

# Análise bayesiana de modelos lineares mistos

Aline Santos Damascena

DISSERTAÇÃO APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO TÍTULO  
DE  
MESTRE EM CIÊNCIAS

Programa: Estatística

Orientadora: Prof. Dr. Julio da Motta Singer

São Paulo, março de 2015

# **Análise bayesiana de modelos lineares mistos**

Esta é a versão original da dissertação elaborada pela  
candidata Aline Santos Damascena, tal como  
submetida à Comissão Julgadora.

# Resumo

Dados com medidas repetidas são muito comuns nas áreas de Agricultura, Medicina ou Biologia e possuem uma fonte extra de variabilidade: a variabilidade intra unidade amostral. Os Modelos Lineares Mistos ou de Efeitos Aleatórios são bem flexíveis quanto à modelagem dessa nova fonte de variabilidade, porém em muitas situações a suposição de que os erros e os efeitos aleatórios têm distribuição Normal pode não ser adequada. Do ponto de vista da inferência clássica, flexibilizar tais distribuições pode ser não trivial pois as integrais envolvidas no processo de estimação não permitem sua expressão de forma analítica. Do ponto de vista da inferência bayesiana, o processo de estimação é facilitado pois trabalhamos com as distribuições a posteriori condicionais, que em geral, são mais fáceis de se trabalhar do que as distribuições a posteriori marginais. Nesse contexto, os métodos de Monte Carlo baseados em cadeia de Markov (MCMC), que vêm se popularizando devido aos avanços computacionais, têm se tornado uma das principais ferramentas para a inferência estatística. Neste trabalho utilizamos a inferência bayesiana para analisar um conjunto de dados para os quais a suposição de que os efeitos aleatórios têm distribuição Normal não é adequada.



# Abstract

Data with repeated measurements are very common in the areas of Agriculture, Medicine and Biology and have an extra source of variability: the variability intra sampling unit. Linear Mixed Models or Random Effects Models are very flexible regarding the modeling of this new source of variability, but in many situations the assumption that the errors and/or random effects are normally distributed may not be adequate. From the point of view of classical inference, flexible such distributions can be nontrivial because the integrals involved in the estimation process does not allow its expression analytically. From the point of view of Bayesian inference, the estimation process is facilitated because we work with the conditional posterior distributions , which are generally easier to work with than the marginal posterior distributions. In this context, Monte Carlo methods based on Markov chain (MCMC), which are gaining popularity due to computational advances, have become a major tool for statistical inference. In this work we will use Bayesian inference to analyze a set of data for which the assumption that the random effects have Normal distribution is not appropriate.



# Lista de Figuras

1.1	Gráfico de perfis para o peso dos bezerros ao longo das semanas . . . . .	2
1.2	Perfis médio para a lactância de vacas Holandesas ao longo de 15 semanas de acordo com a estação do ano (A) e com a paridade (B). . . . .	3
4.1	Análise de resíduos referente ao modelo 4.1 . . . . .	16
4.2	Gráficos referentes à análise de resíduos do modelo linearizado final . . . . .	25
4.3	Gráficos referentes à análise de resíduos do modelo linearizado final . . . . .	26
4.4	Gráficos referentes à análise de resíduos do modelo linearizado final . . . . .	27
A.1	<i>Traceplot</i> e Distribuição a Posteriori de $\alpha$ no modelo 4.1 - abordagem bayesiana	31
A.2	<i>Traceplot</i> e Distribuição a Posteriori de $\beta$ no modelo 4.1 - abordagem bayesiana	32
A.3	<i>Traceplot</i> e Distribuição a Posteriori de $\sigma_a$ no modelo 4.1 - abordagem bayesiana	33
A.4	<i>Traceplot</i> e Distribuição a Posteriori de $\sigma_b$ no modelo 4.1 - abordagem bayesiana	34
A.5	<i>Traceplot</i> e Distribuição a Posteriori de $\sigma_e$ no modelo 4.1 - abordagem bayesiana	35
A.6	<i>Traceplot</i> e Distribuição a Posteriori de $\rho$ no modelo 4.1 - abordagem bayesiana	36





# Lista de Tabelas

1.1	Pesos observados em bezerros em oito semanas após o nascimento . . . . .	2
4.1	Estimativas para os parâmetros do modelo 4.1 . . . . .	15
4.2	Estimativas para os parâmetros do modelo 4.1 - abordagem bayesiana . . . . .	17
4.3	Estimativas (Erro Padrão) para os parâmetros do modelo saturado 4.9 . . . . .	21
4.4	Valores das estatísticas de ajuste para avaliação da inclusão dos efeitos aleatórios . . . . .	22
4.5	Valores das estatísticas de ajuste para avaliação da inclusão dos efeitos de interação . . . . .	22
4.6	Valores das estatísticas de ajuste para avaliação da inclusão dos efeitos da estação do ano e da paridade . . . . .	22
4.7	Estimativas (Erro Padrão) para os parâmetros do modelo final . . . . .	24



# Sumário

<b>Lista de Figuras</b>	<b>vii</b>
<b>Lista de Tabelas</b>	<b>ix</b>
<b>1 Introdução e Motivação</b>	<b>1</b>
<b>2 Abordagem Clássica</b>	<b>5</b>
2.1 Modelo Linear Misto Clássico . . . . .	5
2.1.1 Especificação do Modelo . . . . .	5
2.1.2 Estimação . . . . .	6
2.1.3 Testes de Hipótese e Seleção de Modelos . . . . .	7
2.1.4 Métodos de Diagnóstico para Modelos Lineares Mistos . . . . .	8
<b>3 Abordagem Bayesiana</b>	<b>9</b>
3.1 Introdução . . . . .	9
3.2 Especificação Bayesiana do Modelo Linear Misto . . . . .	9
3.3 Estimação . . . . .	10
3.4 Seleção de Modelos . . . . .	12
<b>4 Aplicação</b>	<b>13</b>
4.1 Exemplo 1 . . . . .	13
4.2 Exemplo 2 . . . . .	17
<b>5 Conclusões e Perspectivas</b>	<b>29</b>
<b>A Convergência e Distribuições a Posteriori</b>	<b>31</b>
<b>Referências Bibliográficas</b>	<b>37</b>



# Capítulo 1

## Introdução e Motivação

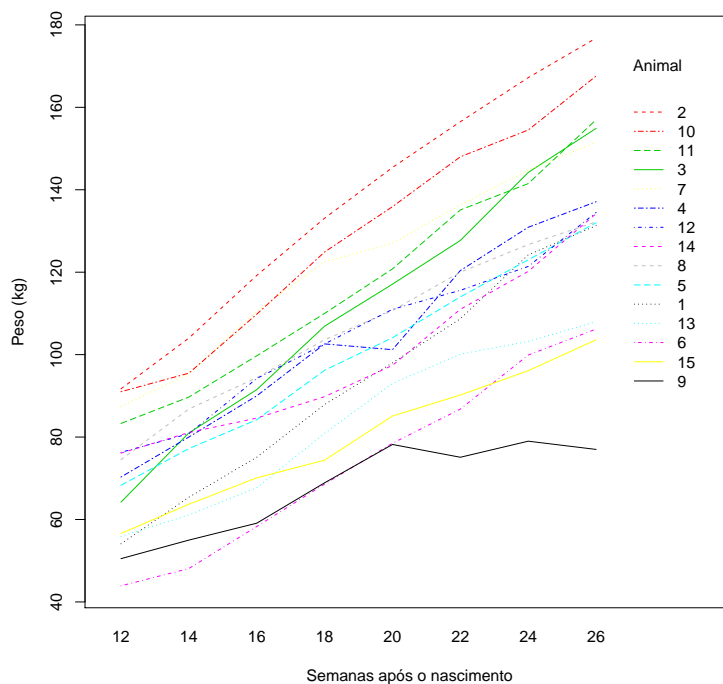
A avaliação de uma mesma unidade amostral em mais de uma condição experimental é muito comum nas áreas de Agricultura, Medicina ou Biologia. Dados desse tipo são denominados dados com medidas repetidas e possuem duas fontes de variabilidade: variabilidade entre e intra unidade amostral. Para ilustração, vamos considerar dois estudos cujos dados serão analisados ao longo desse trabalho.

### **Exemplo 1: Variação do peso de bezerros**

Na Tabela 1.1 são apresentados dados referentes a um estudo cujo objetivo era avaliar a variação do peso de 15 bezerros entre a décima segunda e vigésima sexta semana após o nascimento. É razoável supor, nesse intervalo de tempo, que o peso dos animais, aumenta linearmente com o passar das semanas como evidenciado na Figura 1.1. Note também que animais mais pesados/leves tendem a permanecer mais pesados/leves com o passar do tempo, sugerindo a correlação entre as medidas realizadas em um mesmo animal. Esse tipo de dado apresenta uma estrutura longitudinal que é um caso particular de dados com medidas repetidas: uma mesma unidade amostral é avaliada em diferentes instantes de tempo.

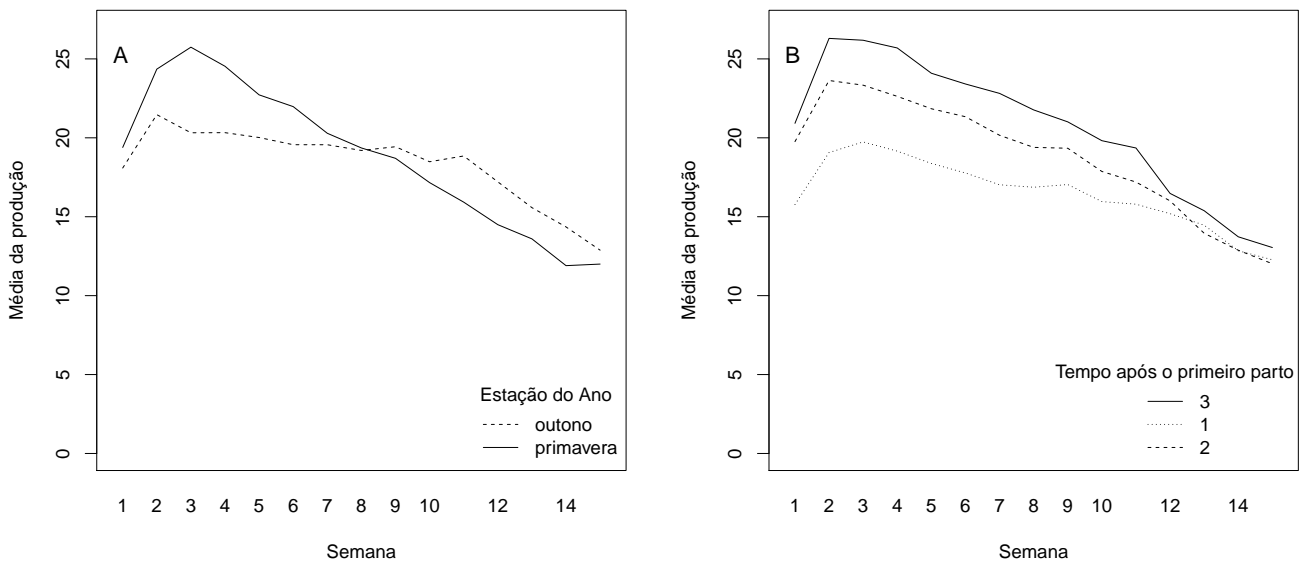
**Tabela 1.1:** *Pesos observados em bezerros em oito semanas após o nascimento*

Animal	Semanas após o nascimento							
	12	14	16	18	20	22	24	26
1	54.1	65.4	75.1	87.9	98	108.7	124.2	131.3
2	91.7	104	119.2	133.1	145.4	156.5	167.2	176.8
3	64.2	81	91.5	106.9	117.1	127.7	144.2	154.9
4	70.3	80	90	102.6	101.2	120.4	130.9	137.1
5	68.3	77.2	84.2	96.2	104.1	114	123	132
6	43.9	48.1	58.3	68.6	78.5	86.8	99.9	106.2
7	87.4	95.4	110.5	122.5	127	136.3	144.8	151.5
8	74.5	86.8	94.4	103.6	110.7	120	126.7	132.2
9	50.5	55	59.1	68.9	78.2	75.1	79	77
10	91	95.5	109.8	124.9	135.9	148	154.5	167.6
11	83.3	89.7	99.7	110	120.8	135.1	141.5	157
12	76.3	80.8	94.2	102.6	111	115.6	121.4	134.5
13	55.9	61.1	67.7	80.9	93	100.1	103.2	108
14	76.1	81.1	84.6	89.8	97.4	111	120.2	134.2
15	56.6	63.7	70.1	74.4	85.1	90.2	96.1	103.6

**Figura 1.1:** *Gráfico de perfis para o peso dos bezerros ao longo das semanas*

## Exemplo 2: Produção de leite de vacas Holandesas

A figura 1.2 representa dados referentes à produção de leite de 150 vacas Holandesas ao longo de 15 semanas após o parto. Para cada vaca foram anotados a paridade (1, 2 e 3 ou mais partos) e a estação do ano (primavera ou outono) correspondentes ao parto. Para cada combinação de paridade e estação do ano foram observadas vacas diferentes. O objetivo do estudo era avaliar se a produção de leite é influenciada pela paridade e/ou pela estação do ano.



**Figura 1.2:** Perfis médio para a lactância de vacas Holandesas ao longo de 15 semanas de acordo com a estação do ano (A) e com a paridade (B).

Nos estudos citados anteriormente notamos que uma mesma unidade amostral (bezerro e vaca, respectivamente) foi avaliada em mais de uma condição (semanas) e por isso esperamos que tais medidas apresentem uma estrutura de correlação. Considerá-la no modelo estatístico utilizado na análise dos dados melhora a precisão das estimativas dos parâmetros do modelo empregado além de permitir o estudo da variabilidade intraunidades amostrais.

Os Modelos mistos ou de Efeitos Aleatórios são bem flexíveis quanto à modelagem dessa nova fonte de variabilidade tanto quando a relação entre  $y$  e  $x$  é linear (Henderson 1984; Laird & Ware 1982; Verbeke & Lesaffre 1987) quanto não linear (Davidian & Giltinan 1995; Pinheiro & Bates 2000; Vonesh & Chinchilli 1997) nos parâmetros desconhecidos do modelo. Para os estudos citados anteriormente, um modelo linear misto parece adequado para o Estudo 1 enquanto que um modelo não linear misto parece adequado para o Exemplo 2. Em ambos os casos, uma das suposições dos modelos é que tanto os efeitos quanto os erros aleatórios têm distribuição Normal, e em muitas situações, tal suposição pode não ser adequada.

Com os modelos lineares generalizados mistos conseguimos flexibilizar a distribuição da

variável resposta (distribuições pertencentes a família exponencial) e ainda contemplar a estrutura de correlação intra unidades amostrais assumindo distribuição gaussiana para o efeito aleatório (Breslow 1993; McCulloch 2001). Lee (1996, 2001) propuseram modelos em que a distribuição do efeito aleatório pode não ser gaussiana, porém deve ser conjugada com a distribuição da variável resposta.

Do ponto de vista da inferência clássica, flexibilizar a distribuição do erro e/ou do efeito aleatório pode não ser trivial, pois facilmente as integrais envolvidas no processo de estimação dos parâmetros do modelo não permitem sua expressão de forma analítica. Já do ponto de vista da inferência bayesiana, o processo de estimação é facilitado pois trabalhamos com as distribuições a posteriori condicionais, que em geral, são mais fáceis de se trabalhar do que as distribuições a posteriori marginais. Nesse contexto, os métodos de Monte Carlo baseados em cadeia de Markov (MCMC) têm se popularizado devido aos avanços computacionais e se tornado uma das principais ferramentas computacionais para a inferência estatística.

O objetivo deste trabalho é utilizar a inferência bayesiana para analisar um conjunto de dados para os quais a suposição de que os efeitos aleatórios têm distribuição Normal não é adequada. No capítulo 2 apresentamos um resumo da teoria dos modelos lineares mistos sob o ponto de vista da inferência clássica. No capítulo 3 apresentamos um resumo da teoria dos modelos lineares mistos sob o ponto de vista da inferência bayesiana. Aplicações das abordagens clássica e bayesiana são apresentadas no capítulo 4. E finalmente, no capítulo 5 apresentamos as conclusões e as perspectivas para trabalhos futuros.



# Capítulo 2

## Abordagem Clássica

### 2.1 Modelo Linear Misto Clássico

#### 2.1.1 Especificação do Modelo

O modelo linear misto pode ser expresso como

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n \quad (2.1)$$

em que  $n$  é o número de unidades amostrais,  $\mathbf{Y}_i$  é o vetor de respostas da  $i$ -ésima unidade amostral com dimensão  $(m_i \times 1)$ ,  $m_i$  é o número de avaliações realizadas na  $i$ -ésima unidade amostral,  $\mathbf{X}_i$  é a matriz de especificação correspondente aos efeitos fixos do modelo com dimensão  $(m_i \times p)$ ,  $\boldsymbol{\beta}$  é um vetor  $(p \times 1)$  de parâmetros desconhecidos,  $\mathbf{Z}_i$  é uma matriz de especificação conhecida  $(m_i \times q)$  correspondente aos efeitos aleatórios,  $\mathbf{b}_i$  é um vetor  $(q \times 1)$  de efeitos aleatórios não observáveis e  $\boldsymbol{\epsilon}_i$  é um vetor  $(m_i \times 1)$  de erros aleatórios. Assume-se que  $\mathbf{b}_i \sim N_q(\mathbf{0}, \mathbf{G})$  e  $\boldsymbol{\epsilon}_i \sim N_{m_i}(\mathbf{0}, \mathbf{R}_i)$ , não correlacionados, sendo  $(q \times q)$  e  $(m_i \times m_i)$  as dimensões de  $\mathbf{G}$  e  $\mathbf{R}_i$ , respectivamente. As matrizes  $\mathbf{R}_i$  e  $\mathbf{G}$  são funções de parâmetros desconhecidos, ou seja,  $\mathbf{G} = \mathbf{G}(\boldsymbol{\theta}_1)$  e  $\mathbf{R}_i = \mathbf{R}_i(\boldsymbol{\theta}_2)$ , de modo que a matriz de covariâncias para a  $i$ -ésima unidade amostral,  $\mathbf{V}_i = \mathbf{V}_i(\boldsymbol{\theta})$ , dependerá de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ . Dessa forma, temos

$$\mathbf{Y}_i \sim N(\mathbf{X}_i\boldsymbol{\beta}, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^\top + \mathbf{R}_i), \quad i = 1, \dots, n \quad (2.2)$$

O modelo descrito em 2.2 pode ser visto como um modelo hierárquico ou modelo em dois estágios (Laird & Ware 1982)

$$(1) : \mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{b}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$

$$(2) : \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}) \quad (2.3)$$

No estágio 1 assumimos que  $\boldsymbol{\epsilon}_i \sim N_{m_i}(\mathbf{0}, \mathbf{R}_i)$  e que os efeitos aleatórios  $\mathbf{b}_i$  são fixos. No

estágio 2, assumimos que os  $\mathbf{b}_i$  são mutualmente independentes e independentes de  $\epsilon_i$ .

Em particular quando  $\mathbf{R}_i = \sigma^2 \mathbf{I}_{m_i}$ , em que  $\mathbf{I}_{m_i}$  é a matriz identidade de dimensão  $m_i$ , o modelo 2.1 é chamado de modelo homocedástico de efeitos aleatórios com independência condicional. Se além disso,  $\mathbf{Z}_i = \mathbf{1}_{m_i}$ , em que  $\mathbf{1}_{m_i}$  é um vetor de dimensão  $m_i$  com todos os elementos iguais a 1, a estrutura de covariância será uniforme. Nesse modelo, todas as variâncias são iguais e todas as covariâncias intraunidades amostrais são constantes.

Mais detalhes sobre modelos lineares mistos podem ser encontrados por exemplo em Davis (2002), Diggle et al. (2002), Singer & Andrade (2000), Vonesh & Chinchilli (1997) entre outros.

### 2.1.2 Estimação

Seja  $\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_n^\top)^\top$ ,  $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_n^\top)^\top$ ,  $\mathbf{Z} = \bigoplus_{i=1}^n \mathbf{Z}_i$ ,  $\mathbf{b} = (\mathbf{b}_1^\top, \dots, \mathbf{b}_n^\top)^\top$  e  $\epsilon = (\epsilon_1^\top, \dots, \epsilon_n^\top)^\top$ . O modelo 2.1 pode ser escrito na forma matricial

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \epsilon \quad (2.4)$$

$$\begin{bmatrix} \mathbf{b} \\ \epsilon \end{bmatrix} \sim N_{n+nq} \left( \begin{bmatrix} \mathbf{0}_n \\ \mathbf{0}_{nq} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{0}_{nq \times n} \\ \mathbf{0}_{n \times nq} & \boldsymbol{\Sigma} \end{bmatrix} \right) \quad (2.5)$$

implicando que  $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$  e  $Var(\mathbf{Y}) = \mathbf{V} = \mathbf{Z}\boldsymbol{\Gamma}\mathbf{Z}^\top + \boldsymbol{\Sigma}$ , em que  $\boldsymbol{\Gamma} = \mathbf{I}_n \otimes \mathbf{G}$  e  $\boldsymbol{\Sigma} = \bigoplus_{i=1}^n \mathbf{R}_i$ . Consequentemente,  $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$  e a verossimilhança marginal dos dados é

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = (2\pi)^{n/2} |V(\boldsymbol{\theta})|^{-n/2} \exp \left[ -\frac{1}{2} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right] \quad (2.6)$$

Quando  $\mathbf{V}(\boldsymbol{\theta})$  é conhecida, o estimador de máxima verossimilhança (EMV) de  $\boldsymbol{\beta}$  é dado por

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = (\mathbf{X}^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \mathbf{X})^{-1} \mathbf{X}^\top [\mathbf{V}(\boldsymbol{\theta})]^{-1} \mathbf{Y} \quad (2.7)$$

Em geral  $\mathbf{V}(\boldsymbol{\theta})$  não é conhecida e também precisa ser estimada. Para isso, precisamos maximizar 2.6 em relação a  $\boldsymbol{\beta}$  e  $\boldsymbol{\theta}$  simultaneamente. Os dois métodos de estimação mais utilizados são o Método de Máxima Verossimilhança (MV) e o Método de Máxima Verossimilhança Restrita (MVR) (Harville 1977; Patterson & Thompson 1971). Devido a natureza não linear da função de máxima verossimilhança, procedimentos iterativos como Newton-Raphson (Lindstrom & Bates 1988) ou algoritmo EM (Dempster et al, 1977) são utilizados na obtenção do estimador de MV de  $\boldsymbol{\theta}$ . Nesses procedimentos, alternam-se iterações para a estimação dos parâmetros de covariância e de  $\boldsymbol{\beta}$ . Em algumas situações o método de MVR é preferido ao método de MV, pois esse último apesar de fornecer estimativa não viesada

para  $\beta$ , fornece estimativas viesadas para os parâmetros de covariância (Diggle et al, 2002). Ambos os estimadores de  $\theta$  são assintoticamente equivalentes, porém o viés do estimador de MVR para  $\theta$  é menor em amostras pequenas/moderadas (Jiang et al, 1996; Miller, 1977; Verbeke & Lesaffre, 1996), sendo por isso mais utilizado na prática.

Métodos bayesianos de estimação também estão disponíveis e têm sido desenvolvidos por muitos autores, como Demidenko (2013), Dey et al. (2000) ou Daniels & Hogan (2008), entre outros.

### 2.1.3 Testes de Hipótese e Seleção de Modelos

O teste da razão de verossimilhanças (TRV) pode ser utilizado para comparar modelos encaixados <sup>1</sup> ajustados utilizando o método de máxima verossimilhança (MV) ou o método de máxima verossimilhança restrita (MVR). No caso da estimação via MVR, como a função de máxima verossimilhança varia com a parametrização escolhida para os efeitos fixos, o TRV só pode ser aplicado se essa parametrização for a mesma nos dois modelos que estão sendo comparados. Sendo assim, o TRV (modelos ajustados via MVR) é geralmente utilizado para testar a significância dos efeitos aleatórios do modelo. A estatística de teste do TRV

$$2\log(L_2/L_1) = 2[\log(L_2) - \log(L_1)] \quad (2.8)$$

em que  $L_i$  é a log-verossimilhança do modelo  $M_i$  e  $k_i$  é o número de parâmetros estimados no modelo  $M_i$ . A estatística 2.8 é positiva e sob a hipótese nula de que o modelo restrito M1 (com menos parâmetros) é adequado sua distribuição assintótica é  $\chi^2_{k_2-k_1}$ .

Stram & Lee (1994) argumentam que o TRV conduzido dessa forma é conservador, ou seja, a hipótese nula não é rejeitada mais vezes do que deveria. Sugerem que os resultados assintóticos para o TRV devem ser corrigidos, pois em certos casos, o modelo restrito se situa na fronteira do espaço paramétrico. Pinheiro & Bates (2000) mostram que nem sempre a correção sugerida por Stram & Lee (1994) é bem sucedida e recomendam conduzir o teste de forma conservadora.

No que se refere às hipóteses sobre os efeitos fixos do modelo, o TRV pode ser utilizado apenas quando a estimação é feita via MV, pois a função de verossimilhança restrita varia com a parametrização utilizada para os efeitos fixos. Pinheiro & Bates (2000) argumentam que, nesse caso, o TRV é não conservador e propõem a utilização de testes condicionais às estimativas das componentes de covariância dos efeitos aleatórios. Por meio de estudos simulação, mostram que o TRV é menos realista do que os testes condicionais.

O Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiana (BIC), muito utilizados na comparação de modelos encaixados ou não, são critérios que penalizam a verossimilhança com o número de parâmetros estimados e são medidas da habilidade preditiva de um modelo estatístico.

---

<sup>1</sup>Dizemos que um modelo M1 é encaixado no modelo M2 quando M1 é um caso particular ou uma redução de M2.

$$AIC = -2\log(L) + 2p$$

$$BIC = -2\log(L) + p\log(n) \quad (2.9)$$

em que  $L$  é o máximo da verossimilhança,  $n$  é o número total de observações e  $p$  é o número de parâmetros estimados. Devemos dar preferência aos modelos com menores valores de AIC e BIC.

Em modelos com efeitos aleatórios podemos considerar dois tipos de verossimilhança: a marginal e a condicional. Em cada caso, o número de parâmetros estimados é diferente: no segundo caso, os efeitos aleatórios são preditos enquanto que no primeiro caso, não. [Vaida & Blanchard \(2005\)](#) mostram que em modelos com efeitos aleatórios o uso do AIC considerando a verossimilhança marginal não é apropriado quando o interesse está em avaliar a habilidade preditiva do modelo para as unidades amostrais consideradas no próprio estudo, ou seja, quando o interesse está nos próprios efeitos aleatórios. [Vaida & Blanchard \(2005\)](#) propõem o AIC condicional, baseado na verossimilhança condicional, que envolve os efeitos aleatórios. Quando o interesse for avaliar a habilidade preditiva do modelo para observações futuras, então a verossimilhança marginal deve ser considerada no cálculo do AIC.

### 2.1.4 Métodos de Diagnóstico para Modelos Lineares Mistos

Por meio da análise dos resíduos podemos verificar se as suposições de normalidade, homocedasticidade e independência são adequadas. Em modelos lineares mistos a análise de resíduos é mais complexa do que em modelos lineares usuais devido existência de duas fontes de variação: a intraunidade amostral e a interunidades amostrais. [Nobre & Singer \(2007\)](#) apresentam as definições a seguir para os resíduos em modelos lineares mistos e resumem o uso de cada tipo de resíduo para a verificação das hipóteses do modelo.

- (i) resíduo marginal  $\widehat{\xi}_i = y_i - x_i^t \widehat{\beta} - Z \widehat{b}_i$ , que prediz o erro marginal  $\epsilon_i = y_i - E(y_i) = y - x_i^t \beta = z_i^t b_i + \epsilon_i$ ;
- (ii) resíduo condicional:  $\widehat{\epsilon}_i = y_i - x_i^t \widehat{\beta}$ , que prediz o erro condicional  $\epsilon_i = y_i - E(y_i | b_i) = y_i - x_i^t \beta - z_i^t b_i$ ;
- (iii) BLUP (*Best Linear Unbiased Predictor*):  $z_i^t \widehat{b}_i$ , que prediz os efeitos aleatórios  $z_i^t b_i = E(y_i | b_i) - E(y_i)$ .

Foi desenvolvida uma sub-rotina em linguagem R para construir os gráficos citados na Tabela 1 de [Nobre & Singer \(2007\)](#) referentes à análise de resíduos em modelos lineares mistos. A sub-rotina está disponível em [www.ime.usp.br/jmsinger](http://www.ime.usp.br/jmsinger) e foi utilizada nas análises apresentadas nesse trabalho.

# Capítulo 3

## Abordagem Bayesiana

### 3.1 Introdução

No contexto bayesiano, inferências sobre parâmetros desconhecidos são feitas por meio da função distribuição a posteriori. De forma bem simplificada, seja  $\mathbf{y}$  uma variável aleatória com densidade  $p(\mathbf{y}|\boldsymbol{\theta})$  que depende do parâmetro a ser estimado  $\boldsymbol{\theta}$ , cuja distribuição a priori é  $p(\boldsymbol{\theta})$ . Podemos traçar inferências sobre  $\boldsymbol{\theta}$  por meio de sua distribuição a posteriori, que de acordo com o Teorema de Bayes, é dada por:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta})}{p(\mathbf{y})} \quad (3.1)$$

A distribuição a posteriori leva em conta tanto a informação proveniente da distribuição a priori  $p(\boldsymbol{\theta})$  quanto a informação observada nos dados por intermédio da função de verossimilhança  $p(\mathbf{y}|\boldsymbol{\theta})$ . Como  $p(\mathbf{y})$  não depende de  $\boldsymbol{\theta}$ , temos

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{y}|\boldsymbol{\theta}) \quad (3.2)$$

em que " $\propto$ " significa "proporcional a". Um estimador bayesiano para  $\boldsymbol{\theta}$  é obtido minimizando o erro esperado com respeito a distribuição a posteriori para  $\boldsymbol{\theta}$ . Para isso devemos escolher alguma função de perda. Se utilizarmos a função de perda quadrática  $h(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^2$ , o estimador de Bayes para  $\boldsymbol{\theta}$  é a média da distribuição a posteriori para  $\boldsymbol{\theta}$ ,

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{y}) \quad (3.3)$$

Mais detalhes sobre inferência bayesiana podem ser encontrados em [DeGroot & Schervish \(2012\)](#) ou em [Paulino et al. \(2003\)](#), por exemplo.

### 3.2 Especificação Bayesiana do Modelo Linear Misto

Do ponto de vista bayesiano, os modelos lineares mistos podem ser vistos como modelos em três estágios, em que no terceiro estágio são especificadas as distribuições a priori  $p(\boldsymbol{\beta})$ ,

$p(\mathbf{R}_i)$  e  $p(\mathbf{G})$  para  $\boldsymbol{\beta}$ ,  $\mathbf{G}$  e  $\mathbf{R}_i$ , respectivamente (Davidian & Giltinan 1995). A formulação bayesiana para o modelo 2.3 é dada por

$$\begin{aligned} \mathbf{Y}_i | \mathbf{b}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i) \\ \mathbf{b}_i &\sim N_q(\mathbf{0}, \mathbf{G}) \\ \boldsymbol{\beta} &\sim p(\boldsymbol{\beta}), \quad \mathbf{G} \sim p(\mathbf{G}) \quad e \quad \mathbf{R}_i \sim p(\mathbf{R}_i) \end{aligned} \quad (3.4)$$

que pode ser reescrito como

$$\begin{aligned} \mathbf{Y}_i &= \boldsymbol{\theta}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i) \\ \boldsymbol{\theta}_i &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i, \quad \mathbf{b}_i \sim N(\mathbf{0}, \mathbf{G}) \\ \boldsymbol{\beta} &\sim p(\boldsymbol{\beta}), \quad \mathbf{G} \sim p(\mathbf{G}) \quad e \quad \mathbf{R}_i \sim p(\mathbf{R}_i) \end{aligned} \quad (3.5)$$

Geralmente, distribuições a priori vagas (com grande variabilidade) ou localmente uniformes são atribuídas para  $\boldsymbol{\beta}$ ,  $\mathbf{G}$  e  $\mathbf{R}_i$ , na tentativa de refletir o pouco conhecimento a priori sobre esses parâmetros. Outra estratégia muito utilizada é a escolha de distribuições a priori conjugadas <sup>1</sup>, o que pode facilitar a obtenção das distribuições marginais dos parâmetros.

### 3.3 Estimação

Definidas as distribuições a priori, e assumindo independência entre elas, de (3.2) temos que a distribuição posteriori conjunta é proporcional a

$$p(\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G} | \mathbf{Y}) \propto \prod_{i=1}^n p(\mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}) \prod_{i=1}^n p(\boldsymbol{\beta}) p(\mathbf{R}_i) p(\mathbf{G}) \quad (3.6)$$

em que  $p(\mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G})$  é a função de verossimilhança de  $\mathbf{Y}_i$ .

Para realizar inferências sobre um determinado parâmetro precisamos obter a densidade a posteriori marginal desse parâmetro integrando a densidade a posteriori conjunta em relação ao demais parâmetros, ou seja

$$p(\boldsymbol{\beta} | \mathbf{Y}, \mathbf{R}_i, \mathbf{G}) = \int_{\mathbf{R}_i, \mathbf{G}} \prod_{i=1}^n p(\mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}) \prod_{i=1}^n p(\boldsymbol{\beta}) p(\mathbf{R}_i) p(\mathbf{G}) \quad (3.7)$$

$$p(\mathbf{G} | \mathbf{Y}, \mathbf{R}_i, \boldsymbol{\beta}) = \int_{\mathbf{R}_i, \boldsymbol{\beta}} \prod_{i=1}^n p(\mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}) \prod_{i=1}^n p(\boldsymbol{\beta}) p(\mathbf{R}_i) p(\mathbf{G}) \quad (3.8)$$

---

<sup>1</sup>Dizemos que uma distribuição é conjugada relativamente a uma família de distribuições se a distribuição a posteriori de interesse pertencer a mesma família de distribuições (tentar melhorar a explicação)

$$p(\mathbf{R}_i | \mathbf{Y}, \boldsymbol{\beta}, \mathbf{G}) = \int_{\boldsymbol{\beta}, \mathbf{G}} \prod_{i=1}^n p(\mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G}) \prod_{i=1}^n p(\boldsymbol{\beta}) p(\mathbf{R}_i) p(\mathbf{G}) \quad (3.9)$$

Na maioria dos casos, as integrais (3.7), (3.8) e (3.9) não permitem sua expressão de forma analítica.

Uma alternativa viável, que nas últimas décadas têm ganhado cada vez mais destaque devido aos avanços computacionais, é o cálculo de integrais por meio de métodos de simulação. Dentre os mais populares estão os métodos de Monte Carlo baseados em cadeias de Markov (MCMC - *Markov Chain Monte Carlo*) que tem se tornado uma das principais ferramentas computacionais para a inferência estatística.

Uma cadeia de Markov é um processo estocástico  $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)}, \dots, \boldsymbol{\theta}^{(T)})$  tal que

$$f(\boldsymbol{\theta}^{(t+1)} | \boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}^{(t-1)}, \dots, \boldsymbol{\theta}^{(1)}) = f(\boldsymbol{\theta}^{(t+1)} | \boldsymbol{\theta}^{(t)})$$

ou seja, a distribuição de  $\boldsymbol{\theta}$  na iteração  $t + 1$ , dado todos os valores anteriores de  $\boldsymbol{\theta}$ , é igual a distribuição de  $\boldsymbol{\theta}$  quando é dado apenas o valor de  $\boldsymbol{\theta}$  na iteração  $t$ ,  $\forall t$ . Quando a cadeia de Markov é irredutível, aperiódica e recorrente-positiva, conforme  $t \rightarrow \infty$ , a distribuição de  $\boldsymbol{\theta}^{(t)}$  converge para sua distribuição de equilíbrio, que é independente de  $\boldsymbol{\theta}^{(0)}$  (valor inicial da cadeia). Mais detalhes podem ser encontrados em [Gilks et al. \(1996\)](#), por exemplo.

Para gerar amostras de  $f(\boldsymbol{\theta} | \mathbf{y})$  precisamos construir uma cadeia de Markov de modo que seja fácil gerar amostras de  $f(\boldsymbol{\theta}^{(t+1)} | \boldsymbol{\theta}^{(t)})$  e sua distribuição de equilíbrio seja a distribuição a posteriori  $f(\boldsymbol{\theta} | \mathbf{y})$  de interesse. Nessas condições, atribuímos um valor inicial  $\boldsymbol{\theta}^{(0)}$ , geramos  $T$  valores até que a convergência da cadeia tenha sido alcançada, descartamos as  $B$  primeiras observações na tentativa de diminuir a influência do valor inicial nos valores gerados (*burnin period*), consideramos as  $(T - B)$  observações subsequentes como amostras da distribuição a posteriori de interesse e calculamos medidas resumo (média, desvio padrão e quantis) e construímos gráficos para representá-la.

Dentre os algoritmos de MCMC, o de Metropolis-Hastings ([Hastings 1970](#); [Metropolis et al., 1953](#)) e o amostrador de Gibbs (*Gibbs Sampler*) ([Gelfand & Smith 1990](#); [Geman & Geman, 1984](#)) são os mais populares. Nesse último, a ideia é gerar amostras das densidades a posteriori condicionais (condicionadas nos dados e nos demais parâmetros) que podem ser mais fáceis de se calcular ao invés de gerar amostras diretamente das densidades a posteriori marginais que podem ser difíceis de se calcular.

Nesse contexto, a distribuição a posteriori conjunta (3.6) pode ser reescrita como

$$p(\boldsymbol{\beta}, \mathbf{R}_i, \mathbf{G} | \mathbf{Y}) \propto \prod_{i=1}^n p(\mathbf{Y}_i | \boldsymbol{\beta}, \mathbf{R}_i) p(\mathbf{b}_i | \mathbf{G}) \prod_{i=1}^n p(\boldsymbol{\beta}) p(\mathbf{R}_i) p(\mathbf{G}) \quad (3.10)$$

Assumindo-se que os erros e os efeitos aleatórios possuem distribuição normal e que as componentes de variância são conhecidas, [Lindley & Smith \(1972\)](#) e [Fearn \(1975\)](#) mostraram analiticamente que tanto os efeitos fixos quanto os aleatórios também seguem distribuições

normais.

Cálculos das distribuições a posteriori condicionais, quando se assume que os erros e os efeitos aleatórios seguem distribuições normais e quando as componentes de variância são desconhecidas podem ser encontrados em [Seltzer \(1993\)](#), [Seltzer et al. \(1996\)](#), [Sorensen \(2002\)](#) e [Congdon \(2010\)](#), por exemplo.

Quando os modelos são mais complicados, para gerar amostras das distribuições a posteriori, podemos utilizar programas como WinBUGS ([Lunn et al, 2000](#); [Spiegelhalter et al, 2003](#)), JAGS ([Plummer, 2011](#)) ou STAN ([Stan Development Team 2014](#)) nos quais estão implementados os principais algoritmos de MCMC. Nessa dissertação, utilizamos o programa JAGS e o pacote *runjags* do software R ([R Core Team 2014](#)).

### 3.4 Seleção de Modelos

O Critério de Informação Deviance (DIC), proposto por [Spiegelhalter et al. \(2002\)](#) como um critério de seleção de modelos bayesiano, é definido como

$$DIC = D(\bar{\theta}) + 2p_D = \overline{D(\theta)} + p_D \quad (3.11)$$

em que  $D(\theta) = -2\log(p(y|\theta)) + 2\log(f(y))$  é o chamado deviance bayesiano e  $p_D = \overline{D(\theta)} - D(\bar{\theta})$  é a complexidade ou o número de parâmetros do modelo no contexto bayesiano.  $\overline{D(\theta)}$  é a média a posteriori de  $D(\theta)$  enquanto  $D(\bar{\theta})$  é igual a  $D(\theta)$  avaliado na média a posteriori dos parâmetros.  $p(y|\theta)$  é a verossimilhança dos dados condicionado no vetor de parâmetros  $\theta$  e  $f(y)$  é uma função dos dados que independe de  $\theta$  não afetando assim o processo de comparação dos modelos.

Assim como o AIC e o BIC, o DIC pode ser utilizado tanto na comparação de modelos encaixados quanto na comparação de modelos não encaixados e tem como vantagem ser calculado durante o processo de estimação via MCMC. Devemos dar preferência ao modelo com menor valor de DIC.

De modo análogo ao AIC, a escolha da verossimilhança utilizada no cálculo DIC (condicional ou marginal) depende do objetivo da análise: avaliar a habilidade preditiva do modelo para as próprias unidades amostrais (condicional) ou avaliar a habilidade preditiva do modelo para unidades amostrais futuras (marginal). [Vaida & Blanchard \(2005\)](#) mostram que existe uma relação próxima entre o AIC condicional e o DIC, quando são consideradas distribuições a priori não informativas e as componentes de variância são conhecidas.



# Capítulo 4

## Aplicação

### 4.1 Exemplo 1

Para o Exemplo 1, propomos o seguinte modelo para avaliar a evolução do peso dos bezerros

$$y_{ij} = \alpha + \beta(t_{ij} - 12) + a_i + b_i(t_{ij} - 12) + \epsilon_{ij}, \quad i = 1, \dots, 15, \quad j = 1, \dots, 8 \quad (4.1)$$

em que

$y_{ij}$ : peso do  $i$ -ésimo bezerro na  $j$ -ésima semana

$\alpha$ : peso esperado dos bezerros na décima segunda semana de vida

$\beta$ : variação esperada no peso por semana

$t_{ij}$ : semana em que foi realizada a  $j$ -ésima observação no  $i$ -ésimo bezerro

$(a_i, b_i)$ : vetor de efeitos aleatórios associadao ao  $i$ -ésimo bezerro, com  $(a_i, b_i) \sim N_2(\mathbf{0}, \mathbf{G})$

$\epsilon_{ij}$ : erro aleatório associado ao peso do  $i$ -ésimo bezerro na  $j$ -ésima semana, com  $\epsilon_{ij} \sim N(0, \sigma_e^2)$

$(a_i, b_i)$  independente de  $\epsilon_{ij}$

Em notação matricial, podemos escrever o modelo (4.1) como

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i \quad (4.2)$$

em que

$$\begin{aligned}
\mathbf{Y}_i &= (y_{i1}, y_{i2}, \dots, y_{i8})^\top \\
\boldsymbol{\beta} &= (\alpha, \beta)^\top \\
\mathbf{X}_i &= \mathbf{Z}_i = \begin{bmatrix} 1 & t_{i1} - 12 \\ 1 & t_{i2} - 12 \\ \dots & \dots \\ 1 & t_{i8} - 12 \end{bmatrix}^\top \\
\mathbf{b}_i &= (a_i, b_i)^\top \\
\boldsymbol{\epsilon}_i &= (\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{i8})^\top
\end{aligned} \tag{4.3}$$

Para modelar a correlação intraunidades amostrais propomos uma estrutura autorregressiva de primeira ordem (AR1), pois espera-se que a correlação entre duas avaliações subsequentes seja maior do que a correlação entre duas avaliações não subsequentes. Nesse caso temos a seguinte matriz de covariâncias para  $\mathbf{Y}_i$

$$\mathbf{V}_i = \mathbf{Z}_i \mathbf{G} \mathbf{Z}_i^\top + \mathbf{R}_i$$

em que  $\mathbf{R}_i = r_{ijj'}$  é definida como

$$r_{ijj'} = \left\{ \begin{array}{ll} \sigma_e^2 & \text{se } j = j', \\ \rho^{|j-j'|} & \text{se } j \neq j' \end{array} \right\}$$

sendo  $\rho$  a correlação entre medidas subsequentes.

Para a definição da estrutura de covariâncias dos efeitos aleatórios  $\mathbf{b}_i = (a_i, b_i)$  realizamos um estudo descritivo, em que para cada bezerro, estimamos os coeficientes linear e angular da reta associada e calculamos a correlação linear entre o vetor de coeficientes angulares e o vetor de coeficientes lineares. Como o valor de correlação encontrado foi pequeno ( $r = 0,42$ ), propomos uma estrutura de independência para o vetor de efeitos aleatórios  $\mathbf{b}_i = (a_i, b_i)$ , ou seja,

$$\mathbf{G} = \begin{bmatrix} \sigma_a^2 & 0 \\ 0 & \sigma_b^2 \end{bmatrix}$$

Sob esse modelo

$$\begin{aligned}
E(y_{ij}) &= \alpha + \beta(t_{ij} - 12) \\
Var(y_{ij}) &= \sigma_a^2 + \sigma_b^2(t_{ij} - 12)^2 + \sigma_e^2 \\
Cov(y_{ij}, y_{i'j'}) &= \left\{ \begin{array}{ll} \sigma_a^2 + (t_{ij} - 12)(t_{i'j'} - 12)\sigma_b^2 + \sigma_e^2 \rho^{|j-j'|} & \text{se } i = i' \text{ e } j \neq j', \\ 0 & \text{se } i \neq i' \end{array} \right\}
\end{aligned}$$

As estimativas para os parâmetros do modelo 4.1 são apresentadas na Tabela 4.1 e foram obtidas utilizando o comando *lme* do pacote *nlme* (Pinheiro et al, 2015) do software estatístico livre R (R Core Team 2014).

**Tabela 4.1:** *Estimativas para os parâmetros do modelo 4.1*

Parâmetro	Estimativa	IC(95%)		p
$\alpha$	69.60	61.85	77.36	<0.001
$\beta$	4.59	4.01	5.17	<0.001
$\sigma_a$	14.55	9.86	21.45	
$\sigma_b$	1.04	0.68	1.60	
$\sigma_e$	4.44	2.20	9.99	
$\rho$	0.71	0.08	0.93	

Realizando o teste da razão de verossimilhanças para testar a hipótese de que  $\rho = 0$ , não se rejeita essa hipótese, sugerindo que a estrutura de correlação AR1 está adequada aos dados.

A Figura 4.1 apresenta os gráficos referentes a análise de resíduos discutida em Nobre & Singer (2007) para o modelo 4.1. De forma resumida podemos dizer que:

- 1) parece razoável assumir que os resíduos marginais são homocedásticos (Figura 4.1(a))
- 2) os pesos dos animais 2 e 9 podem ser considerados discrepantes (Figura 4.1(b))
- 3) parece razoável assumir distribuição normal para os efeitos aleatórios (Figura 4.1(c))
- 4) a estrutura de covariância parece não ser adequada para os animais 4 e 9 (Figura 4.1(d))
- 5) a suposição de homocedasticidade do resíduo condicional parece ser adequada (Figura 4.1(e))
- 6) a suposição de normalidade do resíduo condicional também parece ser adequada (Figura 4.1(f))

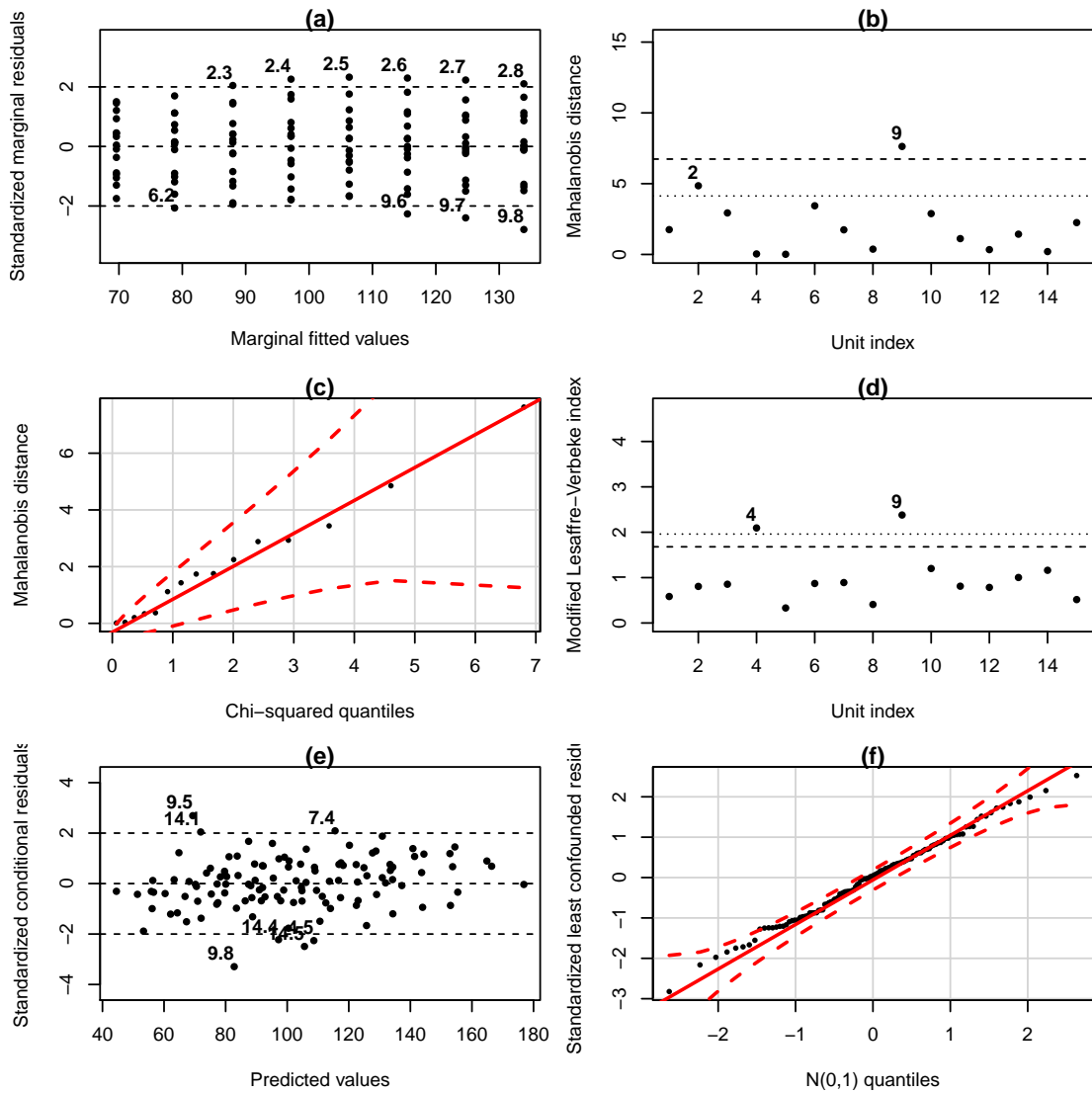


Figura 4.1: Análise de resíduos referente ao modelo 4.1

Na abordagem bayesiana, assumimos distribuições uniformes a priori para todos parâmetros. Foram geradas 3 cadeias de comprimento 15 000 e foram excluídas as 5000 primeiras iterações de cada cadeia (*burnin*). Desse modo, as estimativas apresentadas são baseadas nas 10 000 iterações subsequentes, em cada cadeia. Para avaliar a convergência das cadeias referentes a cada parâmetro, foram contruídos gráficos do tipo *traceplot* e calculada a estatística de Gelman-Rubin (EGB), sendo que valores próximos de 1 indicam convergência das cadeias (Gelman & Rubin 1992).

As estimativas para os parâmetros do modelo 4.1 quando utilizamos a abordagem bayesiana são apresentadas na Tabela 4.2 e foram obtidas utilizando o programa *JAGS* (Just Another Gibbs Sampler, Plummer (2011)) e o pacote *runjags* (Denwood In Review) do software estatístico livre R (R Core Team 2014). Os gráficos *traceplot* e os gráficos referentes às distribuições a posteriori dos parâmetros são apresentados no Apêndice A. Como esperado

as estimativas obtidas nas duas abordagens são equivalentes.

**Tabela 4.2:** Estimativas para os parâmetros do modelo 4.1 - abordagem bayesiana

Parâmetros	Mediana	p25%	p97.5%	Média	Desvio Padrão	erro MC	Gelman-Rubin
$\alpha$	69.45	61.25	78.67	69.61	4.36	0.1863	1.0012
$\beta$	4.62	3.90	5.32	4.62	0.36	0.0136	1.0117
$\sigma_a$	15.06	10.51	19.99	14.99	2.69	0.0427	1.0003
$\sigma_b$	1.16	0.69	1.82	1.21	0.31	0.0047	1.0009
$\sigma_e$	4.99	2.97	9.36	5.49	1.82	0.0996	1.0125
$\rho$	0.76	0.46	0.95	0.74	0.15	0.0074	1.0123

## 4.2 Exemplo 2

No Exemplo 2, cujo objetivo era verificar se a estação do ano e/ou a paridade influenciam a produção de leite de vacas da raça Holandesa, podemos observar que, em média, a relação entre a produção de leite e o tempo é não linear, que quanto maior o número de partos maior a produção de leite (Figura 1.2 (B)) e que a produção do outono é inferior à da primavera até aproximadamente a oitava semana, quando esse comportamento se inverte (Figura 1.2 (A)).

A função mais comumente empregada em estudos de curvas de lactação é a função gama incompleta ou função de Wood (Wood 1967)

$$y_t = \alpha t^\beta \exp^{-\gamma t} \quad (4.4)$$

em que

$y_t$ : média de produção de leite observada na semana  $t$

$\alpha$ : parâmetro relacionado à produção de leite na primeira semana de avaliação, ( $\alpha > 0$ )

$\beta$ : taxa de ascensão média da produção na fase pré pico de produção, ( $0 < \beta < 1$ )

$\gamma$ : taxa de declínio médio da produção na fase pós pico de produção, ( $0 < \gamma < 1$ )

$t$ : tempo (semana)

Uma abordagem muito utilizada no ajuste de modelos não lineares é a linearização, em que alguma transformação é aplicada ao modelo visando obter uma função linear dos parâmetros. Isso facilita a estimação, pois as técnicas convencionais para modelos lineares podem ser empregadas e ainda possibilita a realização da análise de resíduos do modelo proposto. Nesse contexto, propomos o seguinte modelo linearizável para o Exemplo 2:

$$\begin{aligned}
y_{ij}|\phi_{ij} &= \phi_{1i}t_{ij}^{\phi_{2i}} \exp(-\phi_{3i}t_{ij})\epsilon_{ij} \\
\phi_{ij} &= \mathbf{A}_{ij}\boldsymbol{\beta} + \mathbf{B}_{ij}\mathbf{b}_i
\end{aligned} \tag{4.5}$$

em que

$$\begin{aligned}
\phi_{ij} &= \left[ \phi_{1i}, \phi_{2i}, \phi_{3i} \right]^\top \\
\mathbf{A}_{ij} &= \begin{bmatrix} 1 & E_{ij} & P_{2ij} & P_{3ij} & EP_{2ij} & EP_{3ij} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & E_{ij} & P_{2ij} & P_{3ij} & EP_{2ij} & EP_{3ij} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 & E_{ij} & P_{2ij} & P_{3ij} & EP_{2ij} & EP_{3ij} \end{bmatrix}, \mathbf{0} = (0, 0, 0, 0, 0, 0) \\
\boldsymbol{\beta} &= \left[ \beta_1, \beta_2, \dots, \beta_{18} \right]^\top \\
\mathbf{B}_{ij} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
\mathbf{b}_i &= \left[ b_{1i}, b_{2i}, b_{3i} \right]^\top
\end{aligned}$$

que pode ser escrito como

$$\begin{aligned}
y_{ij} &= \phi_{1i}t_{ij}^{\phi_{2i}} \exp(-\phi_{3i}t_{ij})\epsilon_{ij}, \\
\phi_{1i} &= \exp(\beta_1 + \beta_2 E_{ij} + \beta_3 P_{2ij} + \beta_4 P_{3ij} + \beta_5 EP_{2ij} + \beta_6 EP_{3ij} + b_{1i}) \\
\phi_{2i} &= \beta_7 + \beta_8 E_{ij} + \beta_9 P_{2ij} + \beta_{10} P_{3ij} + \beta_{11} EP_{2ij} + \beta_{12} EP_{3ij} + b_{2i} \\
\phi_{3i} &= \beta_{13} + \beta_{14} E_{ij} + \beta_{15} P_{2ij} + \beta_{16} P_{3ij} + \beta_{17} EP_{2ij} + \beta_{18} EP_{3ij} + b_{3i}
\end{aligned} \tag{4.6}$$

em que

$y_{ij}$ : média de produção de leite da  $i$ -ésima vaca na semana  $j$

$\phi_{1i}$ : parâmetro relacionado à produção de leite na primeira semana de avaliação para a  $i$ -ésima vaca, ( $\phi_{1i} > 0$ )

$\phi_{2i}$ : taxa de aumento médio da produção na fase pré pico de produção da  $i$ -ésima vaca, ( $0 < \phi_{2i} < 1$ )

$\phi_{3i}$ : taxa de decréscimo médio da produção na fase pós pico de produção da  $i$ -ésima vaca, ( $0 < \phi_{3i} < 1$ )

$t_{ij}$ : semana em que foi avaliada a  $j$ -ésima produção de leite associada à  $i$ -ésima vaca

$E_{ij}$ : variável indicadora da estação do ano (0=outono, 1=primavera) para a  $i$ -ésima vaca na  $j$ -ésima semana

$P_{2ij}$ : variável indicadora de segundo parto (1 = segundo parto, 0 = caso contrário) para

a  $i$ -ésima vaca na  $j$ -ésima semana

$P_{3ij}$ : variável indicadora de terceiro ou mais partos (1 = terceiro ou mais partos, 0 = caso contrário) para a  $i$ -ésima vaca na  $j$ -ésima semana

$EP_{2ij}$ : variável indicadora de segundo parto e primavera (1 = segundo parto e primavera, 0 = caso contrário) para a  $i$ -ésima vaca na  $j$ -ésima semana

$EP_{2ij}$ : variável indicadora de terceiro ou mais partos e primavera (1 = terceiro ou mais partos e primavera, 0 = caso contrário) para a  $i$ -ésima vaca na  $j$ -ésima semana

Em 4.5 os parâmetros  $\phi_{1i}$ ,  $\phi_{2i}$  e  $\phi_{3i}$  não precisam ter um efeito aleatório associado do mesmo modo que nem todas as variáveis explicativas precisam estar relacionadas a esses parâmetros. Na formulação apresentada anteriormente, estamos admitindo que a produção inicial, a taxa de aumento pré-pico de produção e a taxa de decréscimo pós-pico de produção em cada estação do ano dependem da paridade, ou seja, estamos admitindo que existe um efeito de interação entre a estação do ano e a paridade.

O modelo (4.6) pode ser linearizado por meio da transformação

$$\begin{aligned}
 \log(y_{ij}) &= \log(\phi_{1i}) + \phi_{2i}\log(t_{ij}) - \phi_{3i}t_{ij} + \log(\epsilon_{ij}) \\
 \log(\phi_{1i}) &= \beta_1 + \beta_2E_{ij} + \beta_3P_{2ij} + \beta_4P_{3ij} + \beta_5EP_{2ij} + \beta_6EP_{3ij} + b_{1i} \\
 \phi_{2i} &= \beta_7 + \beta_8E_{ij} + \beta_9P_{2ij} + \beta_{10}P_{3ij} + \beta_{11}EP_{2ij} + \beta_{12}EP_{3ij} + b_{2i} \\
 \phi_{3i} &= \beta_{13} + \beta_{14}E_{ij} + \beta_{15}P_{2ij} + \beta_{16}P_{3ij} + \beta_{17}EP_{2ij} + \beta_{18}EP_{3ij} + b_{3i} \quad (4.7)
 \end{aligned}$$

que ainda pode ser reescrito como

$$\begin{aligned}
 \log(y_{ij}) &= \beta_1 + \beta_2E_{ij} + \beta_3P_{2ij} + \beta_4P_{3ij} + \beta_5EP_{2ij} + \beta_6EP_{3ij} + \\
 &\quad (\beta_7 + \beta_8E_{ij} + \beta_9P_{2ij} + \beta_{10}P_{3ij} + \beta_{11}EP_{2ij} + \beta_{12}EP_{3ij})\log(t_{ij}) - \\
 &\quad (\beta_{13} + \beta_{14}E_{ij} + \beta_{15}P_{2ij} + \beta_{16}P_{3ij} + \beta_{17}EP_{2ij} + \beta_{18}EP_{3ij})t_{ij} + \\
 &\quad b_{1i} + b_{2i}\log(t_{ij}) - b_{3i}t_{ij} + \log(\epsilon_{ij}) \quad (4.8)
 \end{aligned}$$

Escrevendo na forma matricial temos

$$\mathbf{Y}_i^* = \mathbf{X}_i^*\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i^*, \quad i = 1, 2, \dots, 150 \quad e \quad j = 1, 2, \dots, 15 \quad (4.9)$$

em que

$$\begin{aligned}
\mathbf{Y}_i^* &= (\log(y_{i1}), \log(y_{i2}), \dots, \log(y_{ij}))^\top, \\
\boldsymbol{\beta} &= (\beta_1, \beta_2, \dots, \beta_{18})^\top, \\
\mathbf{b}_i &= (b_{1i}, b_{2i}, b_{3i})^\top, \\
\boldsymbol{\epsilon}_i^* &= (\log(\epsilon_{i1}), \log(\epsilon_{i2}), \dots, \log(\epsilon_{ij}))^\top, \\
\mathbf{X}_i^* &= \left[ \mathbf{X} \otimes (1, 1, \dots, 1)^\top \mid \mathbf{X} \otimes (\log(t_{i1}), \log(t_{i2}), \dots, \log(t_{ij}))^\top \mid \mathbf{X} \otimes (t_{i1}, t_{i2}, \dots, t_{ij})^\top \right], \\
\mathbf{X} &= \begin{bmatrix} 1 & E_{i1} & P_{2i1} & P_{3i1} & EP_{2i1} & EP_{3i1} \\ 1 & E_{i2} & P_{2i2} & P_{3i2} & EP_{2i2} & EP_{3i2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & E_{i8} & P_{2i8} & P_{3i8} & EP_{2i8} & EP_{3i8} \end{bmatrix}, \\
\mathbf{Z}_i &= \begin{bmatrix} 1 & \log(t_{i1}) & t_{i1} \\ 1 & \log(t_{i2}) & t_{i2} \\ \vdots & \vdots & \vdots \\ 1 & \log(t_{ij}) & t_{ij} \end{bmatrix}
\end{aligned}$$

Em 4.9 estamos assumindo que  $\mathbf{b}_i = (b_{1i}, b_{2i}, b_{3i}) \sim N_3(\mathbf{0}, \mathbf{G})$  e  $\boldsymbol{\epsilon}_i^* \sim N_8(\mathbf{0}, \mathbf{R}_i)$  e são mutuamente independentes. De modo análogo ao Exemplo 1, assumimos uma estrutura de correlação AR(1) para  $\mathbf{R}_i$ .

Ajustamos o modelo saturado 4.9, por meio da função *lme* do pacote *nlme* (Pinheiro et al, 2015), para obter as estimativas do modelo 4.6. Para isso, basta exponenciar as estimativas associadas à produção inicial  $(\beta_1, \beta_2, \dots, \beta_6)$  no modelo 4.9. Os erros padrão dos parâmetros associados à produção inicial foram obtidos por meio do método Delta (Sen & Singer 1994). O teste da razão de verossimilhanças nos permite rejeitar a hipótese de que  $\rho = 0$  ( $p < 0.001$ ) indicando que a estrutura de correlação autorregressiva é mais adequada aos dados. As estimativas para os parâmetros do modelo saturado 4.6 são apresentadas na Tabela 4.3. Para as componentes de variância foram construídos intervalos com 95 % de confiança.

A seguir descrevemos a estratégia de análise utilizada para simplificar o modelo saturado 4.8:

1. Testar a significância dos efeitos aleatórios incluídos no modelo, por meio do teste da razão de verossimilhanças;
2. Testar a significância dos efeitos fixos começando pelos efeitos de interação, ou seja, testar se  $\beta_5 = \beta_6 = \beta_{11} = \beta_{12} = \beta_{17} = \beta_{18} = 0$ , por meio do teste F condicional;
3. Na ausência dos efeitos de interação, testar a significância dos efeitos principais da estação do ano e da paridade, ou seja, testar se  $\beta_2 = \beta_3 = \beta_4 = 0$ ,  $\beta_8 = \beta_9 = \beta_{10} = 0$  e  $\beta_{14} = \beta_{15} = \beta_{16} = 0$ , por meio do teste F condicional;



4. Ajustar o modelo que incorpora as conclusões obtidas em 1, 2 e 3.

**Tabela 4.3:** *Estimativas (Erro Padrão) para os parâmetros do modelo saturado 4.9*

	Parâmetro	Estimativa (Erro Padrão)
Produção inicial	$\beta_1$	15.5186 (0.9281)
	$\beta_2$	1.1079 (0.0894)
	$\beta_3$	1.2129 (0.1027)
	$\beta_4$	1.3253 (0.1092)
	$\beta_5$	1.1353 (0.1275)
	$\beta_6$	1.1136 (0.125)
Taxa de acréscimo pré pico de produção	$\beta_7$	0.1972 (0.057)
	$\beta_8$	0.2333 (0.0633)
	$\beta_9$	0.1319 (0.0744)
	$\beta_{10}$	0.157 (0.0737)
	$\beta_{11}$	-0.0501 (0.0278)
	$\beta_{12}$	-0.0352 (0.0285)
Taxa de decréscimo pós pico de produção	$\beta_{13}$	-0.0407 (0.0117)
	$\beta_{14}$	-0.0764 (0.0136)
	$\beta_{15}$	-0.0525 (0.0155)
	$\beta_{16}$	-0.0535 (0.0153)
	$\beta_{17}$	0.0029 (0.0009)
	$\beta_{18}$	0.0018 (0.0009)
Componentes de variância	$\sigma_{b_1}$	[0.2509; 0.297]
	$\sigma_{b_2}$	[0.2523; 0.309]
	$\sigma_{b_3}$	[0.0529; 0.0635]
	$\text{cor}(\sigma_{b_1}, \sigma_{b_2})$	[-0.2444; -0.0008]
	$\text{cor}(\sigma_{b_1}, \sigma_{b_3})$	[0.2018; 0.4178]
	$\text{cor}(\sigma_{b_2}, \sigma_{b_3})$	[-0.9421; -0.8992]
	$\sigma_e$	[0.3654; 0.4269]
	$\rho$	[0.1941; 0.2042]

A Tabela 4.4 apresenta os valores de AIC, BIC, log-verossimilhança (L), a estatística do teste da razão de verossimilhanças (TRV) e o correspondente valor de p para os diferentes modelos quando consideramos a presença ou ausência dos efeitos aleatórios. O modelo saturado foi o que apresentou menor valor de AIC e BIC e todos os testes de razão de verossimilhanças foram rejeitados indicando que os efeitos aleatórios devem permanecer no modelo.

**Tabela 4.4:** Valores das estatísticas de ajuste para avaliação da inclusão dos efeitos aleatórios

Modelo	Efeitos Aleatórios			AIC	BIC	L	TRV	p
	$b_{1i}$	$b_{2i}$	$b_{3i}$					
modelo 1 (4.9)	x	x	x	-583.9	-435.3	318.0		
modelo 2	x	x		-520.9	-389.4	283.5	69.0	<0.0001
modelo 3	x		x	-535.5	-404.0	290.7	54.5	<0.0001
modelo 4		x	x	-459.1	-327.6	252.6	130.8	<0.0001
modelo 5	x			-503.2	-383.1	272.6	90.8	<0.0001
modelo 6		x		-443.7	-323.6	242.8	150.3	<0.0001
modelo 7			x	-455.0	-334.9	248.5	139.0	<0.0001

A Tabela 4.5 apresenta as estatísticas de ajuste para a avaliação da inclusão dos efeitos de interação. Um dos efeitos de interação associados a taxa de declínio pós pico de produção é significativo, indicando que o efeito da estação do ano na taxa de declínio pós pico de produção depende da paridade.

**Tabela 4.5:** Valores das estatísticas de ajuste para avaliação da inclusão dos efeitos de interação

Modelo	Hipótese	AIC	BIC	L	F	p
modelo 1 (4.9)		-583.9	-435.3	318.0		
modelo 8	$\beta_5 = \beta_6 = 0$	-586.4	-449.2	317.2	0.75	0.4723
modelo 9	$\beta_{11} = \beta_{12} = 0$	-583.9	-446.6	315.9	2.03	0.1319
modelo 10	$\beta_{17} = \beta_{18} = 0$	-575.9	-438.6	311.9	6.00	0.0025
modelo 11	$\beta_5 = \beta_6 = \beta_{11} = \beta_{12} = 0$	-587.4	-461.6	315.7	-	0.3387

A Tabela 4.6 apresenta as estatísticas de ajuste para a avaliação da inclusão dos efeitos da estação do ano e da paridade. Apenas o efeito da paridade associado à taxa de aumento pré pico de produção não foi significativo. Devido à presença dos efeitos de interação associados à taxa de decréscimo pós pico de produção os coeficientes associados à estação do ano e à paridade permanecerão no modelo.

**Tabela 4.6:** Valores das estatísticas de ajuste para avaliação da inclusão dos efeitos da estação do ano e da paridade

Modelo	Hipótese	AIC	BIC	L	F	p
modelo 11		-587.4	-461.6	315.7		
modelo 12	$\beta_2 = 0$	-577.1	-457.0	309.6	12.8	<0.0001
modelo 13	$\beta_3 = \beta_4 = 0$	-560.7	-446.3	300.4	16.9	<0.0001
modelo 14	$\beta_7 = 0$	-579.5	-459.4	310.8	10.1	0.0015
modelo final	$\beta_8 = \beta_9 = 0$	-587.9	-473.6	314.0	1.7	0.1753

A Tabela (4.7) apresenta as estimativas para o modelo final. Concluimos que:

1. A produção inicial esperada de vacas que têm o primeiro parto no outono foi estimada em 14.57 (EP=0.07) litros.

2. A produção inicial esperada ( $exp(\hat{\beta}_1)$ ) é multiplicada por 1.20 (EP=0.06) quando o primeiro parto ocorre na primavera, ou seja, espera-se um aumento de 20% na produção inicial de vacas que têm o primeiro parto na primavera.
3. A produção inicial esperada ( $exp(\hat{\beta}_1)$ ) é multiplicada por 1.33 (EP=0.08) quando o segundo parto ocorre no outono, ou seja, espera-se um aumento de 33% na produção inicial de vacas que têm o segundo parto no outono.
4. A produção inicial esperada ( $exp(\hat{\beta}_1)$ ) é multiplicada por 1.46 (EP=0.08) quando o terceiro parto ou superior ocorre no outono, ou seja, espera-se um aumento de 46% na produção inicial de vacas que têm o terceiro parto ou superior no outono.
5. A taxa de aumento pré pico de produção esperada de vacas que têm o primeiro parto no outono foi estimada em 0.29 (EP=0.04).
6. O acréscimo esperado na taxa de aumento pré pico de produção quando o primeiro parto ocorre na primavera foi estimado em 0.17 (EP=0.06).
7. A taxa de declínio pós pico de produção esperada quando o primeiro parto ocorre no outono foi estimada em -0.06 (EP=0.009).
8. O decréscimo esperado na taxa de declínio pós pico de produção quando o primeiro parto ocorre na primavera foi estimado em -0.07 (EP=0.013).
9. O decréscimo esperado na taxa de declínio pós pico de produção quando o segundo parto ocorre no outono foi estimado em -0.03 (EP=0.006).
10. O decréscimo esperado na taxa de declínio pós pico de produção quando o terceiro parto ou superior ocorre no outono foi estimado em -0.03 (EP=0.006).
11. O acréscimo esperado na taxa de declínio pós pico de produção quando o segundo parto ocorre na primavera foi estimado em 0.0016 (EP=0.0005).
12. O acréscimo esperado na taxa de declínio pós pico de produção quando o terceiro parto ou superior ocorre na primavera foi estimado em 0.0009 (EP=0.0005).

**Tabela 4.7:** *Estimativas (Erro Padrão) para os parâmetros do modelo final*

	Parâmetro	Estimativa (Erro Padrão)
Produção inicial	$exp(\widehat{\beta}_1)$	14.5678 (0.7045)
	$exp(\widehat{\beta}_2)$	1.1982 (0.0606)
	$exp(\widehat{\beta}_3)$	1.3315 (0.0767)
	$exp(\widehat{\beta}_4)$	1.4574 (0.0835)
Taxa de acréscimo pré pico produção	$\beta_7$	0.2948 (0.0389)
	$\beta_8$	0.1702 (0.0575)
Taxa de decréscimo pós pico de produção	$\beta_{13}$	-0.0569 (0.0086)
	$\beta_{14}$	-0.0682 (0.0132)
	$\beta_{15}$	-0.0315 (0.0064)
	$\beta_{16}$	-0.0265 (0.0062)
	$\beta_{17}$	0.0016 (0.0005)
	$\beta_{18}$	0.0009 (0.0005)
Componentes de variância	$\sigma_{b_1}$	[0.2525; 0.2987]
	$\sigma_{b_2}$	[0.2571; 0.3137]
	$\sigma_{b_3}$	[0.0537; 0.0644]
	$cor(\sigma_{b_1}, \sigma_{b_2})$	[-0.2541; -0.0174]
	$cor(\sigma_{b_1}, \sigma_{b_3})$	[0.2100; 0.4208]
	$cor(\sigma_{b_2}, \sigma_{b_3})$	[-0.9434; -0.9017]
	$\sigma_e$	[0.1945; 0.2047]
	$\rho$	[0.3668; 0.4284]

As Figuras 4.2, 4.3 e 4.4 apresentam os gráficos referentes à análise de resíduos do modelo linearizado final. A suposição de normalidade para os efeitos aleatórios parece não ser adequada (Figura 4.3(c)).

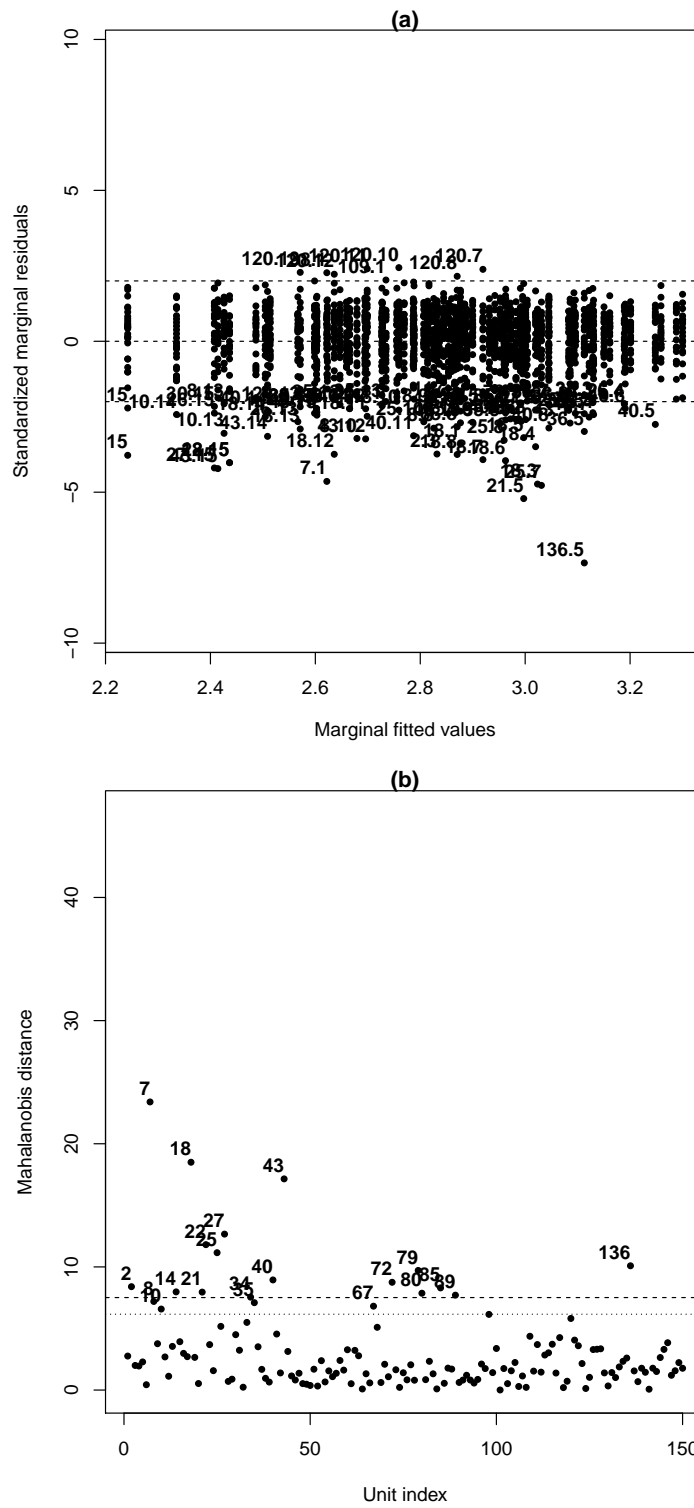


Figura 4.2: Gráficos referentes à análise de resíduos do modelo linearizado final

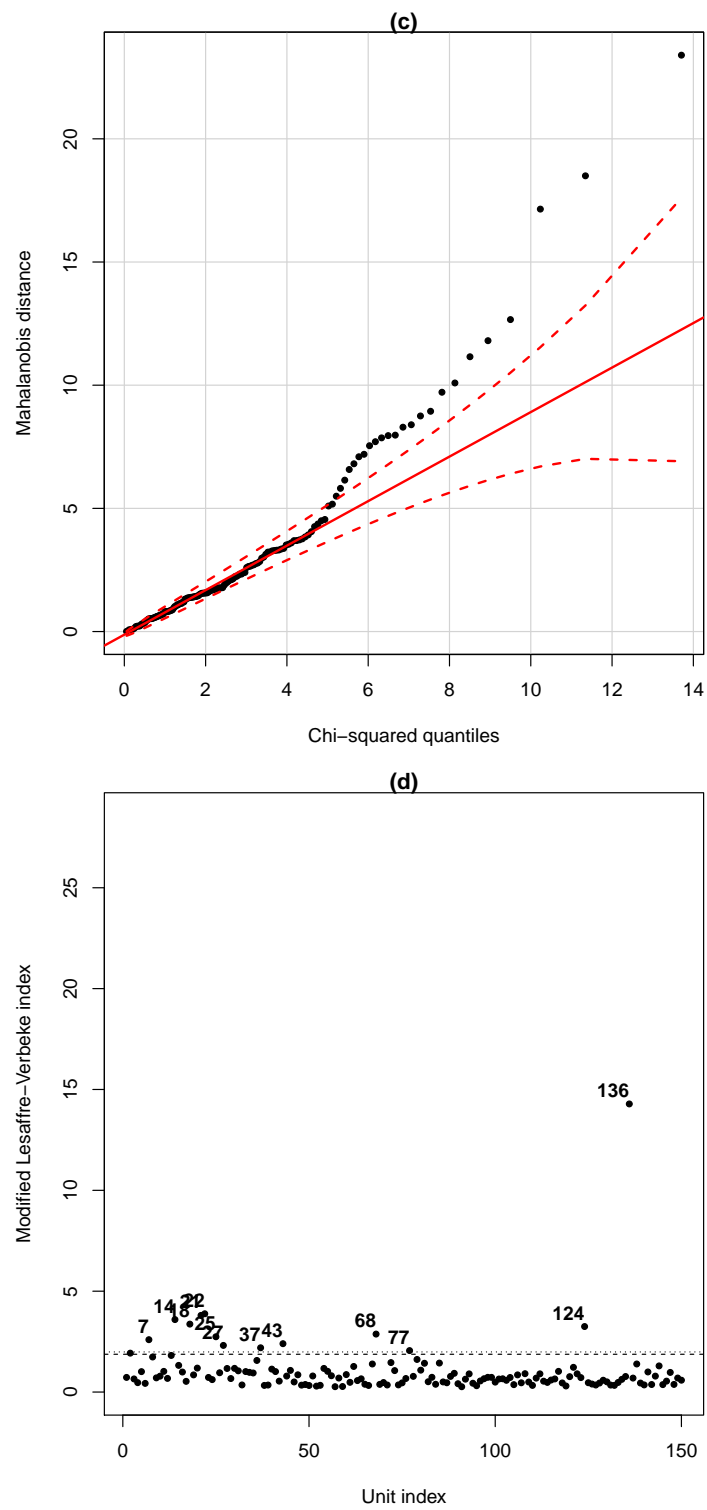


Figura 4.3: Gráficos referentes à análise de resíduos do modelo linearizado final

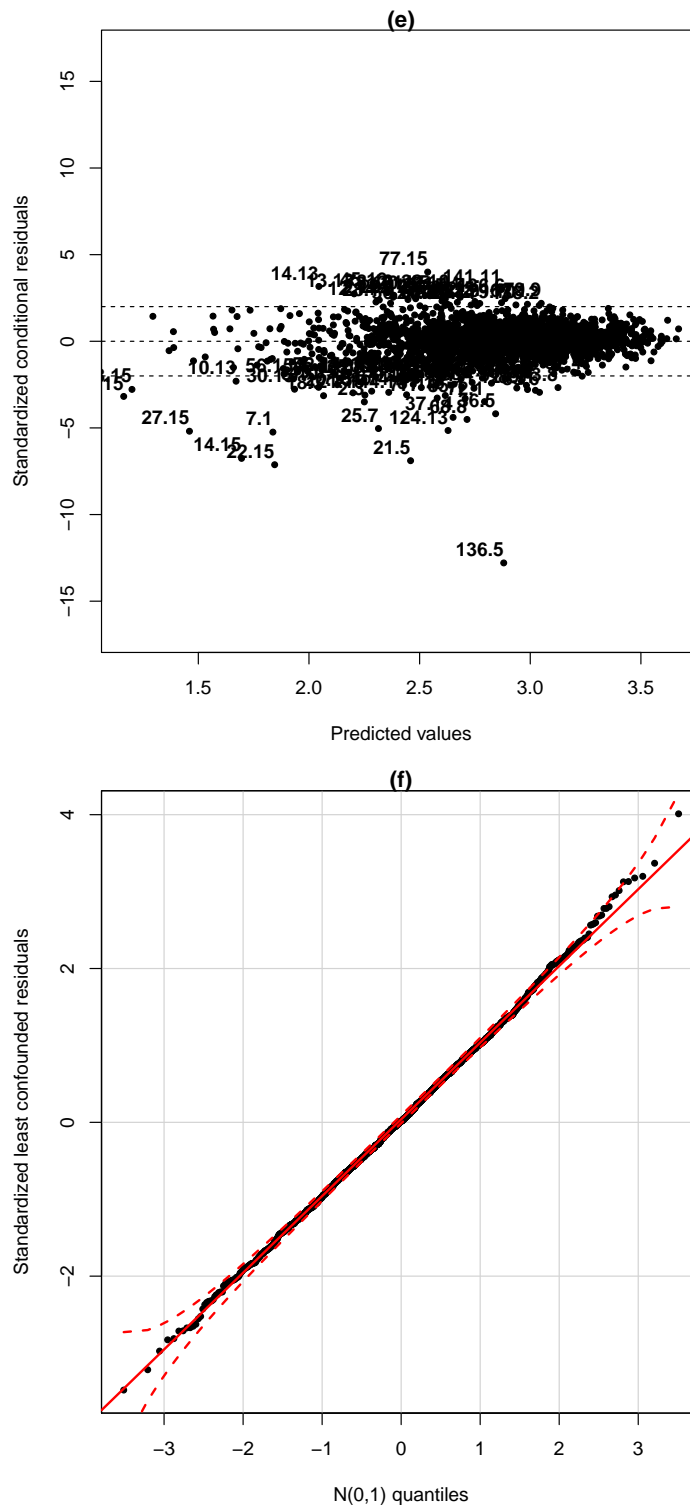


Figura 4.4: Gráficos referentes à análise de resíduos do modelo linearizado final





# Capítulo 5

## Conclusões e Perspectivas

Neste trabalho utilizamos a abordagem bayesiana para analisar um conjunto de dados para os quais a suposição de normalidade dos efeitos aleatórios não estava adequada. Utilizando tal abordagem foi possível supor outras distribuições para os efeitos aleatórios e ajustar modelos mais adequados aos dados. No entanto, existem alguns aspectos que precisam ser explorados, tais como:

- Desenvolver metodologias para análise de resíduos em modelos lineares mistos na abordagem bayesiana.
- Desenvolver um algoritmo para o ajuste de modelos com mais de dois efeitos aleatórios. Atualmente os softwares como JAGS e WinBUGS possuem essa limitação.



# Apêndice A

## Convergência e Distribuições a Posteriori

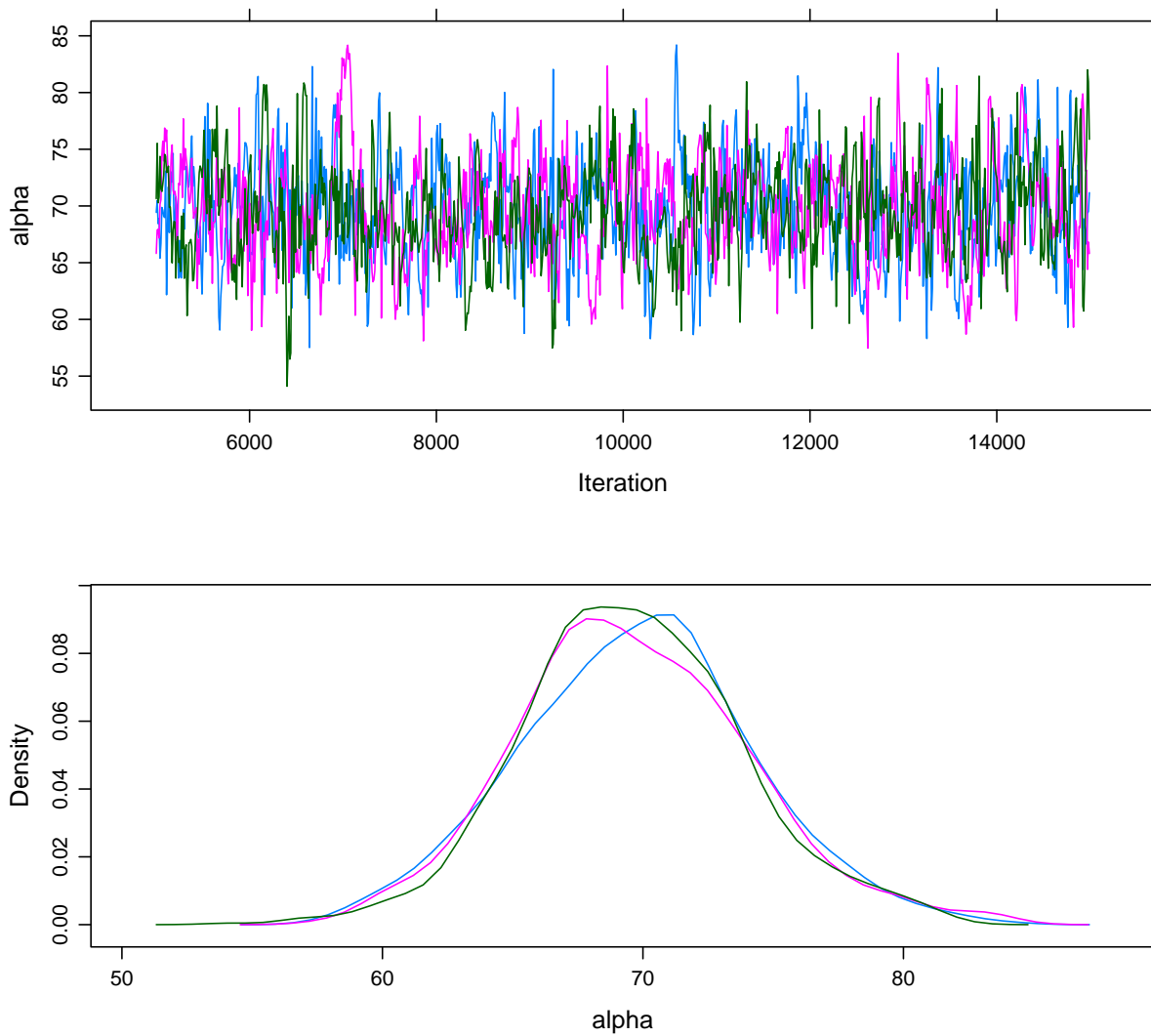
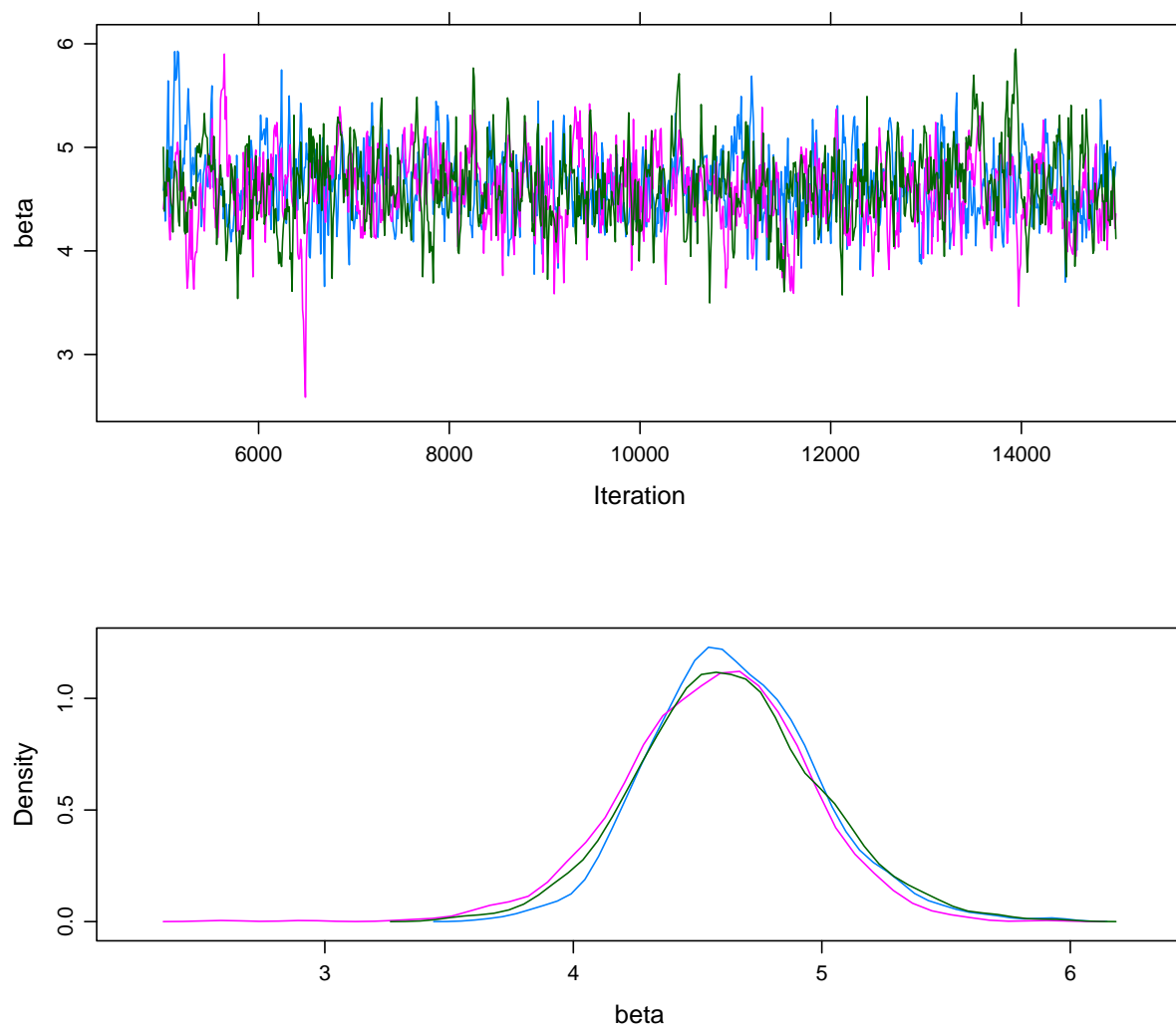


Figura A.1: Traceplot e Distribuição a Posteriori de  $\alpha$  no modelo 4.1 - abordagem bayesiana



**Figura A.2:** Traceplot e Distribuição a Posteriori de  $\beta$  no modelo 4.1 - abordagem bayesiana

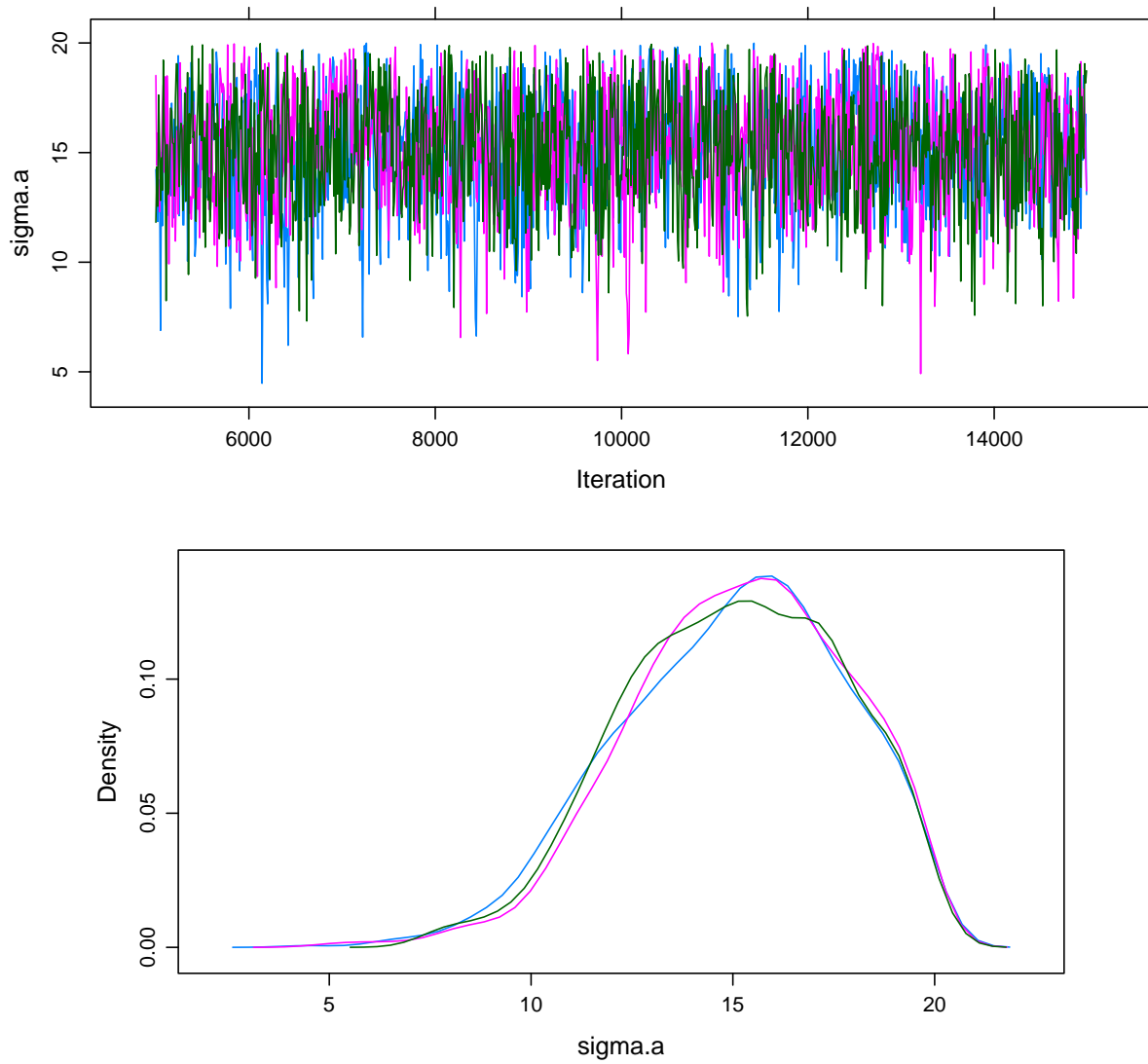


Figura A.3: Traceplot e Distribuição a Posteriori de  $\sigma_a$  no modelo 4.1 - abordagem bayesiana

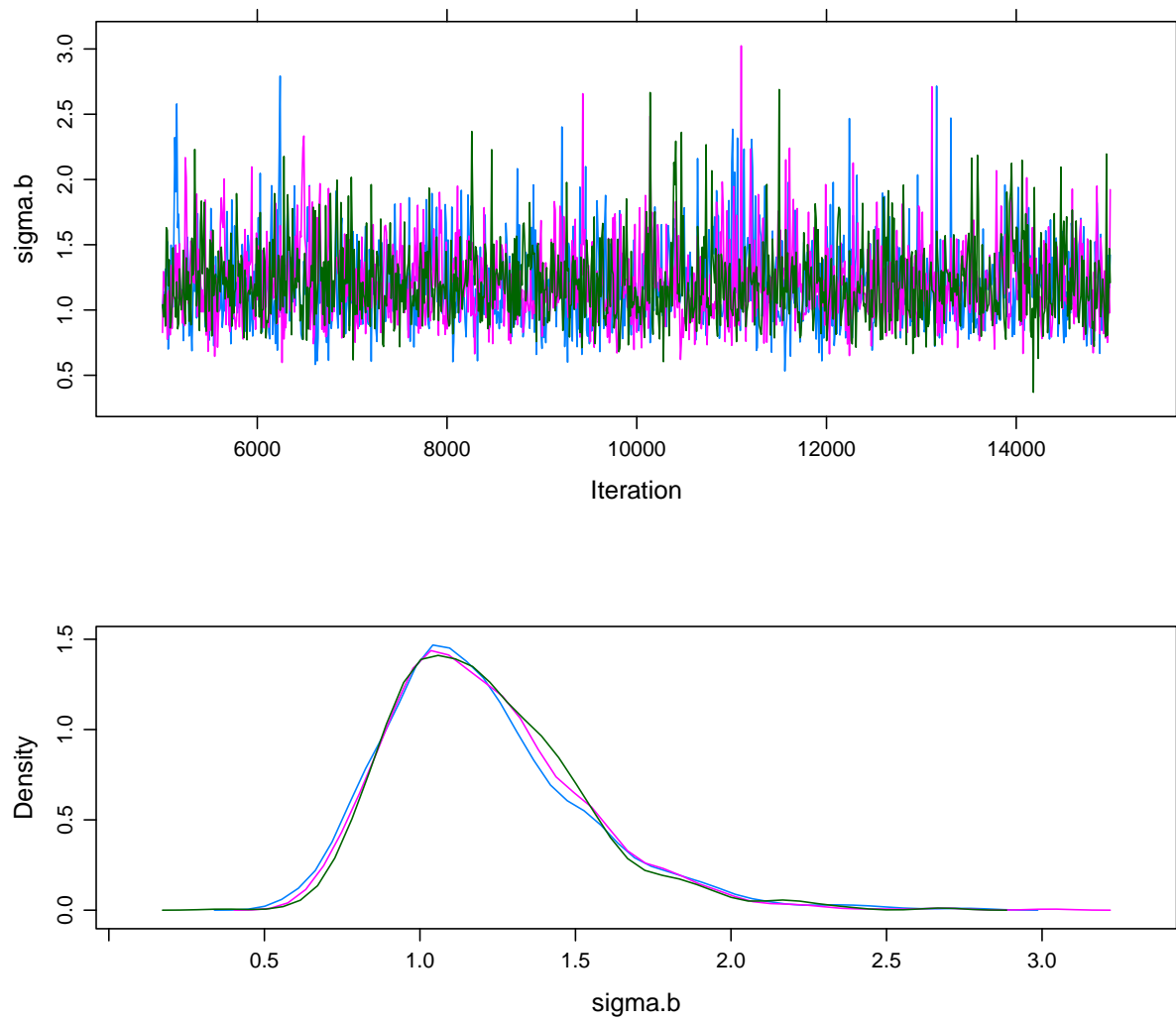


Figura A.4: Traceplot e Distribuição a Posteriori de  $\sigma_b$  no modelo 4.1 - abordagem bayesiana

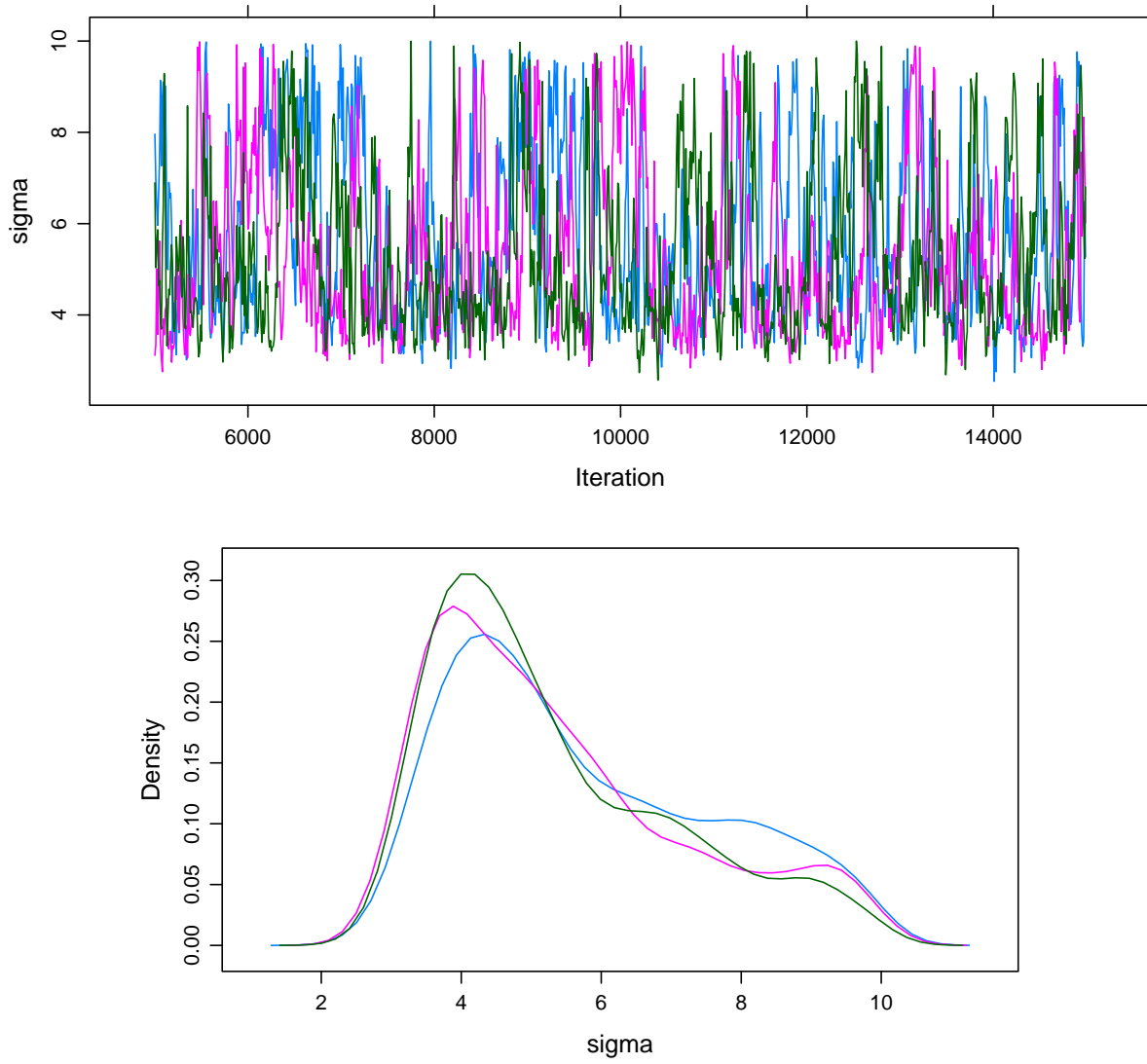
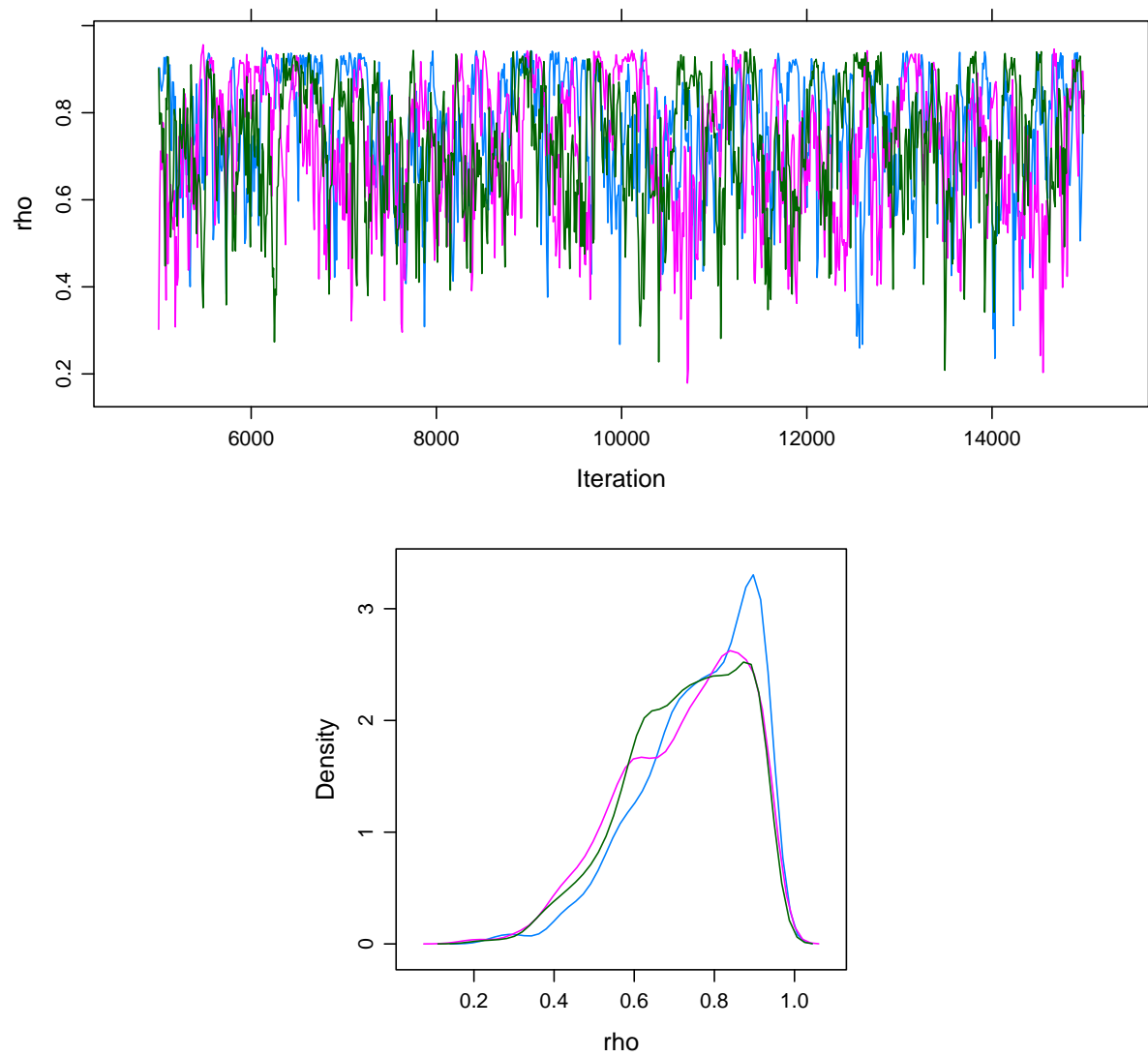


Figura A.5: Traceplot e Distribuição a Posteriori de  $\sigma_e$  no modelo 4.1 - abordagem bayesiana



**Figura A.6:** Traceplot e Distribuição a Posteriori de  $\rho$  no modelo 4.1 - abordagem bayesiana



# Referências Bibliográficas

- Azzalini, A. and Capitanio, A. (1999). Statistical Applications of the Multivariate Skew Normal Distributions. *Journal of Royal Statistics Society* **61**, 579–602.
- Breslow, N. E. e Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88**, 9–25. 4
- Congdon, P. D. (2010). *Applied Bayesian hierarchical methods*. CRC Press. 12
- Daniels, M. J. and Hogan, J. W. (2008). *Missing data in longitudinal studies: Strategies for Bayesian modeling and sensitivity analysis*. CRC Press. 7
- Davidian, M. and Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data*. Vol. 62. CRC Press. 3, 10
- Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. Springer. 6
- DeGroot, M. H. and Schervish, M. J. (2012). *Probability and Statistics*. fourth edn. Addison-Wesley. 9
- Demidenko, E. (2013). *Mixed models: theory and applications with R*. John Wiley & Sons. 7
- Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)* pp. 1–38. 6
- Denwood, M. J. (In Review). runjags: An R package providing interface utilities, parallel computing methods and additional distributions for MCMC models in JAGS. *Journal of Statistical Software*. 16
- Dey, D. K., Ghosh, S. K. and Mallick, B. K. (2000). *Generalized linear models: A Bayesian perspective*. CRC Press. 7
- Diggle, P. J., Heagerty, P. J., Liang, K.-Y. and Zeger, S. L. (2002). Analysis of Longitudinal Data. *Oxford Statistical Science Series*. 6, 7
- Fearn, T. (1975). A Bayesian approach to growth curves. *Biometrika* **62**, 89–100. 11
- Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association* **85**, 398–409. 11
- Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science* pp. 457–472. 16

- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* pp. 721–741. [11](#)
- Gilks, W., Richardson, S. and Spiegelhalter, D. (1996). Markov Chain Monte Carlo in Practice, ser. Interdisciplinary Statistics Series . [11](#)
- Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**(358), 320–338. [6](#)
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109. [11](#)
- Henderson, C. R. (1984). *Applications of linear models in animal breeding*. terceira edn. University of Guelph Press. [3](#)
- Jiang, J. et al. (1996). REML estimation: Asymptotic behavior and related topics. *The Annals of Statistics* **24**(1), 255–286. [7](#)
- Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* pp. 963–974. [3](#), [5](#)
- Lee, Y. e Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society B* **58**, 619–678. [4](#)
- Lee, Y. e Nelder, J. A. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika* **88**, 987–1006. [4](#)
- Lindley, D. V. and Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 1–41. [11](#)
- Lindstrom, M. J. and Bates, D. M. (1988). Newton—Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* **83**(404), 1014–1022. [6](#)
- Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and computing* **10**, 325–337. [12](#)
- McCulloch, C. E. e Searle, S. R. (2001). *Linear and Generalized Linear Mixed Models*. first edn. Wiley. [4](#)
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics* **21**, 1087–1092. [11](#)
- Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *The Annals of Statistics* pp. 746–762. [7](#)
- Nobre, J. S. and Singer, J. M. (2007). Residual Analysis for Linear Mixed Models. *Biometrical Journal* **38**, 1063–1072. [8](#), [15](#)

- Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58**(3), 545–554. 6
- Paulino, C. D. M., Turkman, M. A. A. and Murteira, B. (2003). *Estatística bayesiana*. Fundação Calouste Gulbenkian. 9
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team (2015). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-119. 15, 20
- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Springer. 3, 7
- Plummer, M. (2011). JAGS Version 3.1.0 user manual. *International Agency for Research on Cancer*. 12, 16
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. 12, 15, 16
- Seltzer, M. H. (1993). Sensitivity analysis for fixed effects in the hierarchical model: A Gibbs sampling approach. *Journal of Educational and Behavioral Statistics* **18**, 207–235. 12
- Seltzer, M. H., Wong, W. H. and Bryk, A. S. (1996). Bayesian analysis in applications of hierarchical models: Issues and methods. *Journal of Educational and Behavioral Statistics* **21**, 131–167. 12
- Sen, P. K. and Singer, J. M. (1994). *Large sample methods in statistics: An introduction with applications*. Vol. 25. CRC Press. 20
- Singer, J. M. and Andrade, D. F. (2000). Analysis of longitudinal data. *Handbook of statistics* **18**, 115–160. 6
- Sorensen, D. e Gianola, D. (2002). *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. second edn. Springer. 12
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P. and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 583–639. 12
- Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). WinBUGS user manual . 12
- Stan Development Team (2014). *Stan Modeling Language Users Guide and Reference Manual, Version 2.5.0*. 12
- Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* pp. 1171–1177. 7
- Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika* **92**, 351–370. 8, 12
- Verbeke, G. and Lesaffre, E. (1987). The effect of misspecifying the random effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* **23**, 541–556. 3

Verbeke, G. and Lesaffre, E. (1996). Large sample properties of the maximum likelihood estimators in linear mixed models with misspecified random-effects distributions. Technical report. Technical Report 1996.1, Catholic University of Leuven, Biostatistical Centre for Clinical Trials, Leuven. 7

Vonesh, E. and Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements*. CRC press. 3, 6

Wood, P. D. P. (1967). Algebraic model of the lactation curve in cattle. *Nature* **216**, 164–165.  
17