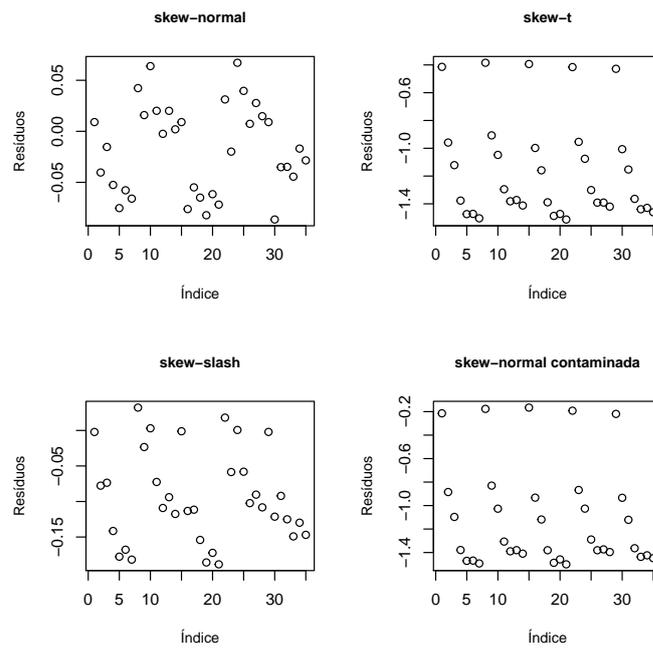


Figura 27 – Gráficos dos resíduos quantílicos aleatorizados para os dados das laranjeiras.

Nos gráficos dos resíduos quantílicos aleatorizados na Figura 27, notamos que nenhuma observação está fora do intervalo de confiança de 95%, ou seja, fora do intervalo  $[-1,96; 1,96]$ , porém sob as distribuições ST, SSL e SNC há tendência para assimetria negativa.

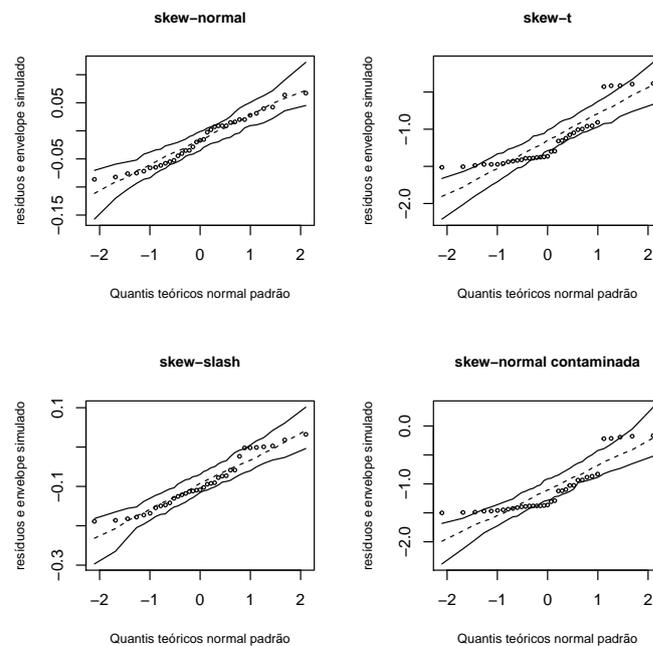


Figura 28 – Gráficos dos resíduos quantílicos aleatorizados e envelopes simulados para os dados das laranjeiras.

Os gráficos de resíduos e envelopes simulados da Figura 28 mostram relativa semelhança

entre os gráficos sob as distribuições SN e SSL e entre as distribuições ST e SNC. No ponto de vista de análise de resíduos, os ajustes mais adequados, por ter resíduos dentro das bandas de confiança, são os ajustes sob as distribuições SN e SSL. É possível identificar, também no gráfico de resíduos e envelopes simulados, que os resíduos têm tendência para assimetria negativa.

### • Influência local

Considerando os esquemas de perturbação apresentados neste Capítulo, construímos gráficos do vetor  $\mathbf{m}(0)$  para verificar a influência devido à contribuição agregada de todos os autovetores nos ajustes sob cada distribuição pertencente à classe SMSN. Os pontos de corte adotados em todos os casos foram obtidos por meio de  $\bar{m}(0) + c^*sm(0)$ , em que  $c^* = 0,5$ .

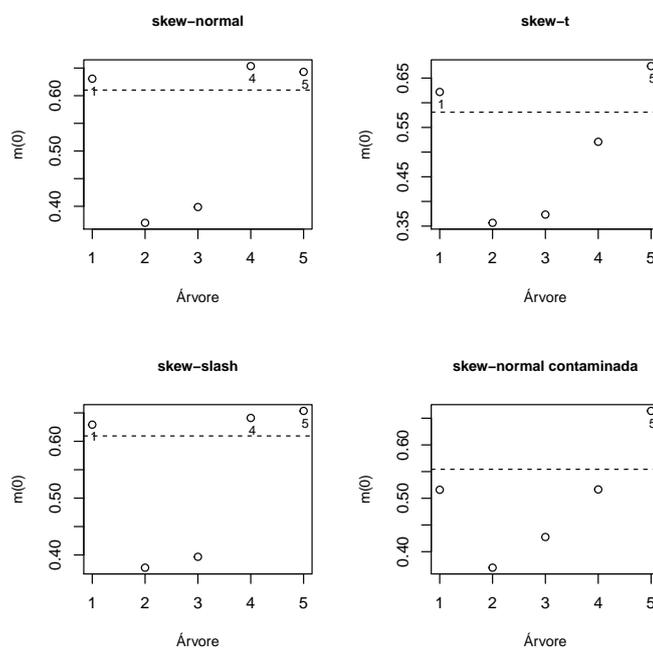


Figura 29 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação ponderação de casos com  $c^* = 0,5$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 29, sob o esquema de perturbação ponderação de casos, verificamos que a árvore 5 possui medições influentes sob todas as distribuições, em conformidade com os gráficos das distâncias de Mahalanobis apresentados anteriormente. Sendo que, sob as distribuições SN e SSL, podemos considerar como influentes também as medições das árvores 1 e 4.

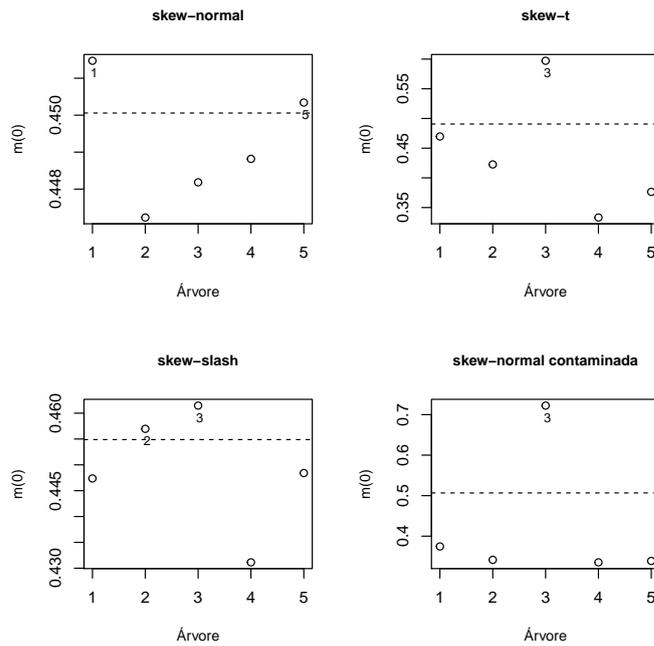


Figura 30 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação na variável resposta com  $c^* = 0,5$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 30, sob o esquema de perturbação na variável resposta, verificamos que a árvore 3 possui as medições mais influentes sob as distribuições ST, SSL e SNC. Sob a distribuição SN, a árvore 1 possui as medições mais influentes, seguida da árvore 5 que é uma árvore com uma das maiores distâncias de Mahalanobis.

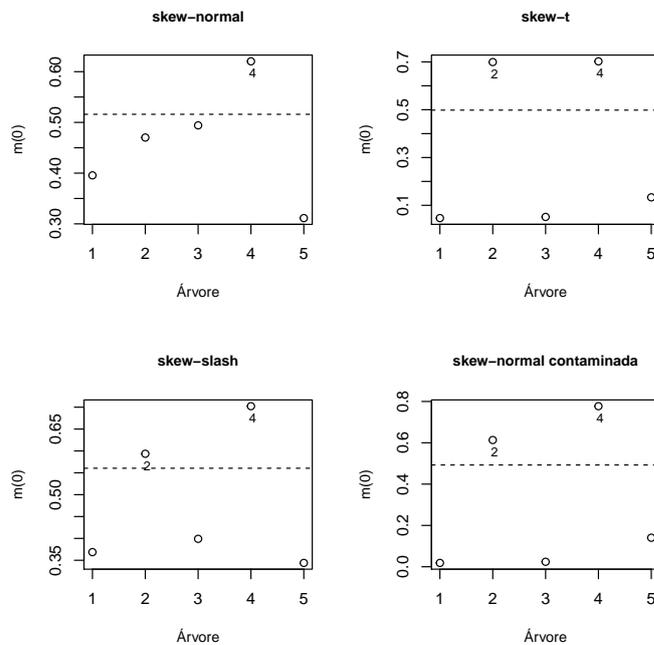


Figura 31 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação na matriz de dispersão dos efeitos aleatórios com  $c^* = 0,5$ .

Na Figura 31, em que são mostrados os gráficos de índices de  $\mathbf{m}(0)$  sob o esquema

de perturbação na matriz de dispersão dos efeitos aleatórios, verificamos que as árvores 2 e 4 possuem as medições mais influentes.

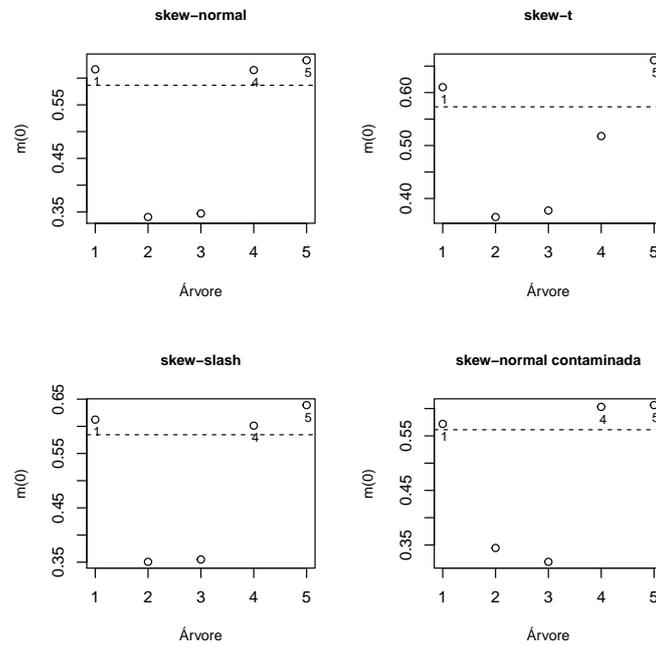


Figura 32 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação na matriz de dispersão dos erros com  $c^* = 0,5$ .

Na Figura 32, em que são mostrados os gráficos de índices de  $\mathbf{m}(0)$  sob o esquema de perturbação na matriz de dispersão dos erros, verificamos que as árvores 1, 4 e 5 possuem as medições mais influentes.

Fizemos algumas perturbações, como propostas na Seção 5.2.1.1, para avaliar o grau de sensibilidade das estimativas dos parâmetros, por meio dos percentuais de mudança nestas estimativas. As perturbações foram introduzidas na direção de uma determinada árvore  $i$ ,  $i = 1, \dots, n$ , considerando, por exemplo, o esquema de perturbação na variável resposta para obter  $\hat{\boldsymbol{\theta}}_{\omega_0 + a\mathbf{d}_i}$ , com  $a$  arbitrário e  $\mathbf{d}_i$  com valor 1 nas posições relativas a árvore  $i$  e zero nas demais posições. Nas análises apresentadas notamos que os efeitos fixos não sofreram grandes mudanças. Omitimos os gráficos de  $d_{12}$ , pois suas estimativas apresentaram comportamento semelhante às estimativas de  $d_{11}$  ou  $d_{22}$ . Os percentuais de mudanças foram obtidos com a expressão  $(\hat{\boldsymbol{\theta}}_{\omega_0 + a\mathbf{d}_i} - \hat{\boldsymbol{\theta}}) / \hat{\boldsymbol{\theta}}$ .

Sob o esquema de perturbação na matriz de dispersão dos erros, a perturbação foi realizada na direção da árvore 4, conforme Figura 33, e podemos observar que o parâmetro  $d_{22}$  e sob a distribuição SNC, sofreu as maiores alterações. O valor de  $a$  adotado pertence ao intervalo  $[-0,2; 0,2]$ . Constatamos que, neste esquema de perturbação, as estimações são bastante sensíveis, visto que o valor de  $a$  adotado é pequeno.

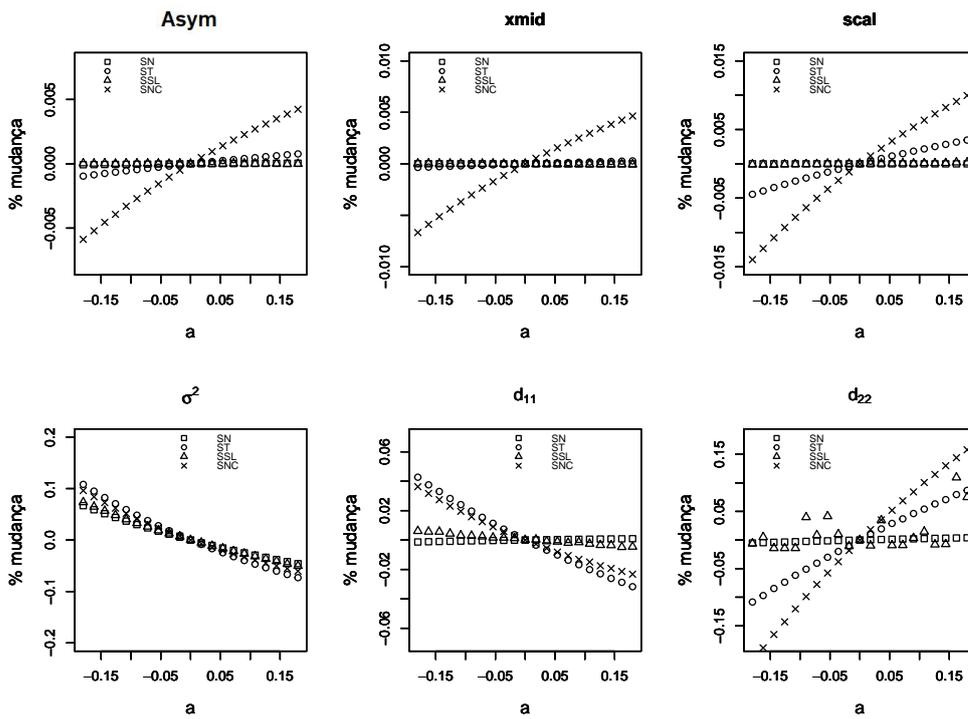


Figura 33 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na matriz de dispersão dos erros na direção da árvore 4.

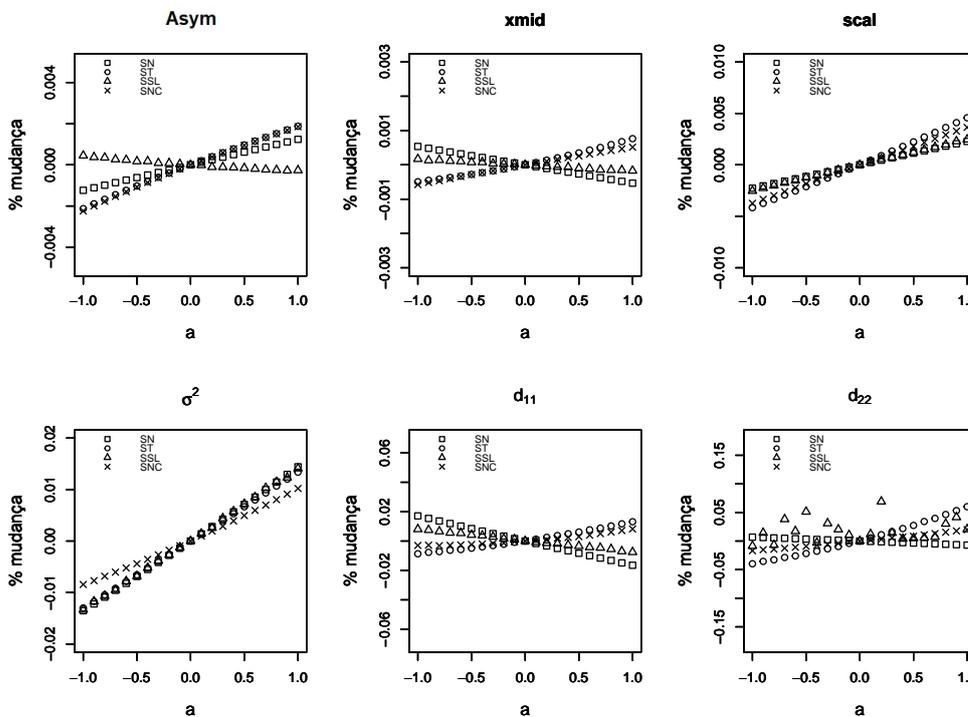


Figura 34 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na variável resposta na direção da árvore 1.

Podemos observar nos gráficos da Figura 34, com perturbação na direção da árvore 1 e no esquema de perturbação na variável resposta, que o parâmetro  $d_{22}$  também sofreu as maiores

alterações, porém tais mudanças não passaram de 5% em valor absoluto, e que o valor de  $a$  adotado pertence ao intervalo  $[-1; 1]$ . Sob o este esquema de perturbação, as estimações não sofreram mudanças expressivas, mesmos para maiores valores de  $a$ , em relação ao esquema anterior. Nas estimações do parâmetro  $d_{22}$  e sob a distribuição SSL, percebemos uma não monotonicidade.

### • Exclusão de casos

Para verificar a influência global de observações, realizamos a exclusão de algumas laranjeiras. As estimativas dos efeitos fixos não sofreram mudanças maiores do que 25%, em valor absoluto, e o parâmetro de dispersão  $\sigma^2$  não sofreu mudança maior do que 10%, em valor absoluto. Os demais parâmetros sofreram maiores mudanças e os resultados destas mudanças estão apresentados na Tabela 20 a seguir.

Tabela 20 – Percentuais de mudanças nas estimativas dos parâmetros quando árvores foram excluídas para os dados de crescimento de laranjeiras.

Distribuição	Árvore excluída			
	1	2	4	5
	$d_{11}$			
skew-normal	-2	-8	-27	24
skew-t	-51	-35	-42	7
skew-slash	-8	71	37	45
skew-normal contaminada	-55	-21	-33	14
	$d_{12}$			
skew-normal	3	6	30	-25
skew-t	88	8	63	11
skew-slash	13	-86	-340	-46
skew-normal contaminada	94	5	48	-20
	$d_{22}$			
skew-normal	-4	-3	-30	23
skew-t	-97	24	-76	-27
skew-slash	-16	96	-97	-29
skew-normal contaminada	-99	12	-221	26
	$\lambda_1$			
skew-normal	116	116	-150	16
skew-t	-98	74	-21	6
skew-slash	-43	99	-97	-45
skew-normal contaminada	-98	74	221	-1
	$\lambda_2$			
skew-normal	-22	-11	22	33
skew-t	99	-82	49	12
skew-slash	106	-99	99	40
skew-normal contaminada	99	-81	-156	-5

Verificamos na Tabela 20 que, com a exclusão das medições da árvore 4, as estimações dos parâmetros apresentam as maiores alterações, sendo esta a árvore mais influente globalmente, entre as árvores estudadas. O parâmetro  $\lambda_1$ , com ajustes sob a distribuição skew-normal, é o parâmetro mais sensível a retirada de observações. Notamos também que na análise de resíduos nos gráficos das distâncias de Mahalanobis e dos resíduos condicionados, Figuras 25 e 26, a

árvore 4 possui as maiores distâncias de Mahalanobis.

### 5.2.3 Dados de antibiótico cefamandol

Os dados do antibiótico cefamandol estão apresentados na Seção 2.2.

#### • Análise de resíduos

Nos gráficos da Figura 35, podemos observar que as medições referentes aos indivíduos 5 e 6 possuem as maiores distâncias de Mahalanobis, sendo que, no ajuste sob a distribuição skew-normal, o indivíduo 6 é possivelmente aberrante.

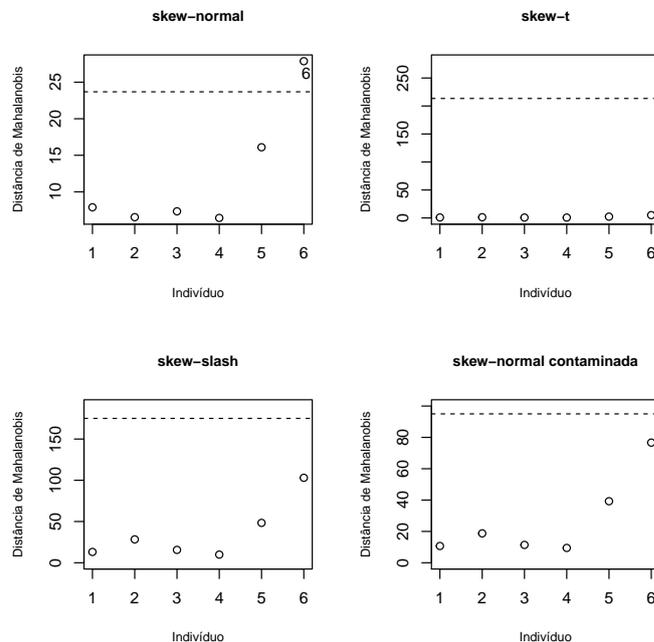


Figura 35 – Distâncias de Mahalanobis dos modelos ajustados aos dados do antibiótico cefamandol.

Os gráficos da Figura 36 mostram resíduos condicionados de cada medição e por indivíduo. Os diâmetros das circunferências são proporcionais aos valores absolutos dos resíduos, assim as medições com resíduos mais destoantes são a terceira medição dos indivíduos 1, 2 e 6, a segunda medição dos indivíduos 3 e 5, a primeira medição dos indivíduos 2 e 6, a quinta medição dos indivíduos 5 e 6 e a oitava medição do indivíduo 5.

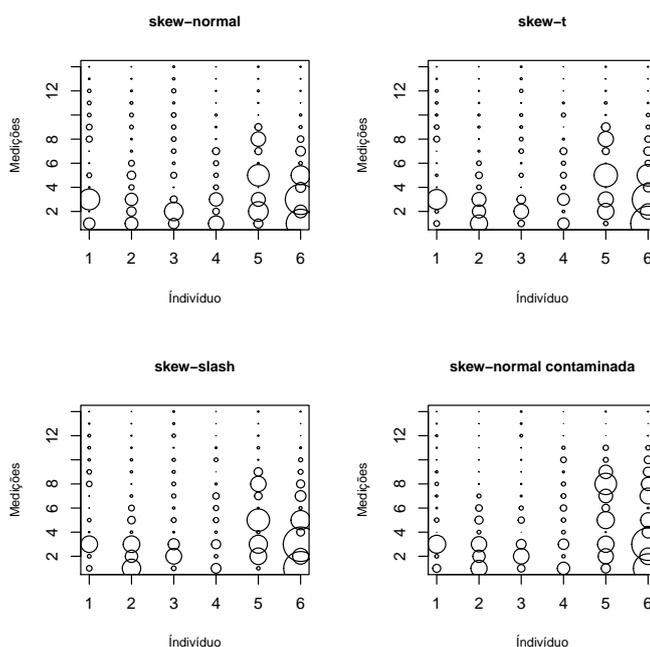


Figura 36 – Gráfico de resíduos condicionados das medições por indivíduos dos dados de cefamandol.

Nos gráficos dos resíduos da Figura 37, consideramos como observações atípicas os resíduos com valores absolutos maiores que o quantil  $z_{0,975} = 1,96$  da distribuição normal padrão e, para este conjunto de dados, não houve tais observações. Podemos notar uma tendência a assimetria negativa.

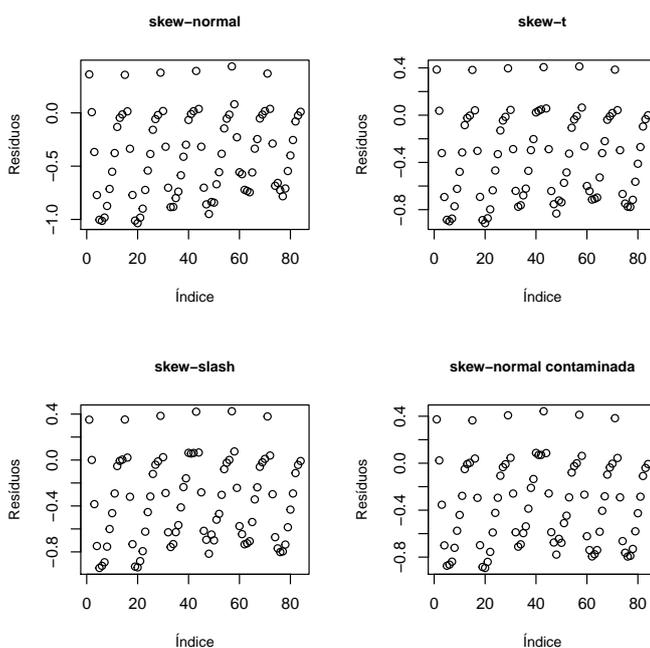


Figura 37 – Gráficos dos resíduos quantílicos aleatorizados para os dados de cefamandol.

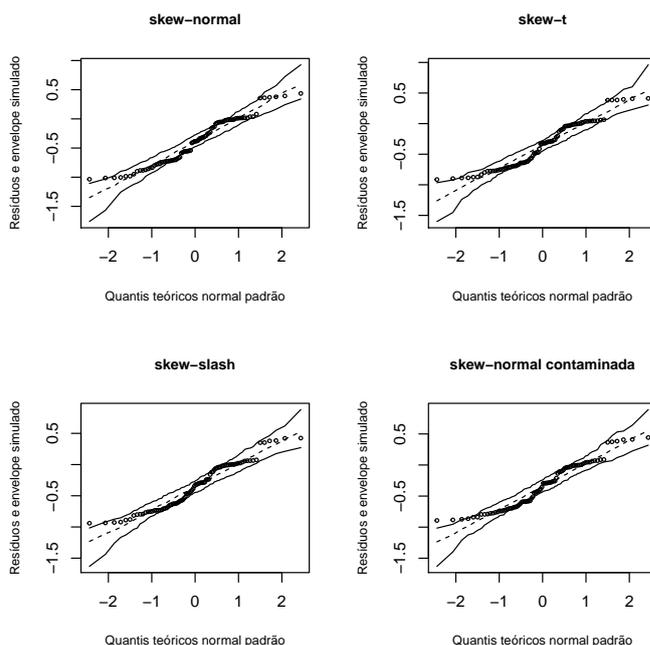


Figura 38 – Gráficos dos resíduos quantílicos aleatorizados e envelopes simulados para os dados de cefamandol.

Os gráficos de resíduos e envelopes simulados da Figura 38 podemos observar que a grande maioria dos resíduos estão dentro das bandas de confiança simuladas em todos os ajustes.

#### • Influência local

Construímos gráficos do vetor  $\mathbf{m}(0)$ , de acordo com os esquemas de perturbação estudados, para analisar a influência devido à contribuição agregada de todos os autovetores nos ajustes com o modelo SMSN-NLME. Os pontos de corte adotados em todos os casos foram obtidos por meio de  $\bar{m}(0) + c^* sm(0)$ , em que  $c^* = 0,5$  e  $c^* = 1$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 39, sob o esquema de perturbação ponderação de casos, verificamos que o indivíduo 6 é o mais influente sob as distribuições SN e ST e o indivíduo 2 é mais influente sob as distribuições SSL e SNC.

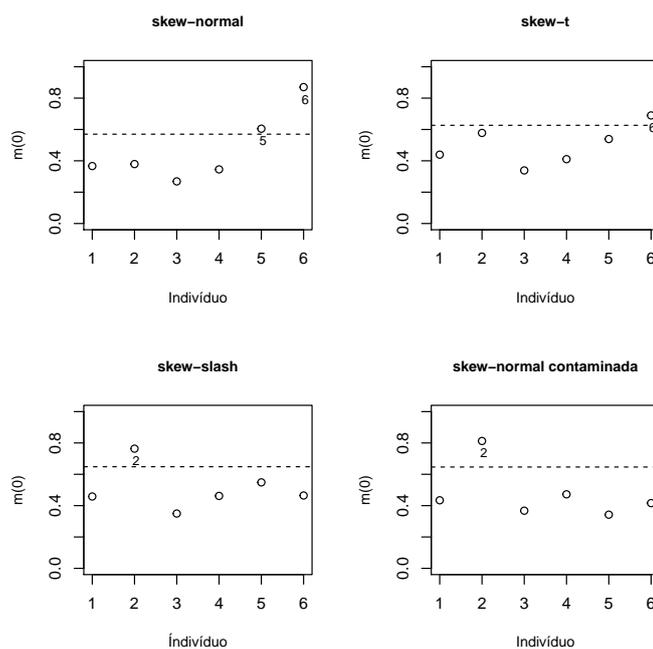


Figura 39 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação ponderação de casos com  $c^* = 1$ .

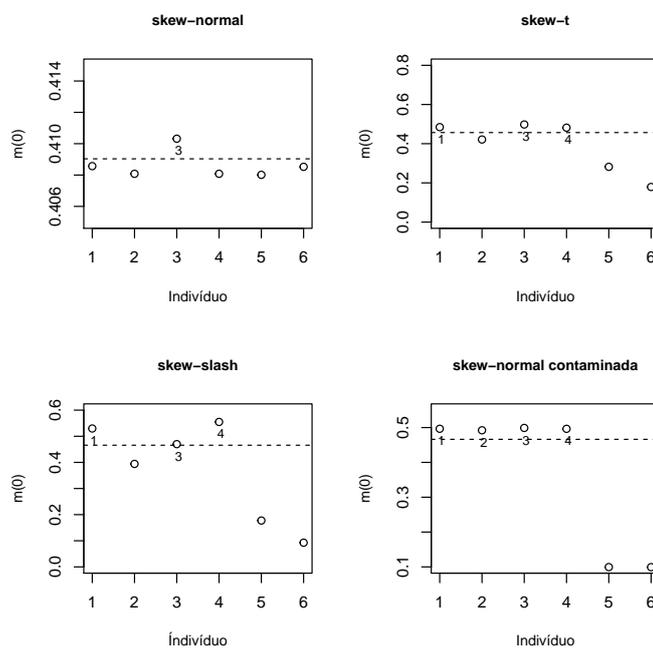


Figura 40 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação na variável resposta com  $c^* = 0,5$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 40, sob o esquema de perturbação na variável resposta, o indivíduo 3 possui medições mais influentes, seguido pelos indivíduos 1 e 4.

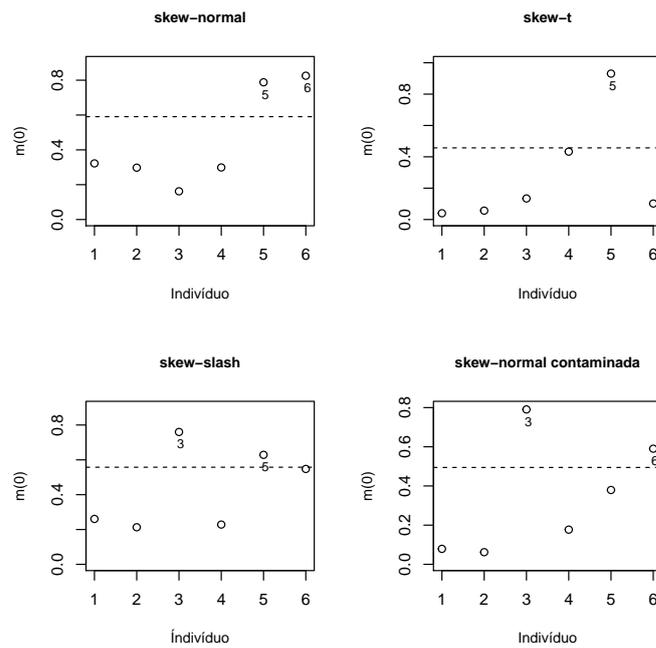


Figura 41 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação na matriz de dispersão dos efeitos aleatórios com  $c^* = 0,5$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 41, sob o esquema de perturbação na matriz de dispersão dos efeitos aleatórios, os indivíduos 3, 5 e 6 são os mais influentes.

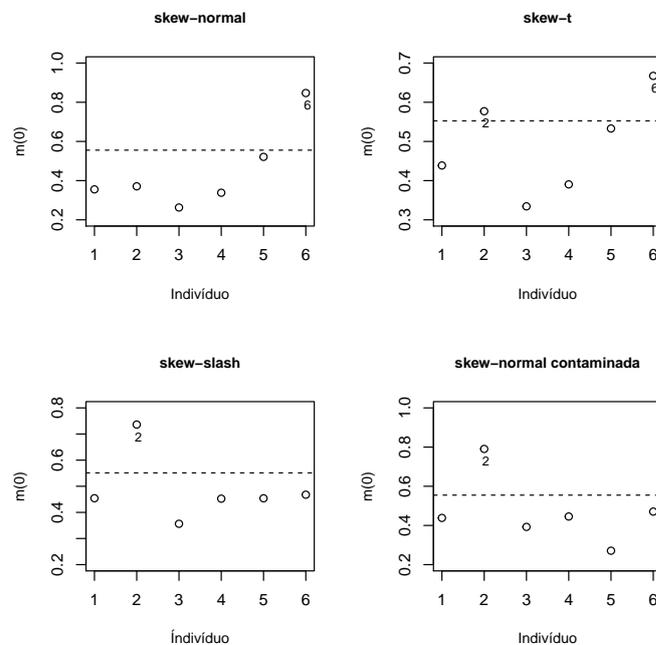


Figura 42 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação na matriz de dispersão dos erros com  $c^* = 0,5$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 42, sob o esquema de perturbação na matriz de dispersão dos erros, o indivíduo 6 é o mais influente sob as distribuições SN e ST e o indivíduo 2

é mais influente sob as distribuições SSL e SNC, assim como sob o esquema de perturbação ponderação de casos.

Avaliamos o grau de sensibilidade das estimativas dos parâmetros, introduzindo perturbações, conforme esquemas apresentados na Seção 5.2.1.1, na direção de um indivíduo de interesse e obtendo  $\hat{\theta}_{\omega_0+a\mathbf{d}_i}$ , com  $a$  arbitrário e, para o caso de perturbação na variável resposta,  $\mathbf{d}_i$  tem valor 1 nas posições relativas ao indivíduo  $i$  e zero nas demais posições. Como medida para análise utilizamos os percentuais de mudanças nas estimativas, obtidos com a expressão  $(\hat{\theta}_{\omega_0+a\mathbf{d}_i} - \hat{\theta})/\hat{\theta}$ .

Nos gráficos da Figura 43, com esquema de perturbação na variável resposta na direção do indivíduo 3 e com  $a$  no intervalo  $[-1; 1]$ , verificamos que os efeitos fixos não apresentaram mudanças significativas. Entre as componentes aleatórias estudadas, as estimativas de  $d_{22}$  sofreram as maiores mudanças, mas que não ultrapassaram 15%, em valor absoluto.

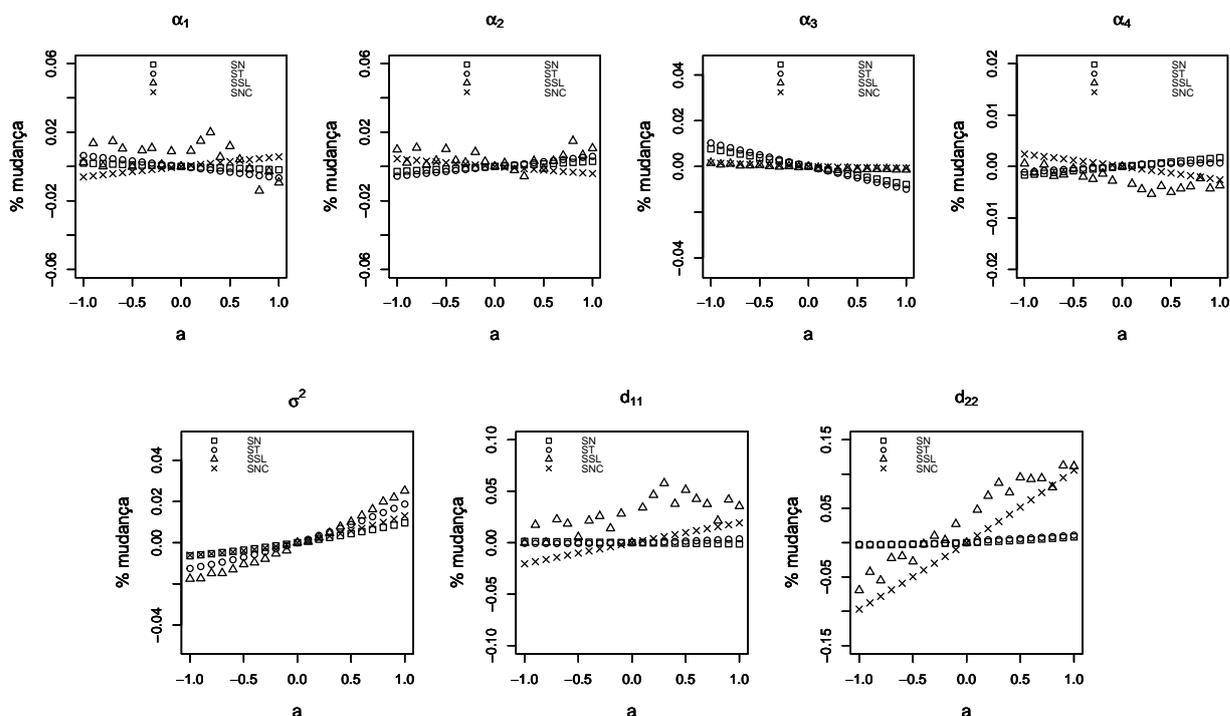


Figura 43 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na variável resposta na direção do indivíduo 3.

Nos gráficos da Figura 44, com perturbação na direção do indivíduo 4 no esquema de perturbação ponderação de casos e com  $a$  no intervalo  $[-0,025; 0,025]$ , os efeitos fixos não apresentaram mudanças significativas. Entre as componentes aleatórias estudadas, a maior mudança ocorreu com o parâmetro  $d_{11}$ , porém as mudanças são inferiores a 15% em valor absoluto.

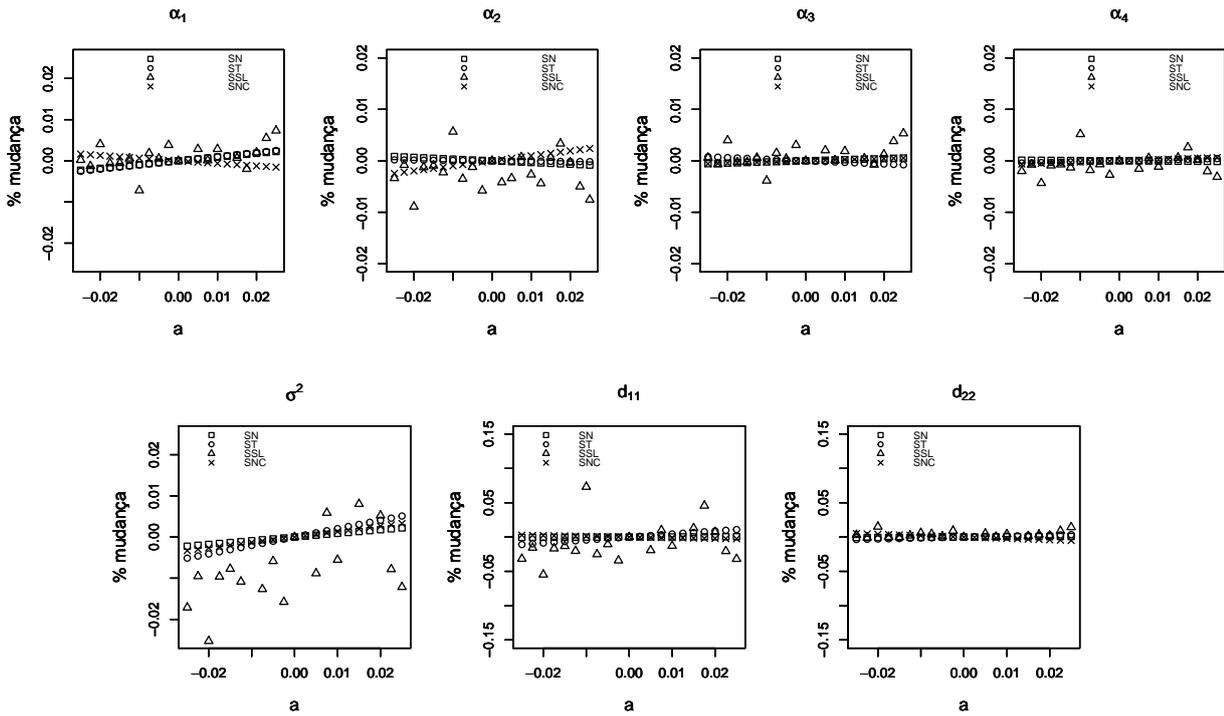


Figura 44 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação ponderação de casos na direção do indivíduo 4.

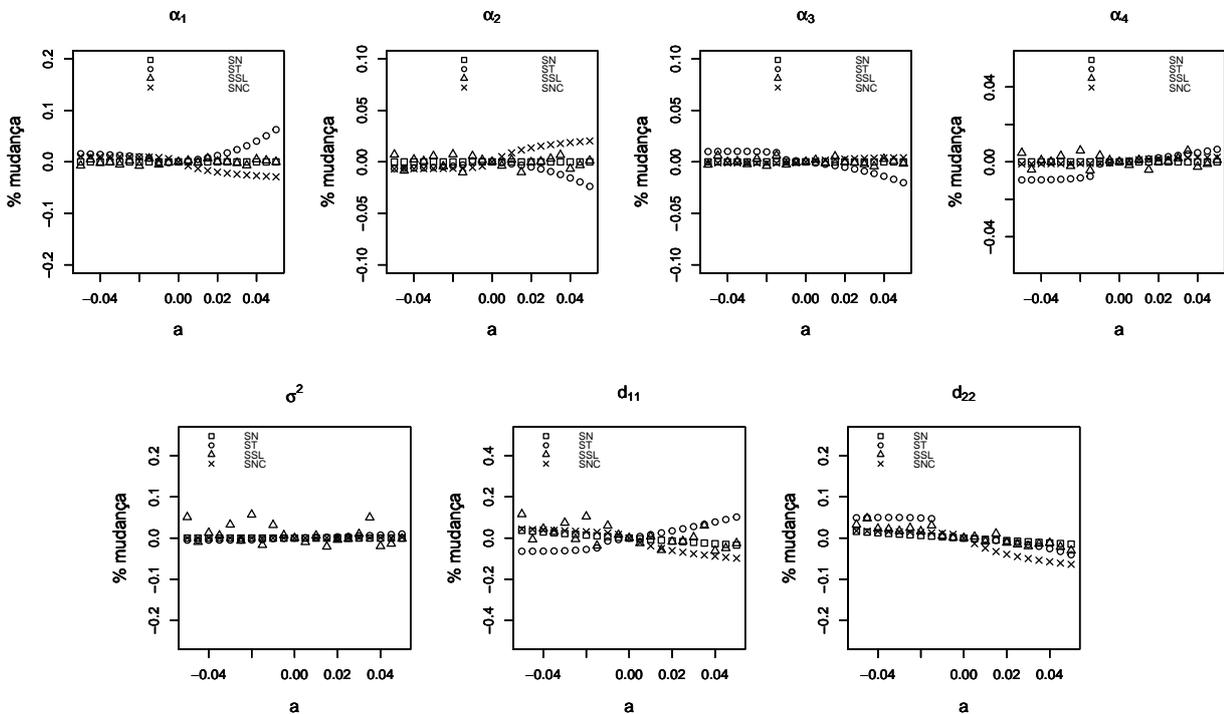


Figura 45 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na matriz de dispersão dos efeitos aleatórios na direção do indivíduo 5.

Nos gráficos da Figura 45, com perturbação na direção do indivíduo 5 no esquema na matriz de dispersão dos efeitos aleatórios e com  $a$  no intervalo  $[-0,05; 0,05]$ , os efeitos fixos não

apresentaram mudanças significativas. Entre as componentes aleatórias estudadas, as mudanças não ultrapassaram 20% em valor absoluto. Observamos nos gráficos dos parâmetros estimados com perturbação que há falta de monotonicidade nas estimações da maioria dos parâmetros nos ajustes sob a distribuição SSL.

### • Exclusão de casos

Para verificar a influência global de observações, realizamos a exclusão de alguns indivíduos. As estimativas dos efeitos fixos e do parâmetro de dispersão  $\sigma^2$  não sofreram mudanças maiores do que 40%, em valor absoluto. Os demais parâmetros sofreram maiores mudanças e os resultados destas mudanças estão apresentados na Tabela 21 seguir.

Tabela 21 – Percentuais de mudanças nas estimativas dos parâmetros quando árvores foram excluídas para os dados de cefamandol.

Distribuição	Indivíduo excluído			
	2	3	4	6
	$d_{11}$			
skew-normal	8	15	15	123
skew-t	-50	10	-42	72
skew-slash	0	50	75	50
skew-normal contaminada	-2	0	31	50
	$d_{12}$			
skew-normal	-28	14	42	200
skew-t	-35	10	-8	23
skew-slash	-50	50	250	100
skew-normal contaminada	-20	-20	58	0
	$d_{22}$			
skew-normal	85	85	142	128
skew-t	32	8	82	-21
skew-slash	-25	-25	101	20
skew-normal contaminada	-14	-43	28	-50
	$\lambda_1$			
skew-normal	205	365	491	-389
skew-t	113	-6	107	-30
skew-slash	-63	10	-189	-145
skew-normal contaminada	96	203	-507	-437
	$\lambda_2$			
skew-normal	321	337	-422	-502
skew-t	-411	-8	-241	91
skew-slash	-96	0	-66	-82
skew-normal contaminada	9	44	2	82

Observamos na Tabela 21 que os indivíduos 4 e 6 são os mais influentes globalmente, e que os parâmetros de assimetria foram os que mais sofreram alterações, principalmente sob a distribuição skew-normal. Na análise de resíduos, verificamos que os valores das distâncias de Mahalanobis e dos resíduos condicionados referentes às medições do indivíduo 6, gráficos das Figuras 35 e 36 são, possivelmente, aberrantes. E na análise de influência local, sob os esquemas de perturbação ponderação de casos, matriz de dispersão dos efeitos aleatórios e matriz

de dispersão dos erros, as medições do indivíduo 6 são consideradas influentes para algumas distribuições.

### 5.2.4 Dados do antiasmático *theophylline*

Os dados do antiasmático *theophylline* estão apresentados na Seção 2.3.

#### • Análise de resíduos

Observamos nos gráficos da Figura 46, que as medições referentes aos indivíduos 1, 2, 5 e 9 apresentam as maiores distâncias de Mahalanobis, sendo que o indivíduo 5 no ajuste sob a distribuição skew-normal, é possivelmente aberrante.

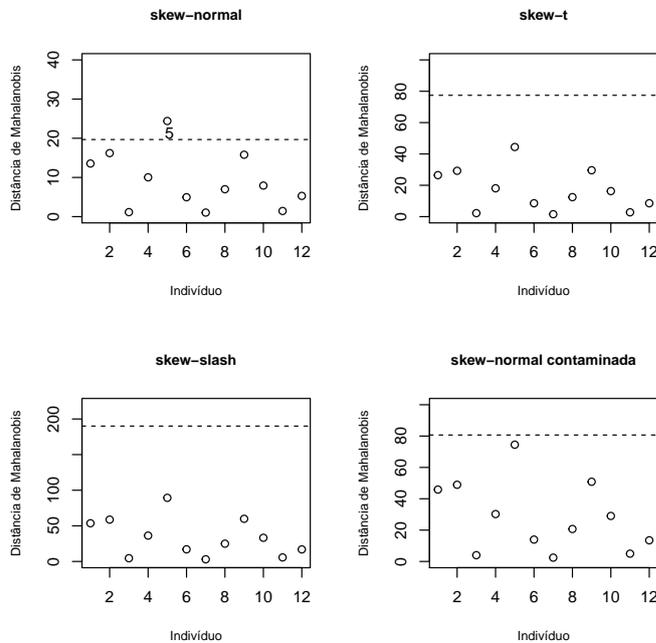


Figura 46 – Distância de Mahalanobis dos dados de *theophylline*.

Os gráficos da Figura 47 mostram resíduos condicionados de cada medição e por indivíduo. Os diâmetros das circunferências são proporcionais aos valores absolutos dos resíduos, assim as medições com resíduos mais destoantes são a segunda medição dos indivíduos 2, 5 e 9, a quarta medição dos indivíduos 5 e 9 e a quinta medição do indivíduo 10.

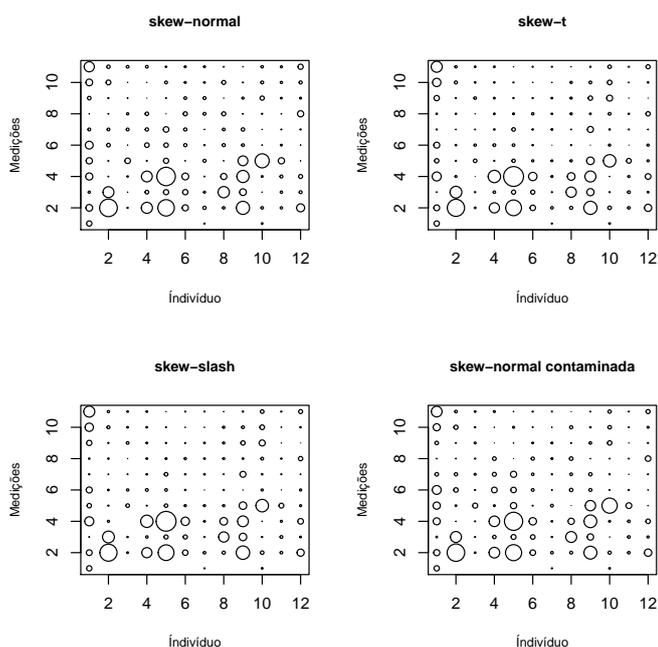


Figura 47 – Gráfico de resíduos condicionados das medições por indivíduos dos dados de *theophylline*.

Nos gráficos dos resíduos da Figura 48, observamos que poucas observações são consideradas atípicas, ou seja, com valores absolutos de resíduos maiores que o quantil  $z_{0,975} = 1,96$  da distribuição normal padrão.

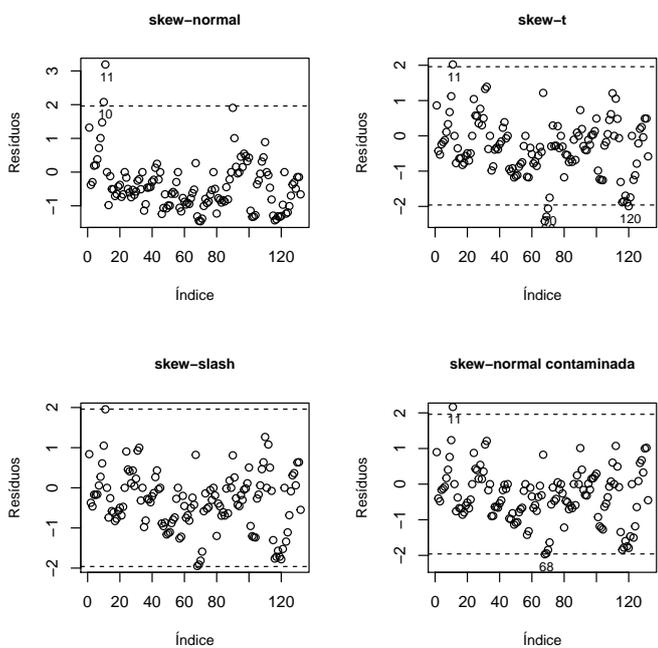


Figura 48 – Gráficos dos resíduos quantílicos aleatorizados para os dados de *theophylline*.

Os gráficos de resíduos e envelopes simulados da Figura 49 a seguir mostram que o comportamento dos resíduos nos ajustes sob as distribuições ST, SSL e SNC são semelhantes e que o ajuste menos adequado ocorre sob a distribuição SN, por apresentar resíduos fora das bandas de confiança simuladas.

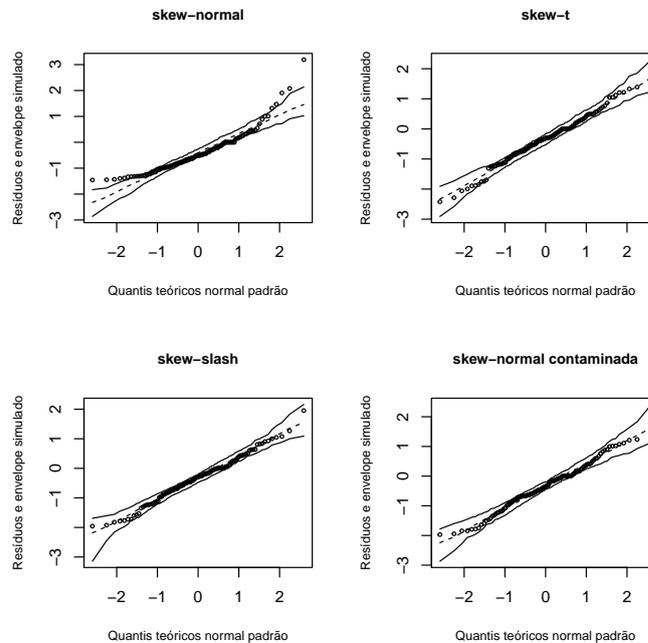


Figura 49 – Gráficos dos resíduos quantílicos aleatorizados e envelopes simulados para os dados de *theophylline*.

Para efeito de comparação entre os ajustes com distribuições assimétricas e simétricas, nesta aplicação construímos o gráfico de envelope simulado, conforme Figura 50, com o ajustes sob distribuições simétricas e verificamos que, na perspectiva de análise de resíduos, os ajustes simétricos possuem mais observações fora das bandas de confiança simuladas, sendo estes ajustes, portanto, menos adequados, principalmente quando comparamos os gráficos sob as distribuições normal e skew-normal contaminada, ajustes menos adequado e mais adequado, respectivamente, em concordância também com os valores dos critérios de seleção de modelos AIC e BIC e dos resultados dos testes de hipóteses apresentados nas Tabelas 12 e 13, respectivamente.

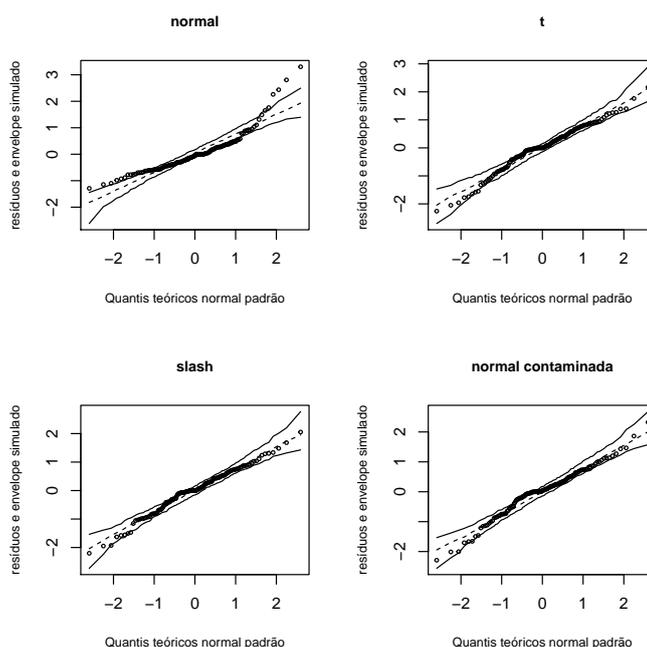


Figura 50 – Gráfico dos resíduos quantílicos aleatorizados e envelopes simulados de ajustes sob distribuições simétricas para os dados de *theophylline*.

### • Influência local

Construímos gráficos do vetor  $\mathbf{m}(0)$ , de acordo com os esquemas de perturbação estudados, para analisar a influência devido à contribuição agregada de todos os autovetores nos ajustes com o modelo SMSN-NLME. Os pontos de corte adotados em todos os casos foram obtidos por meio de  $\bar{m}(0) + c^*sm(0)$ , em que  $c^* = 0,5$  e  $c^* = 0,7$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 51, sob o esquema de perturbação ponderação de casos, verificamos que os indivíduos 3, 7 e 11 possuem as medições mais influentes em todas as distribuições.

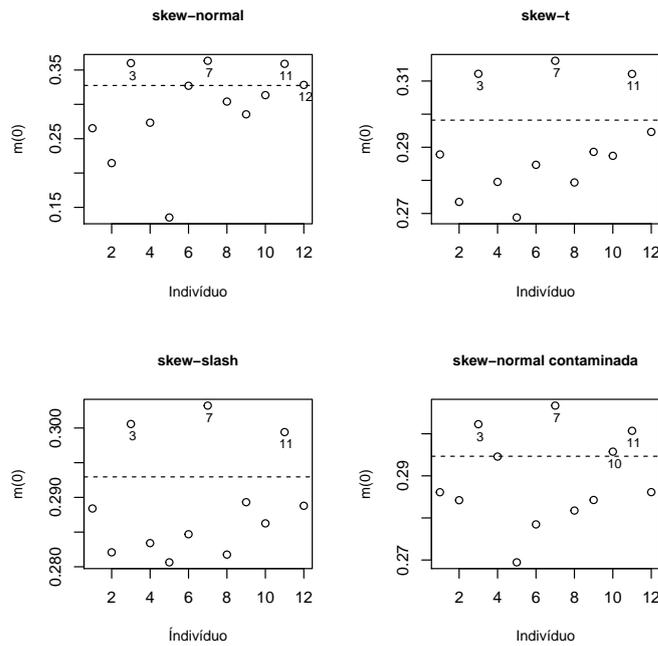


Figura 51 – Gráficos de  $m(0)$  sob esquema de perturbação ponderação de casos com  $c^* = 0,5$ .

Nos gráficos de índices de  $m(0)$  na Figura 52, sob o esquema de perturbação na variável resposta, os indivíduos 3, 7 e 11 possuem as medições mais influentes sob as distribuições ST, SSL e SNC. Na distribuição SNC adicionalmente o indivíduo 12 mostra-se influente também. Sob a distribuição SN, são influentes os indivíduos 5, 10 e 12.

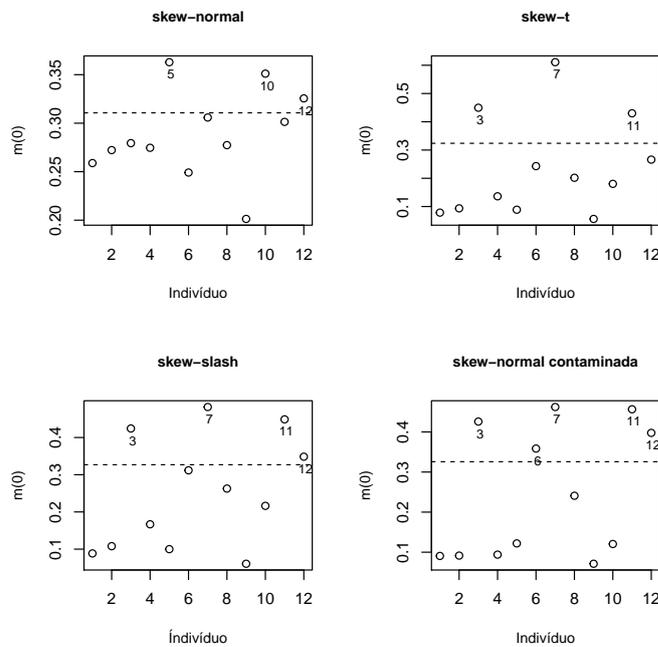


Figura 52 – Gráficos de  $m(0)$  sob esquema de perturbação na variável resposta com  $c^* = 0,5$ .

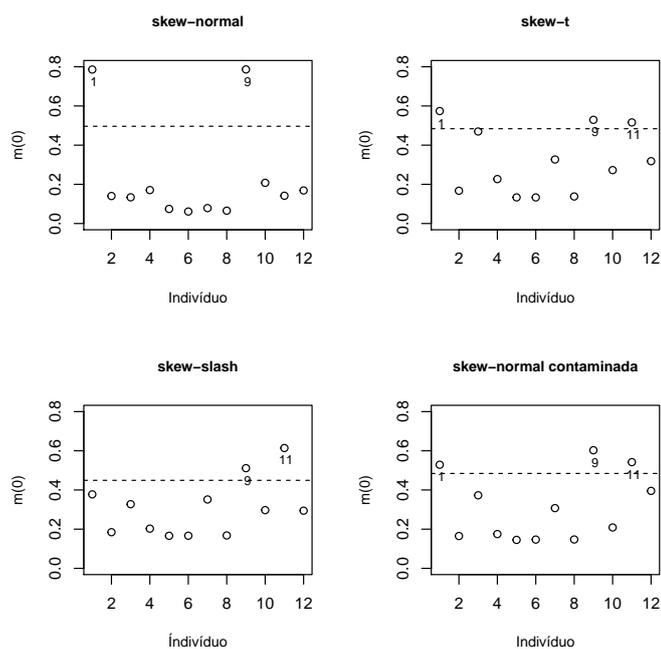


Figura 53 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação na matriz de dispersão dos efeitos aleatórios com  $c^* = 0,7$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 53, sob o esquema de perturbação na matriz de dispersão dos efeitos aleatórios, os indivíduos 1, 9 e 11 possuem as medições mais influentes.

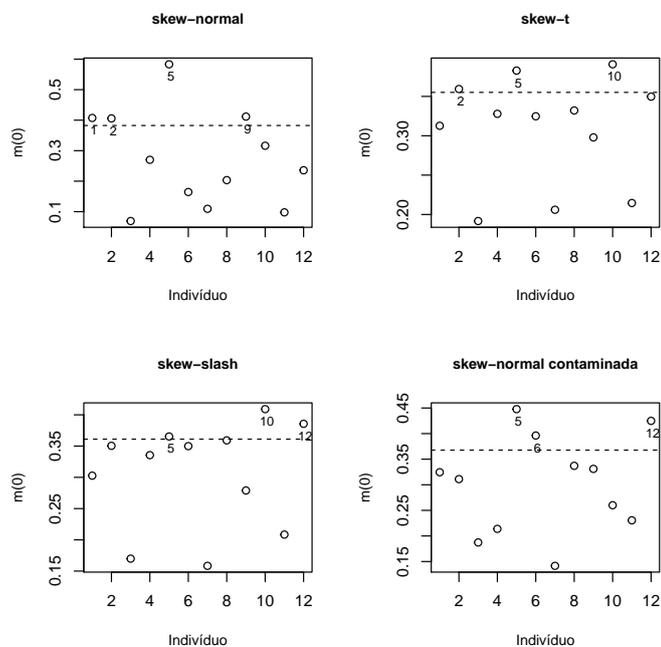


Figura 54 – Gráficos de  $\mathbf{m}(0)$  sob esquema de perturbação na matriz de dispersão dos erros com  $c^* = 0,7$ .

Nos gráficos de índices de  $\mathbf{m}(0)$  na Figura 54, sob o esquema de perturbação na matriz de dispersão dos erros, o indivíduo 5 é o mais influente sob a distribuição SN, 5 e 10 sob a ST, 10 e 12 na SSL e 5 e 12 sob a SNC.

Avaliamos o grau de sensibilidade das estimativas dos parâmetros, introduzindo perturbações, segundo esquemas apresentados na Seção 5.2.1.1, na direção de um indivíduo de interesse e obtendo  $\hat{\theta}_{\omega_0+ad_i}$ , com  $a$  arbitrário e, para o caso de perturbação na variável resposta,  $\mathbf{d}_i$  tem valor 1 nas posições relativas ao indivíduo  $i$  e zero nas demais posições. Como medida para análise utilizamos os percentuais de mudanças nas estimativas, obtidos com a expressão  $(\hat{\theta}_{\omega_0+ad_i} - \hat{\theta})/\hat{\theta}$ .

Como podemos verificar nos gráficos da Figura 55, com perturbação na direção do indivíduo 7 no esquema ponderação de casos e com  $a$  no intervalo  $[-0,5; 0,5]$ , entre os efeitos fixos, o  $lka$  apresentou maiores mudanças, mas inferiores a 30%, em valor absoluto. Entre as componentes aleatórias estudadas, as mudanças não ultrapassaram 20%, em valor absoluto.

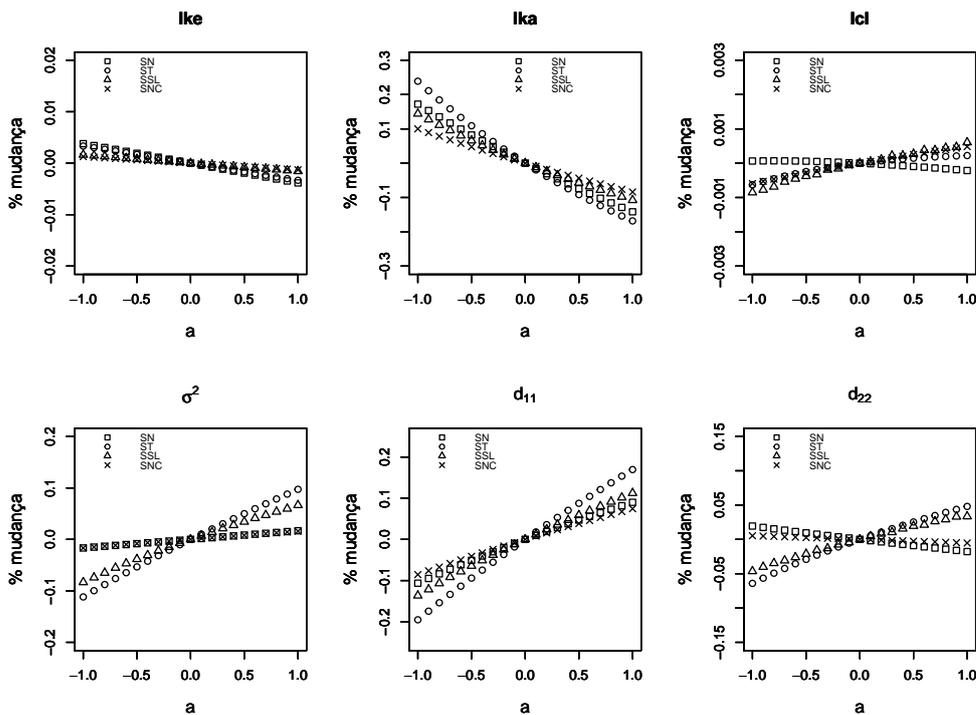


Figura 55 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação ponderação de casos na direção do indivíduo 7.

Nos gráficos da Figura 56, com perturbação na direção do indivíduo 5 no esquema de perturbação na matriz de dispersão dos erros e com  $a$  no intervalo  $[-0,25; 0,25]$ , os efeitos fixos não apresentaram mudanças significativas. Entre as componentes aleatórias estudadas, a maior mudança ocorreu com o parâmetro  $\sigma^2$  que apresentou maiores mudanças próximas de 25%, em valor absoluto.

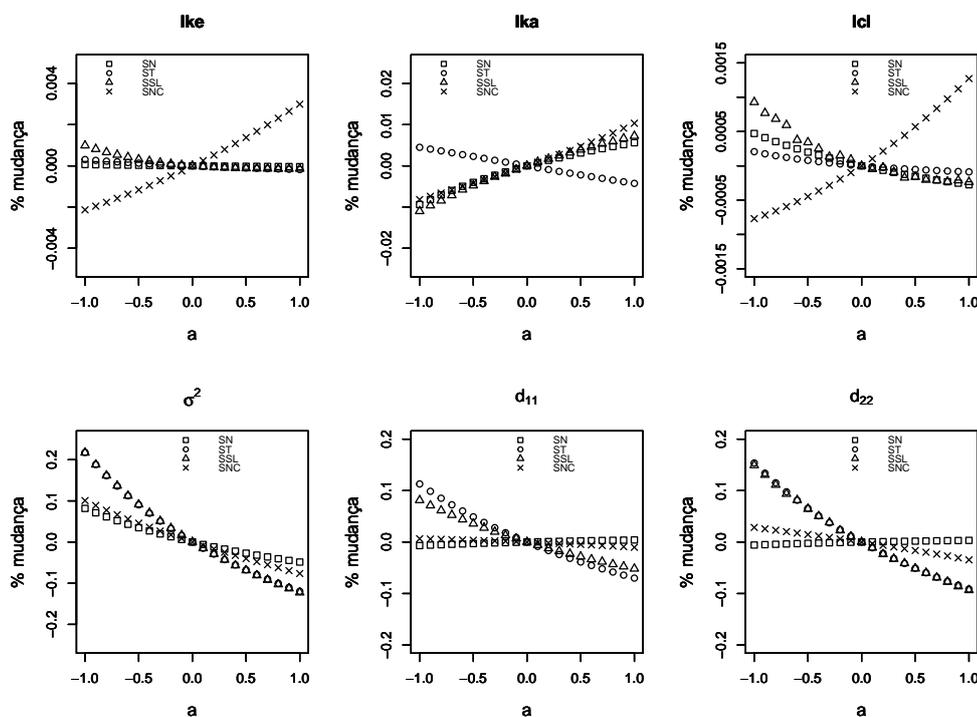


Figura 56 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na matriz de dispersão dos erros na direção do indivíduo 5.

Nos gráficos da Figura 57, com esquema de perturbação na variável resposta na direção do indivíduo 3 e com  $a$  no intervalo  $[-1; 1]$ , os efeitos fixos também não apresentaram mudanças significativas. Entre as componentes aleatórias estudadas, as estimativas de  $\sigma^2$  e  $d_{22}$  sofreram mudanças inferiores a 40%, em valor absoluto.

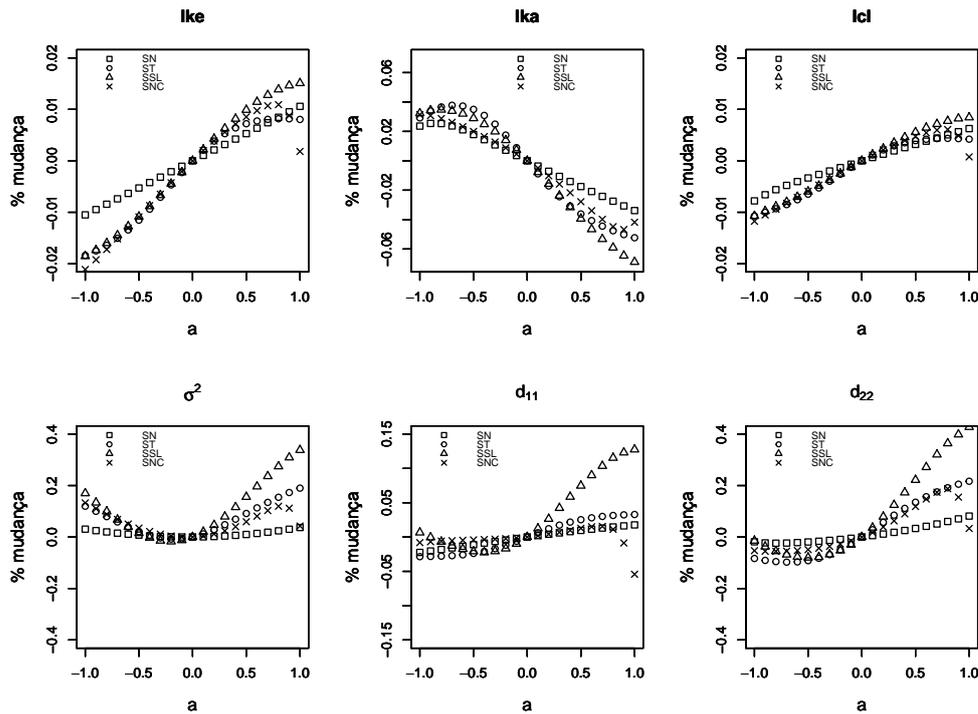


Figura 57 – Gráficos da porcentagem de mudança na estimação dos parâmetros com perturbação na variável resposta na direção do indivíduo 3.

### • Exclusão de casos

Realizamos o procedimento de diagnóstico por eliminação de casos, excluindo alguns indivíduos dos dados de *theophylline* que foram considerados influentes nos estudos anteriores, com o objetivo de verificar a influência global destas observações. As estimativas dos efeitos fixos não sofreram mudanças maiores do que 15%, em valor absoluto, e o parâmetro de dispersão  $\sigma^2$  não sofreu mudança maior do que 35%, em valor absoluto. Os demais parâmetros sofreram maiores mudanças e os resultados destas mudanças estão apresentados na Tabela 22 a seguir.

Tabela 22 – Percentuais de mudanças nas estimativas dos parâmetros quando indivíduos foram excluídos para os dados de *theophylline*.

Distribuição	Indivíduo excluído						
	1	2	3	5	7	9	12
	$d_{11}$						
skew-normal	11	5	3	13	6	-48	5
skew-t	5	2	3	4	-4	-20	5
skew-slash	6	2	5	7	7	-12	0
skew-normal contaminada	12	6	0	-16	-8	-16	-1
	$d_{12}$						
skew-normal	168	16	20	-20	20	204	-64
skew-t	-37	25	262	-92	137	275	-175
skew-slash	0	33	150	16	50	180	-116
skew-normal contaminada	-40	20	200	0	80	280	-220
	$d_{22}$						
skew-normal	-56	6	6	9	-30	313	2
skew-t	-33	0	5	2	-9	-5	-5
skew-slash	-33	-4	14	5	-9	-5	-14
skew-normal contaminada	-28	-4	0	-24	-20	0	-16
	$\lambda_1$						
skew-normal	-58	1	-4	19	61	-102	20
skew-t	37	12	-71	39	134	-96	91
skew-slash	49	-5	-69	22	343	-137	109
skew-normal contaminada	30	-7	-64	66	124	-114	78
	$\lambda_2$						
skew-normal	166	-7	-4	-14	28	100	-1
skew-t	0	-20	23	-30	-7	9	-27
skew-slash	-10	-6	-16	-16	-38	-11	-15
skew-normal contaminada	8	1	4	-101	-1	-2	0

Observamos na Tabela 22 que o indivíduo 9 é o indivíduo mais influente globalmente, pois as estimações dos parâmetros apresentam as maiores alterações, assim como está entre os mais influentes localmente e sob esquemas de perturbação na matriz de dispersão dos efeitos aleatórios e dos erros. Notamos também que o parâmetro mais sensível a retirada de observações é o parâmetro  $d_{12}$ . De forma geral, entre os parâmetros de assimetria,  $\lambda_1$  sofreu maiores alterações, principalmente com a exclusão das medições do indivíduo 7. Na análise de resíduos, podemos verificar que o indivíduo 9 possui uma das maiores distâncias de Mahalanobis, conforme Figura 46.



## MODELO NÃO LINEAR MISTO SKEW-NORMAL

Apresentamos neste Capítulo modelos não lineares com efeitos mistos que, assim como nos modelos apresentados no Capítulo 4, são não lineares nos efeitos fixos mas lineares nos efeitos aleatórios e levam em consideração a presença de observações destoantes e a assimetria em conjuntos de dados. Nestes modelos consideramos uma distribuição assimétrica tanto para os efeitos aleatórios quanto para os erros. Diferentemente dos modelos anteriores, não condicionamos a variável resposta aos efeitos aleatórios.

Desta forma, temos o objetivo de apresentar um modelo NLME em que as componentes aleatórias seguem a distribuição skew-normal, pois a suposição de normalidade dos erros e dos efeitos aleatórios, como discutimos anteriormente, pode não ser realista e, assim, omitir informações importantes presentes nos dados. Para isso implementamos também uma rotina computacional por meio dos algoritmos EM e Newton-Raphson para a estimação dos parâmetros do modelo.

Um modelo não linear nos efeitos fixos e linear nos efeitos aleatórios, com o vetor de resposta observadas  $\mathbf{y}_i$  de dimensão  $m_i \times 1$ ,  $i = 1, \dots, n$ , conforme em Vonesh e Carter (VONESH; CARTER, 1992) e Russo et al. (RUSSO; PAULA; AOKI, 2009) pode ser assim ajustado

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (6.1)$$

em que  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top$  é o vetor de parâmetros desconhecidos,  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)$  é uma função conhecida do vetor de respostas,  $\mathbf{x}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{im_i})^\top$  é um vetor de variáveis explanatórias de dimensão  $m_i \times 1$ ,  $\mathbf{Z}_i$ , de dimensão  $m_i \times q$ ,  $1 \leq q \leq p$ , é uma matriz de constantes conhecidas, os efeitos aleatórios  $\mathbf{b}_i = (b_{i1}, \dots, b_{iq})^\top$  e os erros  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^\top$  são independentes com  $\mathbf{b}_i \stackrel{\text{i.i.d.}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}_b)$  e  $\boldsymbol{\varepsilon}_i \stackrel{\text{ind.}}{\sim} SN_{m_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i}, \boldsymbol{\lambda}_{\varepsilon_i})$ ,  $i = 1, \dots, n$ . As matrizes  $\mathbf{D} = \mathbf{D}(\boldsymbol{\beta})$ , que depende do vetor de parâmetros desconhecidos  $\boldsymbol{\beta}$ , e  $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}_{m_i}$  são matrizes de dispersão e

positivas definidas, com  $\mathbf{D}$  não estruturada.

A matriz de constantes  $\mathbf{Z}_i$ ,  $i = 1, \dots, n$ , conforme discutido em Vonesh e Carter (VONESH; CARTER, 1992) em *modelos mistos pseudo não lineares* (PNMEM), é definida como

$$\mathbf{Z}_i = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_1}, \dots, \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_p} \right] \Big|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}}, \quad i = 1, \dots, n \quad (6.2)$$

com  $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \dots, \tilde{\alpha}_p)^\top$  representando, por exemplo, a estimativa de mínimos quadrados ordinários de  $\boldsymbol{\alpha}$ , e analogamente foram definidos para os casos mais simples considerando a inclusão de um ou mais efeitos aleatórios.

Considerando os momentos da distribuição skew-normal, por meio de (3.21), podemos observar em (6.1) que

$$E[\mathbf{Y}_i] = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + E[\mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i] = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \sqrt{\frac{2}{\pi}} (\mathbf{Z}_i \mathbf{D}^{1/2} \boldsymbol{\delta}_b + \sigma \boldsymbol{\delta}_{\varepsilon_i}), \quad (6.3)$$

em que  $\boldsymbol{\delta}_b = \frac{\boldsymbol{\lambda}_b}{\sqrt{1 + \boldsymbol{\lambda}_b^\top \boldsymbol{\lambda}_b}}$  e  $\boldsymbol{\delta}_{\varepsilon_i} = \frac{\boldsymbol{\lambda}_{\varepsilon_i}}{\sqrt{1 + \boldsymbol{\lambda}_{\varepsilon_i}^\top \boldsymbol{\lambda}_{\varepsilon_i}}}$ .

Da forma que apresentamos o modelo (6.1), em que tanto os efeitos aleatórios quanto os erros têm distribuição skew-normal, a distribuição marginal de  $\mathbf{Y}_i$  está definida na classe de distribuições skew-normal fundamental (FSN) introduzida por Arellano-Valle e Genton (ARELLANO-VALLE; GENTON, 2005). Assim, apresentamos dois casos especiais de interesse prático para o modelo SN-NLME, conforme (LACHOS, 2004). Um caso quando  $\boldsymbol{\lambda}_{\varepsilon_1} = \dots = \boldsymbol{\lambda}_{\varepsilon_m} = \mathbf{0}$  e outro quando  $\boldsymbol{\lambda}_b = \mathbf{0}$ .

- Quando  $\boldsymbol{\lambda}_{\varepsilon_i} = \mathbf{0}$ ,  $i = 1, \dots, n$ .

A distribuição marginal de  $\mathbf{Y}_i$  no modelo (6.1), associado ao vetor de parâmetros de assimetria  $\tilde{\boldsymbol{\lambda}}_{b_i}$ , na qual baseamos a inferência sobre o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \sigma^2, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}_b^\top)^\top$ , segue a distribuição skew-normal,  $\mathbf{Y}_i \sim SN_{m_i}(\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i), \boldsymbol{\Sigma}_i, \tilde{\boldsymbol{\lambda}}_{b_i})$ , com fdp, conforme (3.19), dada por

$$g(\mathbf{y}_i, \boldsymbol{\theta}) = 2\phi_{m_i}\{\mathbf{y}_i | \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i), \boldsymbol{\Sigma}_i\} \Phi\{\tilde{\boldsymbol{\lambda}}_{b_i}^\top \boldsymbol{\Sigma}_i^{-1/2} (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i))\}, \quad (6.4)$$

em que  $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \sigma^2 \mathbf{I}_{m_i}$  e

$$\tilde{\boldsymbol{\lambda}}_{b_i} = \frac{\boldsymbol{\Sigma}_i^{-1/2} \mathbf{Z}_i \mathbf{D} \boldsymbol{\zeta}_b}{\sqrt{1 + \boldsymbol{\zeta}_b^\top \boldsymbol{\Lambda}_i \boldsymbol{\zeta}_b}}, \quad \text{com } \boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \sigma^{-2} \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \quad \text{e} \quad \boldsymbol{\zeta}_b = \mathbf{D}^{-1/2} \boldsymbol{\lambda}_b.$$

Podemos, então, ajustar o modelo skew-normal com as suposições

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (6.5)$$

em que  $\mathbf{b}_i \stackrel{\text{i.i.d.}}{\sim} SN_q(\mathbf{0}, \mathbf{D}, \boldsymbol{\lambda}_b)$  e  $\boldsymbol{\varepsilon}_i \stackrel{\text{ind.}}{\sim} N_{m_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i})$ .

Tendo em vista o modelo (6.5) e o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \sigma^2, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}_b^\top)^\top$ , assumimos o vetor aleatório

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{pmatrix} \sim SN_{q+m_i} \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{m_i} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\lambda}_b \\ \mathbf{0} \end{pmatrix} \right), \quad i = 1, \dots, n. \quad (6.6)$$

• Quando  $\boldsymbol{\lambda}_b = \mathbf{0}$ .

A distribuição marginal de  $\mathbf{Y}_i$  no modelo (6.1), associado ao vetor de parâmetros de assimetria  $\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\varepsilon}_i}$ , na qual baseamos a inferência sobre o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \sigma^2, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}_{\boldsymbol{\varepsilon}_i}^\top)^\top$ , segue a distribuição skew-normal,  $\mathbf{Y}_i \sim SN_{m_i}(\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i), \boldsymbol{\Sigma}_i, \tilde{\boldsymbol{\lambda}}_{\boldsymbol{\varepsilon}_i})$ , com fdp, de acordo com (3.19), dada por

$$g(\mathbf{y}_i, \boldsymbol{\theta}) = 2\phi_{m_i}\{\mathbf{y}_i | \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i), \boldsymbol{\Sigma}_i\} \Phi\{\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\varepsilon}_i}^\top \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i))\}, \quad (6.7)$$

em que  $\boldsymbol{\Sigma}_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i^\top + \sigma^2 \mathbf{I}_{m_i}$  e

$$\tilde{\boldsymbol{\lambda}}_{\boldsymbol{\varepsilon}_i} = \frac{\sigma \boldsymbol{\Sigma}_i^{-1/2} \boldsymbol{\lambda}_{\boldsymbol{\varepsilon}_i}}{\sqrt{1 + \boldsymbol{\zeta}_{\boldsymbol{\varepsilon}_i}^\top \mathbf{Z}_i \boldsymbol{\Lambda}_i \mathbf{Z}_i^\top \boldsymbol{\zeta}_{\boldsymbol{\varepsilon}_i}}}, \quad \text{com } \boldsymbol{\Lambda}_i = (\mathbf{D}^{-1} + \sigma^{-2} \mathbf{Z}_i^\top \mathbf{Z}_i)^{-1} \quad \text{e} \quad \boldsymbol{\zeta}_{\boldsymbol{\varepsilon}_i} = \sigma^{-1} \boldsymbol{\lambda}_{\boldsymbol{\varepsilon}_i}.$$

Assim, teremos um modelo skew-normal com as seguintes suposições

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (6.8)$$

em que  $\mathbf{b}_i \stackrel{\text{i.i.d.}}{\sim} N_q(\mathbf{0}, \mathbf{D})$  e  $\boldsymbol{\varepsilon}_i \stackrel{\text{ind.}}{\sim} SN_{m_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{m_i}, \boldsymbol{\lambda}_{\boldsymbol{\varepsilon}_i})$ .

Tendo em vista o modelo (6.8) e o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \sigma^2, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}_{\boldsymbol{\varepsilon}_i}^\top)^\top$ , assumimos o vetor aleatório

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{pmatrix} \sim SN_{q+m_i} \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{m_i} \end{pmatrix}, \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\lambda}_{\boldsymbol{\varepsilon}_i} \end{pmatrix} \right), \quad i = 1, \dots, n. \quad (6.9)$$

A abordagem considerando  $\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}_i} = \mathbf{0}$  é mais interessante, pois com erros assimétricos, o vetor de parâmetros de assimetria dos erros tem um número de elementos igual ao de observações. Dessa forma, o modelo terá mais parâmetros do que observações, ou seja, teremos um modelo saturado.

Em ambos casos, se considerarmos os vetores de parâmetros de assimetria nulo,  $\boldsymbol{\lambda}_{\mathbf{b}} = \mathbf{0}$  ou  $\boldsymbol{\lambda}_{\boldsymbol{\varepsilon}_i} = \mathbf{0}$ , estaremos trabalhando com um modelo não linear misto em que as componentes aleatórias tem distribuição normal multivariada (N-NLME), em que assumimos o seguinte vetor aleatório

$$\begin{pmatrix} \mathbf{b}_i \\ \boldsymbol{\varepsilon}_i \end{pmatrix} \sim N_{q+m_i} \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{m_i} \end{pmatrix} \right), \quad i = 1, \dots, n. \quad (6.10)$$

## 6.1 Estimação por máxima verossimilhança

Para a estimação dos parâmetros no modelo vamos utilizar uma solução numérica via o algoritmo do tipo EM. Considerando o modelo (6.5) com erros simétricos, vamos obter as estimativas de máxima verossimilhança por meio da distribuição marginal da variável resposta.

De acordo com a representação estocástica (3.3) e segundo (ARELLANO-VALLE; BOLFARINE; LACHOS, 2005), teremos que

$$\mathbf{b}_i \stackrel{d}{=} \mathbf{D}^{1/2} \boldsymbol{\delta}_{\mathbf{b}} |T_{0i}| + \mathbf{D}^{1/2} (\mathbf{I}_q - \boldsymbol{\delta}_{\mathbf{b}} \boldsymbol{\delta}_{\mathbf{b}}^\top)^{1/2} \mathbf{T}_{1i}, \quad i = 1, \dots, n \quad (6.11)$$

onde  $|T_{0i}|$  representa o valor absoluto de  $T_{0i}$ ,  $T_{0i} \sim N(0, 1)$  e  $\mathbf{T}_{1i} \sim N_q(0, \mathbf{I}_q)$  são independentes e  $\boldsymbol{\delta}_{\mathbf{b}} = \frac{\boldsymbol{\lambda}_{\mathbf{b}}}{\sqrt{1 + \boldsymbol{\lambda}_{\mathbf{b}}^\top \boldsymbol{\lambda}_{\mathbf{b}}}}$ .

Substituindo (6.11) em (6.5), e considerando  $t_j = |T_{0j}|$ , teremos

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \mathbf{Z}_i \mathbf{D}^{1/2} \boldsymbol{\delta}_{\mathbf{b}} t_j + \mathbf{Z}_i \mathbf{D}^{1/2} (\mathbf{I}_q - \boldsymbol{\delta}_{\mathbf{b}} \boldsymbol{\delta}_{\mathbf{b}}^\top)^{1/2} \mathbf{T}_{1i} + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n. \quad (6.12)$$

A expressão (6.12) tem uma representação hierárquica, como apresentada na Seção 3.4 (3.41)-(3.43). Na representação hierárquica a seguir não condicionamos a variável resposta ao efeito aleatório

$$\mathbf{Y}_i | t_i \stackrel{\text{ind.}}{\sim} N_{mi}(\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) + \boldsymbol{\Delta}_i t_i, \boldsymbol{\Psi}_i) \quad \text{e} \quad (6.13)$$

$$t_i \stackrel{\text{iid.}}{\sim} HN(0, 1), \quad i = 1, \dots, n, \quad (6.14)$$

onde  $\Delta_i = \mathbf{Z}_i \mathbf{D}^{1/2} \boldsymbol{\delta}_b$ ,  $\Psi_i = \Sigma_i - \Delta_i \Delta_i^\top$ , com  $\mathbf{D}^{1/2}$  tal que  $\mathbf{D}^{1/2} \mathbf{D}^{1/2} = \mathbf{D}$  e  $HN$  denota a distribuico *half*-normal univariada.

Seja  $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$  e  $\mathbf{t} = (t_1, \dots, t_n)^\top$ , em que a varivel  $\mathbf{t}$  no  observada e, como consequncia direta da representao hierrquica (6.13)-(6.14), teremos que a funo de log-verossimilhana completa associada a  $\mathbf{y}_c = (\mathbf{y}^\top, \mathbf{t}^\top)^\top$   dada por

$$\begin{aligned} \ell(\boldsymbol{\theta} | \mathbf{y}_c) &\propto \frac{1}{2} \sum_{i=1}^n \left\{ -\log |\Psi_i| - (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) - \Delta_i t_i)^\top \Psi_i^{-1} (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i) - \Delta_i t_i) - t_i^2 \right\} \\ &= \frac{1}{2} \sum_{i=1}^n \left\{ -\log |\Psi_i| - (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i))^\top \Psi_i^{-1} (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)) \right. \\ &\quad \left. + 2t_i (\mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i))^\top \Psi_i^{-1} \Delta_i - t_i^2 \Delta_i^\top \Psi_i^{-1} \Delta_i - t_i^2 \right\}. \end{aligned} \quad (6.15)$$

De acordo com os resultados apresentados em (3.36) e (3.37) sobre os momentos da distribuico normal truncada, e segundo (ARELLANO-VALLE; BOLFARINE; LACHOS, 2005; LACHOS *et al.*, 2007), teremos as expresses

$$\hat{t}_i = \hat{\eta}_i + \frac{\phi(\frac{\hat{\eta}_i}{\hat{\tau}_i})}{\Phi(\frac{\hat{\eta}_i}{\hat{\tau}_i})} \hat{\tau}_i \quad \text{e} \quad (6.16)$$

$$\hat{t}_i^2 = \hat{\eta}_i^2 + \hat{\tau}_i^2 + \frac{\phi(\frac{\hat{\eta}_i}{\hat{\tau}_i})}{\Phi(\frac{\hat{\eta}_i}{\hat{\tau}_i})} \hat{\tau}_i \hat{\eta}_i, \quad (6.17)$$

em que  $\phi$  e  $\Phi$  so as fdp e fda da distribuico normal padro, e

$$\begin{aligned} \hat{\eta}_i &= \frac{\hat{\Delta}_i^\top \hat{\Psi}_i^{-1} (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}, \mathbf{x}_i))}{1 + \hat{\Delta}_i^\top \hat{\Psi}_i^{-1} \hat{\Delta}_i} \quad \text{e} \\ \hat{\tau}_i^2 &= \frac{1}{1 + \hat{\Delta}_i^\top \hat{\Psi}_i^{-1} \hat{\Delta}_i}. \end{aligned}$$

Considerando  $\hat{\boldsymbol{\theta}}^{(k)} = (\hat{\boldsymbol{\alpha}}^{(k)\top}, \hat{\sigma}^2^{(k)}, \hat{\boldsymbol{\beta}}^{(k)\top}, \hat{\boldsymbol{\lambda}}_b^{(k)\top})^\top$  a estimativa de  $\boldsymbol{\theta}$  na  $k$ -sima iterao do algoritmo do tipo EM, a esperana condicional da funo de log-verossimilhana dos dados completos tem a forma

$$\begin{aligned} Q(\boldsymbol{\theta} | \hat{\boldsymbol{\theta}}^{(k)}) &= E[\ell(\boldsymbol{\theta} | \mathbf{y}_c) | \mathbf{y}, \hat{\boldsymbol{\theta}}^{(k)}] = \\ &\frac{1}{2} \sum_{i=1}^n \left\{ -\log |\hat{\Psi}_i^{(k)}| - (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i))^\top \hat{\Psi}_i^{(k)-1} (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i)) \right. \\ &\quad \left. + 2\hat{t}_i^{(k)} (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i))^\top \hat{\Psi}_i^{(k)-1} \hat{\Delta}_i^{(k)} - \hat{t}_i^{(k)2} \hat{\Delta}_i^{(k)\top} \hat{\Psi}_i^{(k)-1} \hat{\Delta}_i^{(k)} - \hat{t}_i^{(k)2} \right\}. \end{aligned}$$

**Passo E:** Calcular o valores de  $\hat{t}_i^{(k)}$  e  $\hat{t}_i^2^{(k)}$  com as expressões (6.16) e (6.17).

Maximizando a função  $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)})$  em relação a  $\boldsymbol{\theta}$ , obtemos as estimativas de  $\hat{\boldsymbol{\theta}}^{(k+1)}$  cujas expressões são apresentadas no passo M do algoritmo EM

**Passo CM:** Atualizar  $\hat{\boldsymbol{\alpha}}^{(k)}$ ,  $\hat{\boldsymbol{\Psi}}_i^{(k)}$ ,  $\hat{\boldsymbol{\Delta}}_i^{(k)}$  e  $\hat{\boldsymbol{\Sigma}}_i^{(k)}$  com as expressões

$$\hat{\boldsymbol{\alpha}}^{(k+1)} = \operatorname{argmin}_{\boldsymbol{\alpha}} \left( \sum_{i=1}^n \left\{ \frac{1}{2} \hat{\mathbf{r}}_i^{(k)\top} \hat{\mathbf{r}}_i^{(k)} - \hat{t}_i^{(k)} \hat{\mathbf{r}}_i^{(k)\top} \boldsymbol{\Delta}_i^{(k)} \right\} \right), \quad (6.18)$$

$$\hat{\boldsymbol{\Psi}}_i^{(k+1)} = \hat{\mathbf{r}}_i^{(k)} \hat{\mathbf{r}}_i^{(k)\top} - \hat{t}_i^{(k)} \hat{\mathbf{r}}_i^{(k)} \hat{\boldsymbol{\Delta}}_i^{(k)\top} - \hat{t}_i^{(k)} \hat{\boldsymbol{\Delta}}_i^{(k)} \hat{\mathbf{r}}_i^{(k)\top} + \hat{t}_i^2^{(k)} \hat{\boldsymbol{\Delta}}_i^{(k)} \hat{\boldsymbol{\Delta}}_i^{(k)\top}, \quad (6.19)$$

$$\hat{\boldsymbol{\Delta}}_i^{(k+1)} = \frac{\hat{t}_i^{(k)} (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i))}{\hat{t}_i^2^{(k)}} \mathbf{e} \quad (6.20)$$

$$\hat{\boldsymbol{\Sigma}}_i^{(k+1)} = \hat{\boldsymbol{\Psi}}_i^{(k+1)} + \hat{\boldsymbol{\Delta}}_i^{(k+1)} \hat{\boldsymbol{\Delta}}_i^{(k+1)\top}, \quad (6.21)$$

em que  $\hat{\mathbf{r}}_i^{(k)} = \mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(k)}, \mathbf{x}_i)$ .

Para estimar os elementos do vetor de parâmetros  $\boldsymbol{\alpha}$  da função não linear  $\mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)$ , usamos o algoritmo de Newton-Raphson em  $v$  iterações, que é um algoritmo iterativo de otimização, no qual utilizamos a função escore e a matriz de informação de Fisher observada, cujas expressões são dadas, respectivamente, por

$$\hat{\mathbf{U}}_{\boldsymbol{\alpha}}^{(v+1)} = \frac{1}{\hat{\sigma}^2^{(k)}} \sum_{i=1}^n \left\{ \hat{\mathbf{J}}_i^{(v)\top} \left( \mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x}_i) - \hat{t}_i \hat{\boldsymbol{\Delta}}_i^{(k)} \right) \right\} \mathbf{e} \quad (6.22)$$

$$\hat{\mathbf{K}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(v+1)} = \frac{1}{\hat{\sigma}^2^{(k)}} \sum_{i=1}^n \left\{ \hat{\mathbf{J}}_i^{(v)\top} \hat{\mathbf{J}}_i^{(v)} + \operatorname{vec}^{-1} \left[ \hat{\mathbf{L}}_i^{(v)\top} \left( \hat{t}_i \hat{\boldsymbol{\Delta}}_i^{(k)} - (\mathbf{y}_i - \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x}_i)) \right) \right] \right\}, \quad (6.23)$$

em que as matrizes  $\hat{\mathbf{J}}_i^{(v)}$  e  $\hat{\mathbf{L}}_i^{(v)}$  possuem elementos da forma  $\frac{\partial \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x}_i)}{\partial \hat{\alpha}_r^{(v)}}$ ,  $r = 1, \dots, p$  e  $\frac{\partial^2 \mathbf{f}(\hat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x}_i)}{\partial \hat{\alpha}_r^{(v)} \partial \hat{\alpha}_s^{(v)}}$ ,  $r, s = 1, \dots, p$ , respectivamente.

A expressão do estimador do vetor de parâmetros  $\boldsymbol{\alpha}$  na  $(v+1)$ -ésima iteração do algoritmo Newton-Raphson é dado por

$$\hat{\boldsymbol{\alpha}}^{(v+1)} = \hat{\boldsymbol{\alpha}}^{(v)} + \left[ \hat{\mathbf{K}}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(v+1)} \right]^{-1} \hat{\mathbf{U}}_{\boldsymbol{\alpha}}^{(v+1)}. \quad (6.24)$$

Após a convergência do algoritmo Newton-Raphson, as estimativas dos elementos do vetor de parâmetros  $\boldsymbol{\alpha}^{(v+1)}$  é computada no algoritmo EM na  $k$ -ésima iteração.

Considerando  $\boldsymbol{\Psi}_i = \sigma^2 \mathbf{I}_{m_i}$ , teremos que

$$\hat{\sigma}^2^{(k+1)} = \frac{1}{N} \sum_{i=1}^n \left\{ \hat{\mathbf{r}}_i^{(k)\top} \hat{\mathbf{r}}_i^{(k)} - 2 \hat{t}_i^{(k)} \hat{\mathbf{r}}_i^{(k)\top} \hat{\boldsymbol{\Delta}}_i^{(k)} + \hat{t}_i^2^{(k)} \hat{\boldsymbol{\Delta}}_i^{(k)\top} \hat{\boldsymbol{\Delta}}_i^{(k)} \right\}, \quad (6.25)$$

em que  $N = \sum_{i=1}^n m_i$ .

Por meio da inversa generalizada da matriz de constantes  $\mathbf{Z}_i, \mathbf{Z}_i^-$ , obtemos um estimador para a matriz de dispersão e para o vetor de parâmetros de assimetria dos efeitos aleatórios, de acordo com (6.11) e (6.4), conforme expressões a seguir

$$\widehat{\mathbf{D}}^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left\{ \mathbf{Z}_i^- \left[ \widehat{\boldsymbol{\Sigma}}_i^{(k+1)} - \widehat{\sigma}^{2(k+1)} \mathbf{I}_{m_i} \right] \mathbf{Z}_i^{-\top} \right\} \quad \text{e} \quad (6.26)$$

$$\widehat{\boldsymbol{\lambda}}_b^{(k+1)} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\widehat{\mathbf{D}}^{(k+1)-1/2} \mathbf{Z}_i^- \widehat{\boldsymbol{\Delta}}_i^{(k+1)}}{(1 - \widehat{\boldsymbol{\Delta}}^{(k+1)\top} \mathbf{Z}_i^{-\top} \widehat{\mathbf{D}}^{(k+1)-1} \mathbf{Z}_i^- \widehat{\boldsymbol{\Delta}}_i^{(k+1)})^{1/2}} \right\}. \quad (6.27)$$

## 6.2 Aplicações

Apresentamos a seguir os resultados de ajustes realizados com o modelo proposto em dois conjuntos de dados apresentados no Capítulo 2.

### 6.2.1 Dados de crescimento de laranjeiras

Os dados de crescimento de laranjeiras, descritos na Seção 2.1, foram ajustados com o modelo (6.5) com erros simétricos, considerando a medida da circunferência do tronco  $y$ , em  $mm$ , em função do tempo  $t$ , em dias, por meio da relação não linear

$$f(\boldsymbol{\alpha}, t_{ij}) = \frac{\text{Asym}}{1 + \exp[-(t_{ij} - \text{xmid})/\text{scal}]}, \quad i = 1, \dots, 5, \quad j = 1, \dots, 7, \quad (6.28)$$

com  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top = (\text{Asym}, \text{xmid}, \text{scal})^\top$ .

A matriz de contantes  $\mathbf{Z}_i$  é, conforme (6.2), da forma

$$\mathbf{Z}_i = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \text{scal}} \right] \Big|_{\boldsymbol{\alpha}=\tilde{\boldsymbol{\alpha}}}, \quad i = 1, \dots, 6$$

com  $\tilde{\boldsymbol{\alpha}}$  representando, por exemplo, a estimativa de mínimos quadrados ordinários de  $\boldsymbol{\alpha}$ . Dessa forma, o efeito aleatório está associado ao parâmetro  $\text{scal}$ .

Ajustamos o modelo a seguir

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 12. \quad (6.29)$$

em que a variável resposta  $y_i$  tem distribuição skew-normal multivariada.

Na Tabela 23 a seguir apresentamos os valores ajustados dos parâmetros para o modelo proposto, com  $\boldsymbol{\theta} = (\text{Asym}, \text{xmid}, \text{scal}, \sigma^2, d, \lambda_b)^\top$ .

Tabela 23 – Estimativas dos parâmetros como erros padrões (E.P.) do modelo SN-NLME ajustados aos dados de crescimento de laranjeiras.

$\boldsymbol{\theta}$	Estimativa	(E.P.)
Asym	192,571	(0,006)
xmid	727,623	(0,003)
scal	353,233	(0,005)
$\sigma^2$	0,001	(0,001)
$d$	921,944	(4,800)
$\lambda_b$	0,985	(0,001)

Os valores dos respectivos erros padrões das estimativas foram obtidos por meio da matriz de informação de Fisher observada, cujos elementos estão apresentados no Apêndice D.

Observamos que os erros padrões são bem pequenos, com exceção do erro padrão do parâmetro de dispersão do efeito aleatório.

### 6.2.2 Dados do antibiótico cefamandol

Os dados de cefamandol, descritos na Seção 2.2, foram ajustados com o modelo (6.5), em que a assimetria está associada ao efeito aleatório, considerando sua concentração no sangue  $y$ , em  $mcg/ml$ , em função do tempo  $t$ , em minutos, por meio da relação não linear

$$f(\boldsymbol{\alpha}, t_{ij}) = \alpha_1 e^{-t_{ij} \alpha_2'} + \alpha_3 e^{-t_{ij} \alpha_4'}, \quad i = 1, \dots, 6, \quad j = 1, \dots, 14,$$

com  $\alpha_2' = \log \alpha_2$ ,  $\alpha_4' = \log \alpha_4$ ,  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top$  e  $\alpha_2 > 0$ ,  $\alpha_4 > 0$ .

A matriz de contantes  $\mathbf{Z}_i$  é, conforme (6.2), da forma

$$\mathbf{Z}_i = \left[ \frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_4} \right] \Bigg|_{\boldsymbol{\alpha} = \tilde{\boldsymbol{\alpha}}}, \quad i = 1, \dots, 6$$

com  $\tilde{\boldsymbol{\alpha}}$  representando, por exemplo, a estimativa de mínimos quadrados ordinários de  $\boldsymbol{\alpha}$  e, assim, o efeito aleatório está associado ao parâmetro  $\alpha_4$ .

Assim teremos o modelo

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 12. \quad (6.30)$$

em que a variável resposta  $y_i$  tem distribuição skew-normal multivariada.

Na Tabela 24 a seguir apresentamos os valores ajustados dos parâmetros para o modelo proposto, com  $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \sigma^2, d, \lambda_b)^\top$ .

Tabela 24 – Estimativas dos parâmetros como erros padrões (E.P.) do modelo SN-NLME ajustados aos dados de cefamandol.

$\theta$	Estimativa	(E.P.)
$\alpha_1$	299,690	(0,001)
$\alpha_2$	-2,077	(0,003)
$\alpha_3$	92,359	(0,001)
$\alpha_4$	-3,962	(0,002)
$\sigma^2$	0,001	(0,001)
$d$	17,893	(0,005)
$\lambda_b$	1,471	(0,002)

Os valores dos respectivos erros padrões das estimativas foram obtidos por meio da matriz de informação de Fisher observada, cujos elementos estão apresentados no Apêndice D. Observamos que os erros padrões são bem pequenos para todas as estimativas.

### 6.2.3 Dados do antiasmático theophylline

Ajustamos o modelo (6.5), em que os erros tem distribuição normal multivariada, com os dados de *theophylline* apresentados na Seção 2.3, considerando o vetor de parâmetros  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top = (lke, lka, lcl)^\top$ , sua concentração no sangue  $y$ , em  $mg/l$ , em função do tempo  $t$ , em horas, e da dose administrada em cada paciente  $Ds$ ,  $mg/kg$ , por meio da relação não linear

$$f(\boldsymbol{\alpha}, t_{ij}) = \frac{Ds_{ij} e^{lke+lka-lcl}}{e^{lka} - e^{lke}} (e^{-e^{lke}t_{ij}} - e^{-e^{lka}t_{ij}}), \quad i = 1, \dots, 12, \quad j = 1, \dots, 11.$$

Para esta aplicação consideramos os elementos da matriz  $\mathbf{Z}_i$ , conforme (6.2), é da forma  $\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) / \partial lcl$ , avaliada na estimativa de mínimos quadrados ordinários de  $\alpha_3$  e o efeito aleatório está associado a este parâmetro.

Assim teremos o modelo

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, 12. \quad (6.31)$$

com a resposta  $y_i$  tendo distribuição skew-normal multivariada.

Na Tabela 25 a seguir apresentamos os valores ajustados parâmetros para o modelo proposto, com  $\boldsymbol{\theta} = (lke, lka, lcl, \sigma^2, d, \lambda_b)^\top$ .

Tabela 25 – Estimativas dos parâmetros como erros padrões (E.P.) do modelo SN-NLME ajustados aos dados de *theophylline*.

$\boldsymbol{\theta}$	Estimativa	(E.P.)
<i>lke</i>	-2,524	(0,005)
<i>lka</i>	0,399	(0,006)
<i>lcl</i>	-3,248	(0,003)
$\sigma^2$	0,005	(0,001)
<i>d</i>	3,164	(0,010)
$\lambda_b$	0,147	(0,001)

Os valores dos respectivos erros padrões das estimativas foram obtidos por meio da matriz de informação de Fisher observada, cujos elementos estão apresentados no Apêndice D. Observamos que os erros padrões são bem pequenos para todas as estimativas.

### 6.3 Simulações

Realizamos simulações de Monte Carlo para examinar a performance do modelo proposto em dois cenários. Neste estudo de simulação, temos também o interesse em investigar propriedades assintóticas dos estimadores de  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)^\top$ ,  $\sigma^2$  e  $d$ , com  $\lambda_b = 0, 1$  e  $0, 5$ . O critério de convergência adotado foi  $\max|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}| \leq 10^{-3}$ ,  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \sigma^2, d, \lambda_b)^\top$ . Geramos dados artificiais para a variável resposta  $y$ , sob as distribuições normal multivariada e skew-normal univariada para as componentes aleatórias com a curva

$$y_{ij} = f(\boldsymbol{\alpha}, t_{ij}) + \mathbf{Z}_{ij}b_i + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, \dots, 12, \quad j = 1, \dots, 11, \quad (6.32)$$

em que  $b_i \sim SN(0; d; \lambda_b)$ ,  $\boldsymbol{\varepsilon}_i \sim N_{11}(\mathbf{0}; \sigma^2 \mathbf{I}_{11})$ , portanto a abordagem com erros simétricos, e  $\mathbf{Z}_i$  é da forma  $\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{t}_i) / \partial \alpha_3$ , conforme (6.2), avaliada na estimativa de mínimos quadrados ordinários de  $\boldsymbol{\alpha}$ , e

$$f(\boldsymbol{\alpha}, t_{ij}) = \frac{Ds_{ij} e^{\alpha_1 + \alpha_2 - \alpha_3}}{e^{\alpha_2} - e^{\alpha_1}} (e^{-e^{\alpha_1} t_{ij}} - e^{-e^{\alpha_2} t_{ij}}), \quad i = 1, \dots, 12, \quad j = 1, \dots, 11.$$

onde os valores de  $Ds_{ij}$  e de  $t_{ij}$  permaneceram constantes para  $\lambda_b = 0, 1$  e  $\lambda_b = 0, 5$ .

Este cenário de simulação foi baseado no conjunto de dados do antiasmático *theophylline*, apresentado no Seção 2.3.

Nós geramos 1000 amostras de Monte Carlo em cada cenário e ajustamos o modelo proposto. Os valores verdadeiros considerados para os parâmetros são  $\boldsymbol{\alpha} = (2, 50; 0.45; -3, 20)^\top$ ,  $\sigma^2 = 0,05$  e  $d = 0,05$ . As medidas empíricas analisadas foram a média, o viés e o erro quadrático médio (EQM) por meio das expressões

$$\begin{aligned} \text{média} &= \sum_{i=1}^{1000} \frac{\hat{\theta}_i}{1000}, \\ \text{viés} &= \sum_{i=1}^{1000} \frac{(\hat{\theta}_i - \theta)}{1000} \quad \text{e} \\ \text{EQM} &= \sum_{i=1}^{1000} \frac{(\hat{\theta}_i - \theta)^2}{1000}, \end{aligned}$$

com  $\theta = (\boldsymbol{\alpha}^\top, \sigma^2, d, \lambda_b)^\top$ .

Os resultados das simulações estão apresentados nas Tabela 26 a seguir.

Tabela 26 – Média, viés e erro quadrático médio dos ajustes sob o modelo skew-normal baseadas em 1000 amostras Monte Carlo com  $\lambda_b = 0,1$  e  $\lambda_b = 0,5$ .

$\theta$	média	viés	EQM	média	viés	EQM
	$\lambda_b = 0,1$			$\lambda_b = 0,5$		
$\alpha_1$	-2,498	0,002	0,001	-2,478	0,022	0,001
$\alpha_2$	0,455	0,005	0,051	0,481	0,031	0,002
$\alpha_3$	-3,105	0,098	0,014	-2,667	0,533	0,315
$\sigma^2$	0,003	-0,047	0,002	0,006	-0,044	0,003
$d$	0,279	0,229	0,064	1,886	1,836	0,745
$\lambda_b$	-0,015	0,115	0,187	-0,111	-0,611	0,369

Cada amostra de Monte Carlo foi gerada com  $n = 12$  e  $m_i = 11$ , e obtivemos as estimativas de  $\boldsymbol{\alpha}$ ,  $\sigma^2$  e  $d$ .

Podemos verificar em cada cenário que o viés e o erro quadrático médio de cada parâmetro estimado são relativamente pequenos, indicando para uma convergência para os valores verdadeiros dos parâmetros.



---

## CONSIDERAÇÕES FINAIS

---

### 7.1 Conclusões

O modelo SMSN-NLME é um caso geral de modelos não lineares mistos que incluem distribuições simétricas e assimétricas com caudas leves e pesadas, e que ajusta dados balanceados e não balanceados, podendo ser utilizado quando existirem nos dados observações atípicas ou assimetria, devido a flexibilidade da família SMSN de distribuições que possui uma representação hierárquica que permite a obtenção de estimadores de máxima verossimilhança via algoritmo EM e suas extensões. Assim, o modelo que propomos possui aplicabilidade em diversas áreas do conhecimento.

A implementação do algoritmo do tipo EM nos processos de estimação dos parâmetros nos modelos propostos foi de grande importância, pois facilitou os cálculos e viabilizou o desenvolvimento deste trabalho, mesmo com a existência de dificuldades na obtenção de estimadores explícitos para os parâmetros. Para estimação do vetor de parâmetros  $\alpha$  da função não linear optamos por usar o algoritmo de Newton-Raphson, concomitantemente com o EM. Inicialmente, usamos o algoritmo score de Fisher para estimação de  $\alpha$ , porém as estimativas apresentavam viés devido a distribuição assimétrica não ter como esperança o vetor de locação.

Nas aplicações com ajustes do modelo SMSN-NLME, utilizamos a estatística da razão de verossimilhanças para realização de testes de hipóteses e para a construção de intervalos de confiança aproximados e, baseado nesta inferência, o modelo simétrico foi o mais adequado nos ajustes do conjunto de dados do antibiótico cefamandol. Notamos que as estimações do vetor de parâmetros  $\alpha$  da função não linear interferem na assimetria dos valores preditos da variável resposta e resíduos, de forma a se adequar aos dados de acordo com a distribuição e as estimações dos parâmetros e, principalmente se o conjunto de dados tem um número de observações relativamente pequeno. Na aplicação com o conjunto de dados do antiasmático *theophylline*, os melhores ajustes ocorreram com distribuições de causas pesadas, levando-se

em conta a assimetria, o que pode ser verificado nos valores dos critérios de seleção de modelos AIC e BIC, resultados dos testes de hipóteses e intervalos de confiança apresentados nas Tabelas 12, 13 e 14, como também nos gráficos de envelopes simulados apresentados nas Figuras 49 e 50 na análise de diagnóstico.

Os resultados obtidos no estudo de simulação indicam que as estimações dos parâmetros com o modelo SMSN-NLME possuem boas propriedades assintóticas, pois podemos notar que o viés e o erro quadrático médio diminuem quando o tamanho da amostra aumenta na maioria dos cenários estudados. O intervalo de confiança usual de Wald para os efeitos fixos também mostrou-se eficiente, com probabilidade de cobertura próximo do valor nominal, principalmente com tamanhos de amostras maiores. Verificamos que a estimação intervalar feita com intervalos de confiança do tipo Wald não é apropriada para os parâmetros de dispersão e de assimetria e, para estes parâmetros, pode-se usar, por exemplo, intervalos de confiança aproximados por meio da estatística da razão de verossimilhanças.

No primeiro cenário das simulações, em que usamos uma função não linear que é uma combinação linear de termos exponenciais, obtivemos resultados melhores em relação ao viés e o erro quadrático médio do parâmetro de assimetria, quando comparamos com o segundo cenário. Em ambos cenários usamos sistemas de compartimentos de primeira ordem.

Na análise de diagnóstico aplicadas aos dados apresentados no Capítulo 2, alguns resultados podem ser considerados semelhantes entre algumas distribuições, porém, na análise de resíduos e influência local no conjunto de dados de *theophylline*, por exemplo, a distribuição skew-normal apresentou mais observações aberrantes e influente do que as demais distribuições, como pode ser verificado nos gráficos de resíduos, envelopes simulados e  $\mathbf{m}(0)$ . Na análise de influência global, podemos observar que os parâmetros de assimetria sofreram as maiores influências, especialmente nas análises dos conjuntos de dados de crescimento de laranjeiras e do cefamandol. De maneira geral, fica mais clara a detecção de observações aberrantes e/ou influentes para conjuntos de dados com grandes números de observações.

Propomos também um modelo NLME com erros normalmente distribuídos e efeitos aleatórios sob a distribuição skew-normal e, conseqüentemente, a resposta segue uma distribuição skew-normal, porém a inferência é baseada no modelo marginal da variável resposta, diferentemente do modelo SMSN-NLME, que trabalhamos com a distribuição condicional da variável resposta dado o efeito aleatório. O algoritmo do tipo EM que usamos para estimação dos parâmetros nas aplicações para este modelo mostrou-se mais eficiente do que com o modelo SMSN-NLME, com convergência rápida, em virtude de um número menor de quantidades a estimar, porém o modelo é menos flexível, pois os ajustes são realizados somente sob a distribuição skew-normal. A eficiência do algoritmo do tipo EM foi observada também durante o estudo de simulação, em que o tempo para obtenção dos resultados foi bastante reduzido, em relação ao modelo SMSN-NLME.

## 7.2 Perspectivas de trabalhos futuros

Os modelos que apresentamos neste trabalho são não lineares nos efeitos fixos e lineares nos efeitos aleatórios. Assim, estão abertas algumas possibilidades para os ajustes dos dados apresentados no Capítulo 2, como uma abordagem considerando que os efeitos aleatórios são não lineares, mostrados a seguir

$$\mathbf{y}_i = \mathbf{f}(\boldsymbol{\varphi}_i, \mathbf{x}_i) + \boldsymbol{\varepsilon}_i, \quad i = 1, \dots, n, \quad (7.1)$$

em que  $\boldsymbol{\varphi}_i$  é o vetor de parâmetros de efeitos mistos modelado como

$$\boldsymbol{\varphi}_i = \boldsymbol{\alpha} + \mathbf{b}_i, \quad i = 1, \dots, n, \quad (7.2)$$

onde  $\boldsymbol{\alpha}$  é um vetor de efeitos fixos de dimensão  $p \times 1$ . Os efeitos aleatórios  $\mathbf{b}_i = (b_{i1}, \dots, b_{ip})^\top$ , e os erros  $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{im_i})^\top$  são não correlacionados.

Para as componentes aleatórias  $\mathbf{b}_i$  e  $\boldsymbol{\varepsilon}_i$  pode-se assumir, por exemplo, distribuições com caudas pesadas e considerado assimetria.

Uma extensão do modelo SMSN-NLME, é considerar outras estrutura para a matriz de dispersão dos erros e assim considerar, por exemplo, um modelo heterocedástico.

Nas aplicações apresentados no Capítulo 4 com o modelo SMSN-NLME, utilizamos a estatística da razão de verossimilhanças para realizar testes de hipóteses e construir intervalos de confiança do aproximados para o vetor  $\boldsymbol{\lambda}$ , porém, dadas as dificuldades de obtenção de distribuições exatas tanto para o estimador do vetor  $\boldsymbol{\lambda}$ , quanto para os estimadores dos vetores  $\boldsymbol{\alpha}$  e  $\boldsymbol{\beta}$ , sugerimos a utilização do método *bootstrap* (EFRON; TIBSHIRANI, 1994) para tal fim. Por exemplo, para construção de intervalo de confiança bootstrap para  $\alpha_1$  ( $\text{IC}^*(\alpha_1)$ ), pode ser utilizada o método bootstrap não paramétrico, que consistem em realizar reamostragens, com reposição, dos indivíduos  $i$ ,  $i = 1, \dots, n$ ,  $T$  vezes, e estimar  $T$  vetores  $\boldsymbol{\alpha}^*$ . Em seguida, ordenar os  $T$  valores de  $\alpha_1^*$  estimados e, para um intervalo de confiança com nível de significância de 5%, os limites inferior e superior do intervalo de confiança serão obtidos por

$$\text{IC}^*(\alpha_1, 95\%) = \left[ \widehat{\alpha}_1^*_{\lfloor \frac{0,05}{2} \times T \rfloor}, \widehat{\alpha}_1^*_{\lfloor (1 - \frac{0,05}{2}) \times T \rfloor} \right], \quad (7.3)$$

em que  $\lfloor x \rfloor$  representa a parte inteira do valor  $x$ .

Por simplicidade, pode-se considerar  $T = 1000$  e, assim, os limites inferior e superior do  $\text{IC}^*(\alpha_1, 95\%)$  são os valores que ocuparem as posições 25 e 975, respectivamente, no vetor de valores de  $\alpha_1^*$  estimados e ordenados. Tal procedimento pode ser repetido para os demais elementos de  $\boldsymbol{\alpha}$ .

Outras propostas consistem em desenvolver modelos SMSN-NLMEs em que os efeitos aleatórios são simétricos e os erros são assimétricos, como proposto por Lachos (LACHOS, 2004)

com modelos lineares e sob a distribuição skew-normal. Meza et. al (MEZA; OSORIO; CRUZ, 2012) desenvolveram um modelo NLME que são não lineares nos efeitos fixos e aleatórios e sob distribuições simétricas, cujos parâmetros são estimados com uma versão estocástica do algoritmo EM, chamado de algoritmo *stochastic approximation* EM (SAEM). Consideramos que a ideia de trabalhar com o algoritmo SAEM é bastante promissora e será mais uma boa opção para ajustar dados que não possuem relação linear com observações atípicas e assimetria, além do desenvolvimento de procedimentos estatísticos sob a perspectiva Bayesiana.

---

## REFERÊNCIAS

---

- ANDREWS, D. F.; MALLOWS, C. L. Scale mixtures of normal distributions. **Journal of the Royal Statistical Society. Series B (Methodological)**, JSTOR, p. 99–102, 1974. Citado nas páginas 24, 25, 43 e 45.
- ARELLANO-VALLE, R. B.; BOLFARINE, H.; LACHOS, V. H. Skew-normal linear mixed models. **Journal of Data Science** 3, 2005. Citado nas páginas 120 e 121.
- ARELLANO-VALLE, R. B.; GENTON, M. G. On fundamental skew distributions. **Journal of Multivariate Analysis**, Elsevier, v. 96, n. 1, p. 93–116, 2005. Citado nas páginas 43, 50 e 118.
- ATKINSON, A. Two graphical displays for outlying and influential observations in regression. **Biometrika**, JSTOR, p. 13–20, 1981. Citado nas páginas 90 e 145.
- ATKINSON, A. C. **Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis**. [S.l.]: Clarendon Press Oxford, 1985. Citado na página 146.
- AZZALINI, A. **The R package sn: The Skew-Normal and Related Distributions such as the Skew-t (version 1.5-2)**. Università di Padova, Italia, 2018. Disponível em: <<http://azzalini.stat.unipd.it/SN>>. Citado nas páginas 26, 54 e 90.
- AZZALINI, A.; CAPITANIO, A. Statistical applications of the multivariate skew normal distribution. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 61, n. 3, p. 579–602, 1999. Citado nas páginas 23, 46 e 50.
- AZZALINI, A.; GENTON, M. G. Robust likelihood methods based on the skew-t and related distributions. **International Statistical Review**, Wiley Online Library, v. 76, n. 1, p. 106–129, 2008. Citado na página 51.
- AZZALINI, A.; VALLE, A. D. The multivariate skew-normal distribution. **Biometrika**, Biometrika Trust, v. 83, n. 4, p. 715–726, 1996. Citado na página 44.
- BASSO, R. M. **Misturas finitas de misturas de escala skew-normal**. Dissertação (Mestrado) — Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, 2009. Citado nas páginas 47, 48 e 57.
- BECKMAN, R. J.; NACHTSHEIM, C. J.; COOK, R. D. Diagnostics for mixed-model analysis of variance. **Technometrics**, Taylor & Francis Group, v. 29, n. 4, p. 413–426, 1987. Citado na página 24.
- BRANCO, M. D.; DEY, D. K. A general class of multivariate skew-elliptical distributions. **Journal of Multivariate Analysis**, Elsevier, v. 79, n. 1, p. 99–113, 2001. Citado nas páginas 25, 43, 46 e 50.
- CABRAL, C. R. B.; LACHOS, V. H.; PRATES, M. O. Multivariate mixture modeling using skew-normal independent distributions. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 1, p. 126–142, 2012. Citado na página 46.

- CANCHO, V. G.; LACHOS, V. H.; ORTEGA, E. M. A nonlinear regression model with skew-normal errors. **Statistical papers**, Springer, v. 51, n. 3, p. 547–558, 2010. Citado na página 24.
- CASELLA, G.; BERGER, R. L. **Statistical inference**. California, USA: Duxbury Thompson Learning, 2001. Citado na página 140.
- COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society. Series B (Methodological)**, Wiley-Blackwell for the Royal Statistical Society, v. 48, n. 2, p. p. 133–169, 1986. ISSN 00359246. Citado nas páginas 26, 84, 85 e 86.
- CRUZ, R. De la; BRANCO, M. D. Bayesian analysis for nonlinear regression model under skewed errors, with application in growth curves. **Biometrical Journal: Journal of Mathematical Methods in Biosciences**, Wiley Online Library, v. 51, n. 4, p. 588–609, 2009. Citado na página 25.
- DAVIDIAN, M.; GILTINAN, D. M. **Nonlinear models for repeated measurement data**. [S.l.]: CRC press, 1995. v. 62. Citado na página 34.
- DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the royal statistical society. Series B (methodological)**, JSTOR, p. 1–38, 1977. Citado nas páginas 25, 64, 139 e 140.
- DUNN, P. K.; SMYTH, G. K. Randomized quantile residuals. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 5, n. 3, p. 236–244, 1996. Citado nas páginas 84 e 145.
- EFRON, B.; TIBSHIRANI, R. J. **An introduction to the bootstrap**. [S.l.]: CRC press, 1994. Citado na página 131.
- ESCOBAR, L. A.; MEEKER, W. Q. Assessing influence in regression analysis with censored data. **Biometrics**, JSTOR, p. 507–528, 1992. Citado nas páginas 86 e 87.
- FERREIRA, C. da S.; BOLFARINE, H.; LACHOS, V. H. Skew scale mixtures of normal distributions: properties and estimation. **Statistical Methodology**, Elsevier, v. 8, n. 2, p. 154–171, 2011. Citado na página 46.
- FERREIRA, C. S.; LACHOS, V. H. Nonlinear regression models under skew scale mixtures of normal distributions. **Statistical Methodology**, Elsevier, v. 33, p. 131–146, 2016. Citado na página 24.
- GARAY, A. M.; LACHOS, V. H.; ABANTO-VALLE, C. A. Nonlinear regression models based on scale mixtures of skew-normal distributions. **Journal of the Korean Statistical Society**, Elsevier, v. 40, n. 1, p. 115–124, 2011. Citado na página 24.
- GENZ, A.; BRETZ, F. **Computation of multivariate normal and t probabilities**. [S.l.]: Springer Science & Business Media, 2009. v. 195. Citado nas páginas 26, 79 e 80.
- GIBALDI, M.; PERRIER, D. Noncompartmental analysis based on statistical moment theory. **Pharmacokinetics**, Marcel Dekker New York, v. 2, p. 409–417, 1982. Citado na página 34.
- GILBERT, P.; VARADHAN, R. **numDeriv: Accurate Numerical Derivatives**. [S.l.], 2016. R package version 2016.8-1. Disponível em: <<https://CRAN.R-project.org/package=numDeriv>>. Citado nas páginas 26, 68, 71 e 74.

- GILL, P. S. A robust mixed linear model analysis for longitudinal data. **Statistics in medicine**, Wiley Online Library, v. 19, n. 7, p. 975–987, 2000. Citado na página 24.
- HARTLEY, H. O. Maximum likelihood estimation from incomplete data. **Biometrics**, JSTOR, v. 14, n. 2, p. 174–194, 1958. Citado na página 139.
- HUANG, Y.; DAGNE, G. A bayesian approach to joint mixed-effects models with a skew-normal distribution and measurement errors in covariates. **Biometrics**, Wiley Online Library, v. 67, n. 1, p. 260–269, 2011. Citado na página 24.
- JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. **Continuous univariate distributions**, vol. 1. [S.l.]: Wiley, New York, 1994. Citado na página 55.
- KIM, H.-M. A note on scale mixtures of skew normal distribution. **Statistics & Probability Letters**, Elsevier, v. 78, n. 13, p. 1694–1701, 2008. Citado na página 25.
- LACHOS, V. H. **Modelos lineares mistos assimétricos**. Tese (Doutorado) — Instituto de Matemática e Estatística da Universidade de São Paulo, 27/10/2004., 2004. Citado nas páginas 55, 118 e 131.
- LACHOS, V. H.; BANDYOPADHYAY, D.; DEY, D. K. Linear and nonlinear mixed-effects models for censored hiv viral loads using normal/independent distributions. **Biometrics**, Wiley Online Library, v. 67, n. 4, p. 1594–1604, 2011. Citado na página 24.
- LACHOS, V. H.; BOLFARINE, H.; ARELLANO-VALLE, R. B.; MONTENEGRO, L. C. Likelihood-based inference for multivariate skew-normal regression models. **Communications in Statistics—Theory and Methods**, Taylor & Francis, v. 36, n. 9, p. 1769–1786, 2007. Citado na página 121.
- LACHOS, V. H.; GHOSH, P.; ARELLANO-VALLE, R. B. Likelihood based inference for skew-normal independent linear mixed models. **Statistica Sinica**, JSTOR, p. 303–322, 2010. Citado nas páginas 24, 25, 46, 48, 50, 65 e 66.
- LACHOS, V. H.; LABRA, F. V. **Skew-normal/independent Distributions, with Applications**. [S.l.]: Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica, 2007. Citado na página 144.
- LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. **Biometrics**, JSTOR, p. 963–974, 1982. Citado na página 24.
- LANGE, K.; SINSHEIMER, J. S. Normal/independent distributions and their applications in robust regression. **Journal of Computational and Graphical Statistics**, Taylor & Francis Group, v. 2, n. 2, p. 175–198, 1993. Citado na página 45.
- LEE, S.-Y.; XU, L. Influence analyses of nonlinear mixed-effects models. **Computational Statistics & Data Analysis**, Elsevier, v. 45, n. 2, p. 321–341, 2004. Citado nas páginas 24 e 87.
- LEYDOLD, J.; HÖRMANN, W. **Runuran: R Interface to the UNU.RAN Random Variate Generators**. [S.l.], 2017. R package version 0.24. Disponível em: <<https://CRAN.R-project.org/package=Runuran>>. Citado nas páginas 26 e 52.
- LIN, T. I.; LEE, J. C. Estimation and prediction in linear mixed models with skew-normal random effects for longitudinal data. **Statistics in medicine**, Wiley Online Library, v. 27, n. 9, p. 1490–1507, 2008. Citado na página 49.

- LINDSTROM, M. J.; BATES, D. M. Nonlinear mixed effects models for repeated measures data. **Biometrics**, JSTOR, p. 673–687, 1990. Citado na página 23.
- LIU, C.; RUBIN, D. B. The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. **Biometrika**, JSTOR, p. 633–648, 1994. Citado nas páginas 64 e 140.
- LIU, W.; WU, L. Simultaneous inference for semiparametric nonlinear mixed-effects models with covariate measurement errors and missing responses. **Biometrics**, Wiley Online Library, v. 63, n. 2, p. 342–350, 2007. Citado na página 25.
- LU, X.; HUANG, Y. Bayesian analysis of nonlinear mixed-effects mixture models for longitudinal data with heterogeneity and skewness. **Statistics in medicine**, Wiley Online Library, v. 33, n. 16, p. 2830–2849, 2014. Citado na página 25.
- MATOS, L. A.; LACHOS, V. H.; BALAKRISHNAN, N.; LABRA, F. V. Influence diagnostics in linear and nonlinear mixed-effects models with censored data. **Computational Statistics & Data Analysis**, Elsevier, v. 57, n. 1, p. 450–464, 2013. Citado na página 24.
- MENG, X. L.; RUBIN, D. B. Maximum likelihood estimation via the ECM algorithm: A general framework. **Biometrika**, Biometrika Trust, v. 80, n. 2, p. 267–278, 1993. Citado nas páginas 64 e 140.
- MEZA, C.; OSORIO, F.; CRUZ, R. De la. Estimation in nonlinear mixed-effects models using heavy-tailed distributions. **Statistics and Computing**, Springer, v. 22, n. 1, p. 121–139, 2012. Citado nas páginas 24 e 132.
- MONTENEGRO, L. C.; LACHOS, V. H.; BOLFARINE, H. Inference for a skew extension of the grubbs model. **Statistical Papers**, Springer, v. 51, n. 3, p. 701–715, 2010. Citado na página 67.
- MORAL, R. A.; HINDE, J.; DEMÉTRIO, C. G. B. Half-normal plots and overdispersed models in R: The hnp package. **Journal of Statistical Software**, v. 81, n. 10, p. 1–23, 2017. Citado nas páginas 26 e 90.
- PAULA, G. A. Influence diagnostics in proper dispersion models. **Australian Journal of Statistics**, Wiley Online Library, v. 38, n. 3, p. 307–316, 1996. Citado na página 24.
- PAWITAN, Y. **In all likelihood: statistical modelling and inference using likelihood**. [S.l.]: Oxford University Press, 2001. Citado na página 67.
- PINHEIRO, J.; BATES, D.; DEBROY, S.; SARKAR, D.; R Core Team. **nlme: Linear and Nonlinear Mixed Effects Models**. [S.l.], 2018. R package version 3.1-137. Disponível em: <<https://CRAN.R-project.org/package=nlme>>. Citado nas páginas 26, 29, 32, 34 e 39.
- PINHEIRO, J. C.; BATES, D. M. **Mixed-Effect models in S and S-PLUS**. New York: Springer-Verlag, 2000. Citado nas páginas 29, 34, 35 e 38.
- PINHEIRO, J. C.; LIU, C.; WU, Y. N. Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate t distribution. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 10, n. 2, p. 249–276, 2001. Citado na página 53.
- POON, W.-Y.; POON, Y. S. Conformal normal curvature and assessment of local influence. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 61, n. 1, p. 51–61, 1999. Citado nas páginas 26, 86 e 87.

PRATES, M. O.; CABRAL, C. R. B.; LACHOS, V. H. mixsmsn: Fitting finite mixture of scale mixture of skew-normal distributions. **Journal of Statistical Software**, v. 54, n. 12, p. 1–20, 2013. Disponível em: <<http://www.jstatsoft.org/v54/i12/>>. Citado nas páginas 26, 54, 79, 80 e 90.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>. Citado nas páginas 26, 29, 54 e 60.

RICHARDSON, A. M. Bounded influence estimation in the mixed linear model. **Journal of the American Statistical Association**, Taylor & Francis, v. 92, n. 437, p. 154–161, 1997. Citado na página 24.

RICHARDSON, A. M.; WELSH, A. H. Robust restricted maximum likelihood in mixed linear models. **Biometrics**, JSTOR, p. 1429–1439, 1995. Citado na página 24.

RUSSO, C. M. **Modelos não lineares elípticos para dados correlacionados**. Tese (Tese de Doutorado) — Departamento de Estatística, Universidade de São Paulo, São Paulo, Brasil, 2010. Citado na página 87.

RUSSO, C. M.; PAULA, G. A.; AOKI, R. Influence diagnostics in nonlinear mixed-effects elliptical models. **Comput. Stat. Data Anal.**, Elsevier Science Publishers B. V., Amsterdam, The Netherlands, The Netherlands, v. 53, n. 12, p. p. 4143–4156, Outubro 2009. ISSN 0167-9473. Citado nas páginas 23, 59 e 117.

SARKAR, D. **Lattice: Multivariate Data Visualization with R**. New York: Springer, 2008. ISBN 978-0-387-75968-5. Disponível em: <<http://lmdvr.r-forge.r-project.org>>. Citado nas páginas 26, 33, 69 e 75.

SAVALLI, C.; PAULA, G. A.; CYSNEIROS, F. J. Assessment of variance components in elliptical linear mixed models. **Statistical Modelling**, Sage Publications, v. 6, n. 1, p. 59–76, 2006. Citado na página 24.

TANG, N.-S.; WEI, B.-C.; WANG, X.-R. Local influence in nonlinear reproductive dispersion models. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 30, n. 3, p. 435–449, 2001. Citado na página 24.

TANG, N.-S.; WEI, B.-C.; ZHANG, W.-Z. Influence diagnostics in nonlinear reproductive dispersion mixed models. **Statistics**, Taylor & Francis, v. 40, n. 3, p. 227–246, 2006. Citado na página 24.

VONESH, E. F.; CARTER, R. L. Mixed-effects nonlinear regression for unbalanced repeated measures. **Biometrics**, JSTOR, p. 1–17, 1992. Citado nas páginas 59, 61, 117 e 118.

WANG, K.; NG, A.; MCLACHLAN, G. **EMMIXskew: The EM Algorithm and Skew Mixture Distribution**. [S.l.], 2018. R package version 1.0.3. Disponível em: <<https://CRAN.R-project.org/package=EMMIXskew>>. Citado nas páginas 26 e 68.

WEI, G. C.; TANNER, M. A. A monte carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. **Journal of the American statistical Association**, Taylor & Francis, v. 85, n. 411, p. 699–704, 1990. Citado nas páginas 64 e 140.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. Disponível em: <<http://ggplot2.org>>. Citado nas páginas 26 e 54.

- WILLEMSSEN, S. P.; RUSSO, C. M.; LESAFFRE, E.; LEÃO, D. Flexible multivariate nonlinear models for bioequivalence problems. **Statistical Modelling**, SAGE Publications Sage India: New Delhi, India, v. 17, n. 6, p. 449–467, 2017. Citado na página 39.
- ZACKS, S. **The theory of statistical inference**. [S.l.: s.n.], 1971. Citado na página 65.
- ZELLER, C. B. **Distribuições misturas de escala skew-normal: estimação e diagnóstico em modelos lineares**. Tese (Doutorado) — PhD thesis, Instituto de Matemática, Estatística e Computação Científica, Universidade de Campinas (in Portuguese), 2009. Citado nas páginas 43, 45, 47, 48, 53, 58, 67 e 143.
- ZELLER, C. B.; LACHOS, V. H.; VILCA-LABRA, F. E. Local influence analysis for regression models with scale mixtures of skew-normal distributions. **Journal of Applied Statistics**, Taylor & Francis, v. 38, n. 2, p. 343–368, 2011. Citado nas páginas 25, 43, 45, 47, 48, 50, 51, 52 e 56.
- ZHONG, X.; ZHAO, J.; WANG, H.; WEI, B. Influence analysis on exponential nonlinear models with random effects. **ACTA MATHEMATICA SCIENTIA**, KLUWER ACADEMIC PUBL VAN GODEWIJCKSTRAAT 30, 3311 GZ DORDRECHT, NETHERLANDS, v. 23, n. 3, p. 297–308, 2003. Citado na página 24.
- ZHU, H.-T.; LEE, S.-Y. Local influence for incomplete data models. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 63, n. 1, p. p. 111–126, 2001. Citado nas páginas 26 e 85.

---

## ALGORITMOS EM E NEWTON-RAPHSON

---

### A.1 Algoritmo EM

O algoritmo EM tem sua origem no trabalho realizado por Hartley (HARTLEY, 1958), porém após o trabalho de Dempster et al. (DEMPSTER; LAIRD; RUBIN, 1977), que detalhou sua estrutura básica e ilustrou seu uso em ampla variedade de aplicações, o algoritmo EM ganhou destaque e entrou efetivamente em proeminência estatística. Em alguns problemas estatísticos é possível utilizar uma formulação com dados aumentados, nestes casos o algoritmo EM permite a obtenção das estimativas de máxima verossimilhança de forma simples. As estimativas de máxima verossimilhança são obtidas computacionalmente e de forma iterativa e cada iteração envolve dois passos, um passo E (Esperança) e um passo M (Maximização). Para exemplificar, vamos considerar  $\mathbf{Y}$  e  $\mathbf{u}$  vetores aleatórios de dados observados e não observados, respectivamente, assim um vetor aleatório de dados completos é  $\mathbf{Y}_c = (\mathbf{Y}^\top, \mathbf{u}^\top)^\top$ , cuja função de log-verossimilhança dos dados completos é dada por  $\ell(\boldsymbol{\theta}|\mathbf{Y}_c)$ , com  $\boldsymbol{\theta}$  pertencente ao espaço paramétrico. Dessa forma, o algoritmo EM trata de problemas com dados incompletos, substituindo os dados não observados por suas esperanças condicionais, dado os dados observados e os valores ajustados de  $\boldsymbol{\theta}$ . Teremos então esperança condicional da função de log-verossimilhança dos dados completos com a forma

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = E[\ell(\boldsymbol{\theta}|\mathbf{Y}_c)|\mathbf{Y}, \hat{\boldsymbol{\theta}}]. \quad (\text{A.1})$$

A  $(k + 1)$ -ésima iteração do algoritmo EM tem os seguintes passos

**Passo E:** para  $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}^{(k)}$ , calcular

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}) = E[\ell(\boldsymbol{\theta}|\mathbf{Y}_c)|\mathbf{Y}, \hat{\boldsymbol{\theta}}].$$

**Passo M:** obter  $\hat{\boldsymbol{\theta}}^{(k+1)}$  que maximize  $Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})$  de modo que

$$Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k+1)}) \geq Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}}^{(k)}).$$

Repetir passos E e M até obter convergência para a sequência de estimativas  $\boldsymbol{\theta}$ , segundo algum critério de parada baseado na precisão desejada.

Segundo Casella e Berger (CASELLA; BERGER, 2001), o EM é um algoritmo que certamente converge para o estimador de máxima verossimilhança e tem como base a ideia de substituir uma difícil maximização da função de log-verossimilhança por uma sequência de maximizações mais fáceis, cujo limite é a resposta para o problema original.

No Passo M do processo de estimação dos parâmetros do modelo SMSN-NLME, são utilizadas maximizações condicionais (CM), maximização da função de log-verossimilhança marginal restrita (CML) e maximização da esperança da função de log-verossimilhança marginal restrita, além de integração de Monte Carlo. Por isso, vamos utilizar uma combinação do algoritmo MCEM, (WEI; TANNER, 1990) e do algoritmo ECME (LIU; RUBIN, 1994) que é uma extensão rápida do EM (DEMPSTER; LAIRD; RUBIN, 1977) e do ECM (MENG; RUBIN, 1993), que denominaremos de **algoritmo do tipo EM**.

## A.2 Algoritmo de Newton-Raphson

O método de Newton-Raphson é aplicado para a determinação do ponto em que uma função não linear assume seu máximo ou seu mínimo de forma iterativa. Inicialmente admite-se uma aproximação inicial para a solução e, na sequência do processo iterativo, outros valores na vizinhança do valor inicial são obtidos até a convergência para o valor ótimo (máximo ou mínimo), podendo-se assumir uma determinada precisão. Trata-se de um método baseado na expansão da série de Taylor, por exemplo da função de variável real  $x$  e parâmetro  $\alpha$ ,  $L(\alpha)$ , em torno do ponto ótimo. Apresentamos a seguir os principais passos para obtenção do valor ótimo via algoritmo de Newton-Raphson.

$$L(\alpha) \approx L(\alpha^{(v)}) + U_{\alpha}^{(v+1)}(\alpha - \alpha^{(v)}), \quad (\text{A.2})$$

$$L'(\alpha) \approx U_{\alpha}^{(v+1)} + K_{\alpha\alpha}^{(v+1)}(\alpha - \alpha^{(v)}), \quad (\text{A.3})$$

em que  $U_{\alpha}^{(v+1)}$  e  $K_{\alpha\alpha}^{(v+1)}$  são a primeira e segunda derivadas de  $L(\alpha)$  em relação a  $\hat{\alpha}^{(v)}$ , computadas no passo  $(v+1)$ . Fazendo  $L'(\alpha) = 0$ , teremos

$$L'(\alpha) \approx U_{\alpha}^{(v+1)} + K_{\alpha\alpha}^{(v+1)}(\alpha - \alpha^{(v)}) = 0, \quad (\text{A.4})$$

$$\hat{\alpha}^{(v+1)} = \hat{\alpha}^{(v)} - \frac{U_{\alpha}^{(v+1)}}{K_{\alpha\alpha}^{(v+1)}}, \quad (\text{A.5})$$

em que  $\hat{\boldsymbol{\alpha}}^{(v+1)}$  é uma aproximação para  $\boldsymbol{\alpha}$  no passo  $(v+1)$  e  $\mathbf{K}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(v+1)} \neq 0$ .

Um critério de parada deve ser adotado, para que as iterações do algoritmo não continuem indefinidamente, de acordo com a precisão desejada.

Em estatística, o método de Newton–Raphson é utilizado para maximizar funções log-verossimilhanças  $\mathbf{L}(\boldsymbol{\alpha}, \mathbf{x})$ , em que  $\mathbf{x}$  é uma amostra aleatória e segue uma determinada distribuição de probabilidade e  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^\top$  é o vetor de parâmetros da função densidade de probabilidade, assim

$$\hat{\boldsymbol{\alpha}}^{(v+1)} = \hat{\boldsymbol{\alpha}}^{(v)} - [\mathbf{K}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(v+1)}]^{-1} \mathbf{U}_{\boldsymbol{\alpha}}^{(v+1)}, \quad (\text{A.6})$$

em que  $\hat{\boldsymbol{\alpha}}^{(v+1)}$  é a estimativa de máxima verossimilhança de  $\boldsymbol{\alpha}$  na  $(v+1)$ -ésima iteração e  $\mathbf{U}_{\boldsymbol{\alpha}}^{(v+1)}$  e  $\mathbf{K}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(v+1)}$  são matrizes cujos elementos são da forma  $\frac{\partial \mathbf{L}(\hat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x})}{\partial \hat{\alpha}_r^{(v)}}$ ,  $r = 1, \dots, p$  e  $\frac{\partial^2 \mathbf{L}(\hat{\boldsymbol{\alpha}}^{(v)}, \mathbf{x})}{\partial \hat{\alpha}_r^{(v)} \partial \hat{\alpha}_s^{(v)}}$ ,  $r, s = 1, \dots, p$ , respectivamente, assumindo que  $\mathbf{K}_{\boldsymbol{\alpha}\boldsymbol{\alpha}}^{(v+1)}$  é não singular.

Nos ajustes com o modelo SMSN-NLME, a cada iteração  $k$  do algoritmo EM, o algoritmo de Newton-Raphson calcula estimativas dos parâmetros da função não linear em  $v$  iterações.



## RESULTADOS ADICIONAIS DO CAPÍTULO 3

### B.1 Matriz de informação de Fisher observada

Nesta seção apresentamos as expressões dos elementos da matriz de informação de Fisher observadas para as distribuições skew-normal, skew-t, skew-slash e skew-normal contaminada, como apresentados em Zeller (ZELLER, 2009).

A função de log-verossimilhança para  $\boldsymbol{\theta} = (\boldsymbol{\mu}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\lambda}^\top)^\top$  dada uma amostra observada  $\mathbf{y} = (\mathbf{y}_1^\top, \dots, \mathbf{y}_n^\top)^\top$ , tal que  $\mathbf{Y} \sim SMSN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\beta}), \boldsymbol{\lambda}; H)$ , é dada por

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \log 2 - \frac{p}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| + \log K_i(\boldsymbol{\theta}), \quad \text{com} \quad (\text{B.1})$$

$$K_i(\boldsymbol{\theta}) = \int_0^\infty \kappa^{-p/2}(u_i) \exp\left\{-\frac{1}{2}\kappa^{-1}(u_i)d_i\right\} \Phi(\kappa^{-1/2}(u_i)A_i) dH(u_i; \mathbf{v}) \quad (\text{B.2})$$

com  $d = (\mathbf{Y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$  e  $A_i = \boldsymbol{\lambda}_i^\top \boldsymbol{\Sigma}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i)$ .

Usando a notação

$$I_i^F(w) = \int_0^\infty \kappa^{-w}(u_i) \exp\left\{-\frac{1}{2}\kappa^{-1}(u_i)d_i\right\} F(\kappa^{-1/2}(u_i)A_i) dH(u_i; \mathbf{v}),$$

em que  $F$  é a função  $\Phi(\cdot)$  ou  $\phi(\cdot)$ . Então, podemos escrever que  $K_i(\boldsymbol{\theta}) = I_i^\Phi(\frac{p}{2})$ . Assim, a matriz de segundas derivadas em relação a  $\boldsymbol{\theta}$  é dada por

$$L = \sum_{i=1}^n \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = -\frac{n}{2} \frac{\partial^2 \log |\boldsymbol{\Sigma}|}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \sum_{i=1}^n \frac{1}{K_i^2(\boldsymbol{\theta})} \frac{\partial K_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial K_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top} + \sum_{i=1}^n \frac{1}{K_i(\boldsymbol{\theta})} \frac{\partial^2 K_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}, \quad (\text{B.3})$$

onde

$$\begin{aligned}\frac{\partial K_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= I_i^\phi \left( \frac{p+1}{2} \right) \frac{\partial A_i}{\partial \boldsymbol{\theta}} - \frac{1}{2} I_i^\Phi \left( \frac{p+1}{2} \right) \frac{\partial d_i}{\partial \boldsymbol{\theta}} \quad e \\ \frac{\partial^2 K_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} &= \frac{1}{4} I_i^\Phi \left( \frac{p+4}{2} \right) \frac{\partial d_i}{\partial \boldsymbol{\theta}} \frac{\partial d_i}{\partial \boldsymbol{\theta}^\top} - \frac{1}{2} I_i^\Phi \left( \frac{p+2}{2} \right) \frac{\partial^2 d_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} - \\ &\quad \frac{1}{2} I_i^\phi \left( \frac{p+3}{2} \right) \left( \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial d_i}{\partial \boldsymbol{\theta}^\top} + \frac{\partial d_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^\top} \right) - I_i^\phi \left( \frac{p+3}{2} \right) A_i \frac{\partial A_i}{\partial \boldsymbol{\theta}} \frac{\partial A_i}{\partial \boldsymbol{\theta}^\top} + \\ &\quad I_i^\phi \left( \frac{p+1}{2} \right) \frac{\partial^2 A_i}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}.\end{aligned}$$

Apresentamos a seguir as expressões de  $I_i^F$  para os casos particulares da classe SMSN de distribuições.

- **Skew-normal**

$$\begin{aligned}I_i^\Phi(w) &= \exp\left\{-\frac{1}{2}d_i\right\}\Phi(A_i) \quad e \\ I_i^\phi(w) &= \exp\left\{-\frac{1}{2}d_i\right\}\phi(A_i).\end{aligned}$$

- **Skew-t**

$$\begin{aligned}I_i^\Phi(w) &= \frac{2^w v^{v/2} \Gamma(w+v/2)}{\Gamma(v/2)(v+d_i)^{w+v/2}} T\left(\frac{A_i}{(d_i+v)^{1/2}}; \sqrt{2w+v}; 2w+v\right) \quad e \\ I_i^\phi(w) &= \frac{2^w v^{v/2}}{\sqrt{2\pi}\Gamma(v/2)} \left(\frac{1}{d_i+A_i+v}\right)^{\frac{v+2w}{2}} \Gamma\left(\frac{v+2w}{2}\right).\end{aligned}$$

- **Skew-slash**

$$\begin{aligned}I_i^\Phi(w) &= \frac{v2^{w+v}\Gamma(w+v)}{d_i^{w+v}} P_1\left(w+v, \frac{d_i}{2}\right) E(\Phi(S_i^{1/2}A_i)) \quad e \\ I_i^\phi(w) &= \frac{v2^{w+v}\Gamma(w+v)}{\sqrt{2\pi}(d_i+A_i^2)^{w+v}} P_1\left(w+v, \frac{d_i+A_i^2}{2}\right),\end{aligned}$$

em que  $S_i \sim \text{Gamma}(w+v; \frac{d_i}{2}) \mathbb{I}_{(0,1)}$ .

- **Skew-normal contaminada**

$$\begin{aligned}I_i^\Phi(w) &= \sqrt{2\pi}\{v_1 v_2^{w-1/2} \phi(\sqrt{d_i}|0, v_2^{-1}) \Phi(v_2^{1/2}A_i) + (1-v_1) \phi(\sqrt{d_i}|0, 1) \Phi(A_i)\} \quad e \\ I_i^\phi(w) &= v_1 v_2^{w-1/2} \phi(\sqrt{d_i+A_i^2}|0, v_2^{-1}) + (1-v_1) \phi(\sqrt{d_i+A_i^2}|0, 1).\end{aligned}$$

Resultados adicionais sobre a matriz de informação de Fisher observada podem ser encontrados em Lachos e Vilca-Labra ([LACHOS; LABRA, 2007](#)).

## RESULTADOS ADICIONAIS DO CAPÍTULO 5

---

### C.1 Envelopes simulados

Uma forma de interpretar os resíduos no que diz respeito a adequabilidade de um modelo de regressão, é a construção de envelopes simulados (ATKINSON, 1981), que são gráficos de probabilidade dos resíduos e as respectivas bandas de confiança simuladas. Nestes gráficos, se as bandas de confiança cobrem os resíduos, ou seja, se a maioria dos pontos estão dentro do envelope, o modelo está bem ajustado.

Na construção de um envelope simulado, considerando, por exemplo, o valor absoluto do resíduo quantílico aleatorizado  $\hat{r}q_{lk}$  (DUNN; SMYTH, 1996), podemos seguir os seguintes passos:

1. Ajustar o modelo proposto com os dados observados e calcular os resíduos  $\hat{r}q_{ij}$ ,
2. Para cada perfil  $i$ ,  $i = 1, \dots, n$ , simular um vetor de respostas levando em consideração o vetor de médias e a matriz de covariâncias obtidos no item anterior e ajustar estas respostas simuladas com o modelo proposto.
3. Calcular e ordenar os resíduos simulados  $\hat{r}q_{lk}^*$ ,
4. Repetir os três primeiros passos 19 vezes e definir  $\hat{r}q_{lk}^*$ ,  $l = 1, \dots, N$ , com  $N = \sum_{i=1}^n m_i$ , como sendo o  $l$ -ésimo valor absoluto do resíduo simulado do pertencente a  $k$ -ésima simulação,  $k = 1, \dots, 20$ ,
5. Determinar o mínimo, a mediana e o máximo dos 20 menores valores absolutos dos resíduos simulados de todas as simulações, ou seja, entre os resíduos  $\hat{r}q_{1k}^*$ ,  $k = 1, \dots, 20$ ,
6. Repetir o passo anterior para os 20 segundos menores valores absolutos ( $\hat{r}q_{2k}^*$ ), em seguida para os 20 terceiros ( $\hat{r}q_{3k}^*$ ) e assim sucessivamente até os 20 maiores valores absolutos dos resíduos  $\hat{r}q_{Nk}^*$ ,  $k = 1, \dots, 20$ . Ao final deste passo obtêm-se três vetores de tamanho  $N$ , contendo os mínimos, as medianas e os máximos resíduos, em valores absolutos.

7. Dispor no gráfico de probabilidades meio-normal dos valores absolutos dos resíduos observados  $\widehat{r}_{ij}$ , os mínimos, as medianas e os máximos dos dados simulados  $\widehat{r}_{ik}^*$  versus os correspondentes valores esperados da distribuição normal padrão.

O modelo estará bem ajustado, se os resíduos apresentarem um comportamento aleatório em torno de zero. Atkinson (ATKINSON, 1985) sugere a realização de 20 simulações. Por isso, existe 1 chance em 20 ou 5% de que o maior valor absoluto dos resíduos dos dados estudados esteja fora do envelope simulado. Alternativamente, podemos gerar um número maior de observações simuladas e escolher valores que produzam uma banda de confiança de 95%, em vez de usar o mínimo e o máximo como limites inferiores e superiores, respectivamente.

## C.2 Matriz hessiana

A seguir são apresentadas as expressões dos elementos da matriz hessiana do modelo SMSN-NLME.

Conforme apresentado no Capítulo 5, temos que  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \gamma, \boldsymbol{\beta}, \boldsymbol{\lambda}^\top)^\top$ , com  $\gamma = \sigma^2$ , desta forma, a matriz hessiana é obtida pela expressão

$$\ddot{Q}(\boldsymbol{\theta}) = \sum_{i=1}^n \ddot{Q}_i(\boldsymbol{\theta}) = \sum_{i=1}^n \ddot{Q}_{1i}(\boldsymbol{\theta}_1) + \sum_{i=1}^n \ddot{Q}_{2i}(\boldsymbol{\theta}_2),$$

onde

$$\begin{aligned} \ddot{Q}_{1i}(\boldsymbol{\theta}_1) &= -\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1 | \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_1 \partial \boldsymbol{\theta}_1^\top} \quad \text{e} \\ \ddot{Q}_{2i}(\boldsymbol{\theta}_2) &= -\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2 | \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}_2 \partial \boldsymbol{\theta}_2^\top}, \end{aligned}$$

com  $\boldsymbol{\theta}_1 = (\boldsymbol{\alpha}^\top, \sigma^2)^\top$  e  $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}, \boldsymbol{\lambda}^\top)^\top$ .

Assim,

$$\ddot{Q}_i(\boldsymbol{\theta}) = -\frac{\partial^2 Q_i(\boldsymbol{\theta} | \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = \begin{pmatrix} \ddot{Q}_{1i}(\boldsymbol{\theta}_1) & \mathbf{0} \\ \mathbf{0} & \ddot{Q}_{2i}(\boldsymbol{\theta}_2) \end{pmatrix}.$$

Para cada elemento do vetor  $\boldsymbol{\theta}$ , teremos que

$$\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1 | \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} = -\frac{1}{\sigma^2} \left[ \widehat{u}_i \mathbf{J}_i^\top \mathbf{J}_i + \text{vec}^{-1} \left( \mathbf{L}_i^\top \{ \mathbf{Z}_i(\widehat{u\mathbf{b}})_i - \widehat{u}_i \mathbf{r}_i \} \right) \right], \quad (\text{C.1})$$

$$\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1 | \widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \sigma^2} = \frac{1}{\sigma^4} \left[ \mathbf{J}_i^\top \{ \mathbf{Z}_i(\widehat{u\mathbf{b}})_i - \widehat{u}_i \mathbf{r}_i \} \right], \quad (\text{C.2})$$

$$\frac{\partial^2 Q_{1i}(\boldsymbol{\theta}_1 | \widehat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \sigma^2} = \frac{m_i}{2\sigma^4} - \frac{1}{\sigma^6} \left[ \widehat{u}_i \mathbf{r}_i^\top \mathbf{r}_i - 2\mathbf{r}_i^\top \mathbf{Z}_i(\widehat{u\mathbf{b}})_i + \text{tr} \left( \mathbf{Z}_i(\widehat{u\mathbf{b}\mathbf{b}^\top})_i \mathbf{Z}_i^\top \right) \right], \quad (\text{C.3})$$

$$\begin{aligned}
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2 | \widehat{\boldsymbol{\theta}})}{\partial \beta_r \partial \beta_k} &= \frac{1}{2} \text{tr} \left( [\Gamma^{-1} \dot{\Gamma}_{\beta_r} \Gamma^{-1} \dot{\Gamma}_{\beta_k} - \Gamma^{-1} (\dot{\mathbf{F}}_k \dot{\mathbf{F}}_r + \dot{\mathbf{F}}_r \dot{\mathbf{F}}_k)] (\mathbf{I}_q - \Gamma^{-1} \mathbf{N}_i) \right) \\
&+ \frac{1}{2} \text{tr} \left( \Gamma^{-1} (\dot{\mathbf{F}}_k \boldsymbol{\delta} \boldsymbol{\delta}^\top \dot{\mathbf{F}}_r + \dot{\mathbf{F}}_r \boldsymbol{\delta} \boldsymbol{\delta}^\top \dot{\mathbf{F}}_k) \right) (\mathbf{I}_q - \Gamma^{-1} \mathbf{N}_i) \\
&- \frac{1}{2} \text{tr} \left( \Gamma^{-1} \dot{\Gamma}_{\beta_k} \Gamma^{-1} \dot{\Gamma}_{\beta_r} \Gamma^{-1} \mathbf{N}_i \right) \\
&- \boldsymbol{\delta}^\top (\dot{\Gamma}_{\beta_r} \Gamma^{-1} \Gamma_{\beta_s} + \dot{\mathbf{F}}_s \Gamma^{-1} \Gamma_{\beta_r}) \Gamma^{-1} \left( (\widehat{ut\mathbf{b}})_i - (\widehat{ut^2})_i \boldsymbol{\Delta} \right) \\
&- (\widehat{ut^2})_i \boldsymbol{\delta}^\top \dot{\mathbf{F}}_s \Gamma^{-1} \dot{\Gamma}_{\beta_r} \boldsymbol{\delta}, \tag{C.4}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2 | \widehat{\boldsymbol{\theta}})}{\partial \beta_r \partial \lambda_s} &= \frac{1}{2} \text{tr} \left( [\Gamma^{-1} \dot{\Gamma}_{\lambda_s} \Gamma^{-1} \dot{\Gamma}_{\beta_r} + \Gamma^{-1} (\dot{\mathbf{F}}_r \dot{\mathbf{G}}_s \mathbf{F} + \mathbf{F} \dot{\mathbf{G}}_s \dot{\mathbf{F}}_r)] (\mathbf{I}_q - \Gamma^{-1} \mathbf{N}_i) \right) \\
&- \frac{1}{2} \text{tr} \left( \Gamma^{-1} \dot{\Gamma}_{\beta_r} \Gamma^{-1} \dot{\Gamma}_{\lambda_s} \Gamma^{-1} \mathbf{N}_i \right) \\
&- \dot{\boldsymbol{\delta}}_s^\top \mathbf{F} \Gamma^{-1} \Gamma_{\beta_r} \Gamma^{-1} \left( (\widehat{ut\mathbf{b}})_i - (\widehat{ut^2})_i \boldsymbol{\Delta} \right) \\
&+ \left( \dot{\boldsymbol{\delta}}_s^\top \dot{\mathbf{F}}_r \Gamma^{-1} - \boldsymbol{\delta}^\top \dot{\mathbf{F}}_r \Gamma^{-1} \dot{\Gamma}_{\lambda_s} \Gamma^{-1} \right) \left( (\widehat{ut\mathbf{b}})_i - (\widehat{ut^2})_i \boldsymbol{\Delta} \right) \\
&- (\widehat{ut^2})_i \boldsymbol{\delta}^\top \dot{\mathbf{F}}_r \Gamma^{-1} \mathbf{F} \dot{\boldsymbol{\delta}}_s \quad e \tag{C.5}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 Q_{2i}(\boldsymbol{\theta}_2 | \widehat{\boldsymbol{\theta}})}{\partial \lambda_s \partial \lambda_t} &= \frac{1}{2} \text{tr} \left( \Gamma^{-1} \dot{\Gamma}_{\lambda_s} \Gamma^{-1} \dot{\Gamma}_{\lambda_t} + \Gamma^{-1} \mathbf{F} \dot{\mathbf{G}}_{st} \mathbf{F} (\mathbf{I}_q - \Gamma^{-1} \mathbf{N}_i) \right) \\
&- \frac{1}{2} \text{tr} \left( \Gamma^{-1} \dot{\Gamma}_{\lambda_s} \Gamma^{-1} \dot{\Gamma}_{\lambda_t} \Gamma^{-1} \mathbf{N}_i \right) \\
&- \dot{\boldsymbol{\delta}}_s^\top \mathbf{F} \Gamma^{-1} \dot{\Gamma}_{\lambda_t} \Gamma^{-1} \left( (\widehat{ut\mathbf{b}})_i - (\widehat{ut^2})_i \boldsymbol{\Delta} \right) \\
&+ (\dot{\mathbf{G}}_{st} \mathbf{F} - \dot{\Gamma}_{\lambda_t} \Gamma^{-1} \dot{\Gamma}_{\lambda_s}) \Gamma^{-1} \left( (\widehat{ut\mathbf{b}})_i - (\widehat{ut^2})_i \boldsymbol{\Delta} \right) - (\widehat{ut^2})_i \dot{\boldsymbol{\delta}}_t^\top \Gamma^{-1} \dot{\boldsymbol{\delta}}_s \tag{C.6}
\end{aligned}$$

em que  $\mathbf{J}_i$  e  $\mathbf{L}_i$  são matrizes cujos elementos são  $\frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_r}$   $r = 1, \dots, p$  e  $\frac{\partial^2 \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_r \partial \alpha_s}$ ,  $r, s = 1, \dots, p$  de dimensões  $m_i \times p$  e  $m_i \times p^2$ , respectivamente,  $\mathbf{F}$  é a raiz quadrada de  $\mathbf{D}$ , ou seja,  $\mathbf{D} = \mathbf{F}^2$ ,  $\widehat{\mathbf{r}}_i = \mathbf{y}_i - \mathbf{f}(\widehat{\boldsymbol{\alpha}}, \mathbf{x}_i)$ ,  $\dot{\Gamma}_{\beta_r} = \frac{\partial \Gamma}{\partial \beta_r} = \dot{\mathbf{F}}_r \mathbf{F} + \mathbf{F} \dot{\mathbf{F}}_r - \dot{\mathbf{F}}_r \boldsymbol{\delta} \boldsymbol{\delta}^\top \mathbf{F} - \mathbf{F} \boldsymbol{\delta} \boldsymbol{\delta}^\top \dot{\mathbf{F}}_r$ ,  $\dot{\mathbf{F}}_r = \frac{\partial \mathbf{F}}{\partial \beta_r}$ ,  $r = 1, \dots, \dim(\boldsymbol{\beta})$ ,  $\dot{\Gamma}_{\lambda_s} = \frac{\partial \Gamma}{\partial \lambda_s} = -\mathbf{F} \dot{\mathbf{G}}_s \mathbf{F}$ ,  $\dot{\boldsymbol{\delta}}_s = \frac{\partial \boldsymbol{\delta}}{\partial \lambda_s}$ ,  $s = 1, \dots, q$ ,  $\widehat{\mathbf{N}}_i = (\widehat{u\mathbf{b}\mathbf{b}^\top})_i - (\widehat{ut\mathbf{b}})_i \boldsymbol{\Delta}^\top - \boldsymbol{\Delta} (\widehat{ut\mathbf{b}})_i^\top + (\widehat{ut^2})_i \boldsymbol{\Delta} \boldsymbol{\Delta}^\top$ ,  $\mathbf{I}_q$  é a matriz identidade de ordem  $q$ ,  $\dot{\mathbf{G}}_s = \frac{\partial \boldsymbol{\delta} \boldsymbol{\delta}^\top}{\partial \lambda_s}$  e  $\dot{\mathbf{G}}_{st} = \frac{\partial^2 \boldsymbol{\delta} \boldsymbol{\delta}^\top}{\partial \lambda_s \partial \lambda_t}$ .



## RESULTADOS ADICIONAIS DO CAPÍTULO 6

### D.1 Matriz hessiana

Neste Apêndice apresentamos as expressões dos elementos da matriz hessiana do modelo NLME sob a distribuição skew-normal, conforme apresentado no Capítulo 6. Considerando o vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\alpha}^\top, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\lambda}^\top)^\top$  e a reparametrização da matriz de efeitos aleatórios  $\mathbf{D}$  de forma que  $\mathbf{F} = \mathbf{D}^{1/2}$  e  $\boldsymbol{\delta}_b = \boldsymbol{\delta}$ , temos a seguinte matriz hessiana

$$\frac{\partial^2 Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} = I_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{pmatrix} I_{\alpha_1 \alpha_1} & \cdots & I_{\alpha_1 \alpha_p} & I_{\alpha_1 \sigma^2} & I_{\alpha_1 \beta_1} & \cdots & I_{\alpha_1 \beta_q} & I_{\alpha_1 \lambda_1} & \cdots & I_{\alpha_1 \lambda_q} \\ & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & & I_{\alpha_p \alpha_p} & I_{\alpha_p \sigma^2} & I_{\alpha_p \beta_1} & \cdots & I_{\alpha_p \beta_q} & I_{\alpha_p \lambda_1} & \cdots & I_{\alpha_p \lambda_q} \\ & & & I_{\sigma^2 \sigma^2} & I_{\sigma^2 \beta_1} & \cdots & I_{\sigma^2 \beta_q} & I_{\sigma^2 \lambda_1} & \cdots & I_{\sigma^2 \lambda_q} \\ & & & & I_{\beta_1 \beta_1} & \cdots & I_{\beta_1 \beta_q} & I_{\beta_1 \lambda_1} & \cdots & I_{\beta_1 \lambda_q} \\ & & & & & \ddots & \vdots & \vdots & \ddots & \vdots \\ & & & & & & I_{\beta_q \beta_q} & I_{\beta_q \lambda_1} & \cdots & I_{\beta_q \lambda_q} \\ & & & & & & & I_{\lambda_1 \lambda_1} & \cdots & I_{\lambda_1 \lambda_q} \\ & & & & & & & & \ddots & \vdots \\ & & & & & & & & & I_{\lambda_q \lambda_q} \end{pmatrix}.$$

Desta forma, os elementos da matriz hessiana são obtidos com as expressões

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} = -\frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \mathbf{J}_i^\top \mathbf{J}_i + \text{vec}^{-1} \left[ \mathbf{L}_i^\top (\hat{t}_i \mathbf{Z}_i \mathbf{F} \boldsymbol{\delta} - \mathbf{r}_i) \right] \right\}, \quad (\text{D.1})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \sigma^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n \left\{ \mathbf{r}_i^\top \mathbf{r}_i - 2\hat{t}_i \mathbf{r}_i^\top \mathbf{Z}_i \mathbf{F} \boldsymbol{\delta} + \hat{t}_i^2 \text{tr} \left( \mathbf{Z}_i \mathbf{F} \boldsymbol{\delta} \boldsymbol{\delta}^\top \mathbf{F} \mathbf{Z}_i^\top \right) \right\}, \quad (\text{D.2})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \beta_r \partial \beta_s} = -\frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \hat{t}_i^2 \text{tr} \left( \mathbf{Z}_i \dot{\mathbf{F}}_r \boldsymbol{\delta} \boldsymbol{\delta}^\top \dot{\mathbf{F}}_s \mathbf{Z}_i^\top \right) \right\}, \quad (\text{D.3})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \lambda_s \partial \lambda_t} = \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \hat{t}_i \mathbf{r}_i^\top \mathbf{Z}_i \mathbf{F} \ddot{\boldsymbol{\delta}}_{st} - \frac{\hat{t}_i^2}{2} \text{tr} \left( \mathbf{Z}_i \mathbf{F} \ddot{\mathbf{G}}_{st} \mathbf{F} \mathbf{Z}_i^\top \right) \right\}, \quad (\text{D.4})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \sigma^2} = -\frac{1}{\sigma^4} \sum_{i=1}^n \left\{ \mathbf{J}_i^\top (\mathbf{r}_i - \hat{t}_i \mathbf{Z}_i \mathbf{F} \boldsymbol{\delta}) \right\}, \quad (\text{D.5})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \beta_r} = -\frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \hat{t}_i \mathbf{J}_i^\top \mathbf{Z}_i \dot{\mathbf{F}}_r \boldsymbol{\delta} \right\}, \quad (\text{D.6})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\alpha} \partial \lambda_s} = -\frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \hat{t}_i \mathbf{J}_i^\top \mathbf{Z}_i \mathbf{F} \dot{\boldsymbol{\delta}}_r \right\}, \quad (\text{D.7})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \beta_s} = -\frac{1}{\sigma^4} \sum_{i=1}^n \left\{ \hat{t}_i \mathbf{r}_i^\top \mathbf{Z}_i \dot{\mathbf{F}}_s \boldsymbol{\delta} - \hat{t}_i^2 \text{tr} \left( \mathbf{Z}_i \dot{\mathbf{F}}_s \boldsymbol{\delta} \boldsymbol{\delta}^\top \mathbf{F} \mathbf{Z}_i^\top \right) \right\}, \quad (\text{D.8})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \lambda_s} = -\frac{1}{\sigma^4} \sum_{i=1}^n \left\{ \hat{t}_i \mathbf{r}_i^\top \mathbf{Z}_i \mathbf{F} \dot{\boldsymbol{\delta}}_r - \frac{\hat{t}_i^2}{2} \text{tr} \left( \mathbf{Z}_i \mathbf{F} \dot{\mathbf{G}}_r \mathbf{F} \mathbf{Z}_i^\top \right) \right\}, \quad (\text{D.9})$$

$$\frac{\partial Q(\boldsymbol{\theta}|\hat{\boldsymbol{\theta}})}{\partial \beta_s \partial \lambda_s} = \frac{1}{\sigma^2} \sum_{i=1}^n \left\{ \hat{t}_i \mathbf{r}_i^\top \mathbf{Z}_i \dot{\mathbf{F}}_s \dot{\boldsymbol{\delta}}_r - \hat{t}_i^2 \text{tr} \left( \mathbf{Z}_i \dot{\mathbf{F}}_s \dot{\mathbf{G}}_r \mathbf{F} \mathbf{Z}_i^\top \right) \right\}, \quad (\text{D.10})$$

em que  $\mathbf{J}_i$  e  $\mathbf{L}_i$  são matrizes cujos elementos são  $\frac{\partial \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_r}$   $r = 1, \dots, p$  e  $\frac{\partial^2 \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)}{\partial \alpha_r \partial \alpha_s}$ ,  $r, s = 1, \dots, p$  de dimensões  $m_i \times p$  e  $m_i \times p^2$ ,  $N = \sum_{i=1}^n m_i$ ,  $\mathbf{r}_i = \mathbf{y}_i - \mathbf{f}(\boldsymbol{\alpha}, \mathbf{x}_i)$ ,  $\dot{\mathbf{F}}_r = \frac{\partial \mathbf{F}}{\partial \beta_r}$ ,  $r = 1, \dots, \dim(\boldsymbol{\beta})$ ,  $\dot{\boldsymbol{\delta}}_s = \frac{\partial \boldsymbol{\delta}}{\partial \lambda_s}$ ,  $s = 1, \dots, q$ ,  $\ddot{\boldsymbol{\delta}}_{st} = \frac{\partial^2 \boldsymbol{\delta}}{\partial \lambda_s \partial \lambda_t}$ ,  $\mathbf{I}_q$  é a matriz identidade de ordem  $q$ ,  $\dot{\mathbf{G}}_s = \frac{\partial \boldsymbol{\delta} \boldsymbol{\delta}^\top}{\partial \lambda_s}$  e  $\ddot{\mathbf{G}}_{st} = \frac{\partial^2 \boldsymbol{\delta} \boldsymbol{\delta}^\top}{\partial \lambda_s \partial \lambda_t}$ .

