

UNIVERSIDADE ESTADUAL PAULISTA
Faculdade de Ciências e Tecnologia de Presidente Prudente
Programa de Pós-Graduação em Matemática Aplicada e Computacional

**Análise Bayesiana de Dados Composicionais na
Presença de Covariáveis**

Taciana Kisaki Oliveira Shimizu
Orientador: Prof. Dr. Jorge Alberto Achcar
Coorientador: Prof. Dr. Mário Hissamitsu Tarumoto

Presidente Prudente, Fevereiro de 2014

UNIVERSIDADE ESTADUAL PAULISTA
Faculdade de Ciências e Tecnologia de Presidente Prudente
Programa de Pós-Graduação em Matemática Aplicada e Computacional

**Análise Bayesiana de Dados Composicionais na
Presença de Covariáveis**

Taciana Kisaki Oliveira Shimizu
Orientador: Prof. Dr. Jorge Alberto Achcar
Coorientador: Prof. Dr. Mário Hissamitsu Tarumoto

Dissertação apresentada ao Programa de
Pós-Graduação em Matemática Aplicada e
Computacional da Faculdade de Ciências
e Tecnologia da UNESP para obtenção do
título de Mestre em Matemática Aplicada e
Computacional.

Presidente Prudente, Fevereiro de 2014

FICHA CATALOGRÁFICA

S559a Shimizu, Taciana Kisaki Oliveira.
Análise Bayesiana de dados composicionais na presença de covariáveis /
Taciana Kisaki Oliveira Shimizu. - Presidente Prudente : [s.n], 2014
111 f.

Orientador: Jorge Alberto Achcar
Coorientador: Mário Hissamitsu Tarumoto
Dissertação (mestrado) - Universidade Estadual Paulista, Faculdade de
Ciências e Tecnologia
Inclui bibliografia

1. Dados Composicionais. 2. Inferência Bayesiana. 3. MCMC. I. Achcar,
Jorge Alberto. II. Tarumoto, Mário Hissamitsu. III. Universidade Estadual
Paulista. Faculdade de Ciências e Tecnologia. IV. Análise Bayesiana de
dados composicionais na presença de covariáveis.

BANCA EXAMINADORA



PROF. DR. JORGE ALBERTO ACHCAR
ORIENTADOR



PROF. DR. CARLOS APARECIDO DOS SANTOS
UEM



PROF. DRA. APARECIDA DONISETI PIRES DE SOUZA
UNESP/FCT



TACIANA KISAKI OLIVEIRA SHIMIZU

Presidente Prudente (SP), 24 de fevereiro de 2014.

RESULTADO: A PROVA DO

*Aos meus pais, Rosa e Gilmar e
ao meu marido Marcelo (Hiro)
pelo amor, paciência,
incentivo e companheirismo.*

Agradecimentos

Agradeço a todos que diretamente ou indiretamente contribuíram para a realização desse trabalho, de forma especial:

À Deus por estar sempre presente na minha vida, proporcionando proteção, sabedoria em todos os momentos.

Aos meus pais Gilmar e Rosa, com amor e carinho dedicaram-se na minha educação, aos meus irmãos Emanuelle e Renan, pelo amor fraternal, apoio e alegrias que passamos juntos.

Ao meu marido Marcelo (Hiro) pelo amor, companheirismo, respeito, paciência e por sempre me apoiar à conquistar meus objetivos pessoais e profissionais. Obrigada!

À toda minha família, inclusive à família do meu marido pela ajuda em todos os momentos que necessitei.

Ao meu tio Paulo Shigueru Kísaki, pelo incentivo nos meus estudos desde criança.

Ao meu orientador Prof. Dr. Jorge Alberto Achcar pela orientação, dedicação, paciência e apoio oferecidos para a elaboração desse trabalho, e sobretudo pela oportunidade de aprender e trabalhar juntamente com um exemplo de profissional.

Ao meu coorientador Prof. Dr. Mário Hissamitsu Tarumoto e a Profa. Olga Lyda Anglas Rosales Tarumoto, por tudo que me ensinaram, pela amizade desde do período de graduação, conselhos, orientação, incentivo em todas as atividades que realizei após a graduação.

À FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo), pelo apoio financeiro oferecido, possibilitando a minha dedicação exclusiva ao desenvolvimento desse trabalho.

Aos professores que contribuíram com a minha formação acadêmica e também me auxiliaram no mestrado, em especial, Profa. Aparecida Doniseti Pires de Souza, Profa. Vilma Mayumi Tachibana, Prof. Messias Meneguette Júnior, Prof. Josmar Mazucheli, Profa. Vanessa Avansini Botta Pirani.

Aos professores da banca de qualificação, Dr. Josmar Mazucheli e Dra. Vilma Mayumi Tachibana pelas contribuições relevantes para o trabalho.

Aos professores Dra. Renata Maria Coimbra Libório e Dr. Everaldo Santos Melazzo que colaboraram com o meu crescimento científico, por terem dado a oportunidade de aplicar a Estatística em projetos de pesquisas acadêmicas, conviver e aprender com profissionais de outras áreas.

À amiga Elizabeth Mie Hashimoto, pela sua amizade desde da graduação e sua disposição em me ajudar no que fosse preciso.

Aos colegas do curso de mestrado, Ana, Débora, Leandro, Pedro e Renato, pela amizade e experiências trocadas.

À FCT/UNESP, aos funcionários da Seção de Pós Graduação: Cinthia, André e Ivonete, em especial à Aparecida Tamae Otsuka pela ajuda fundamental no processo de pedido de bolsa e na prestação de contas à FAPESP.

SHIMIZU, T. K. O. **Análise Bayesiana de Dados Composicionais na Presença de Covariáveis**. Dissertação de mestrado - Faculdade de Ciências e Tecnologia - FCT, UNESP, Presidente Prudente - SP, 2014.

Resumo

Dados composicionais consistem em vetores conhecidos como composições cujos componentes são positivos e definidos no intervalo $(0,1)$ representando proporções ou frações de um “todo”. A soma desses componentes deve ser igual a um. Os dados composicionais estão presentes em diferentes áreas, como na geologia, ecologia, economia, medicina entre muitas outras. Desta forma há um grande interesse em novas abordagens de modelar dados composicionais. Neste estudo, introduzimos as transformações logaritmo da razão (*alr*) e Box-Cox em modelos usados para dados composicionais, assumindo erros normais não correlacionados. O objetivo principal deste trabalho é aplicar métodos Bayesianos para estes modelos utilizando os métodos padrões de Monte Carlo via Cadeias de Markov (MCMC) para simular amostras da *posteriori* conjunta de interesse. Nós aplicamos a metodologia proposta em dois conjuntos de dados, sendo que um deles é sobre um experimento de medidas repetidas na qual introduzimos uma variável de efeito aleatório para capturar a dependência para os dados longitudinais e, além disso, a introdução de dois efeitos aleatórios extras no modelo. Estes resultados de modelagem podem ser de grande interesse em trabalhos aplicados que lidam com conjuntos de dados composicionais.

Palavras-chave: Dados Composicionais, Inferência Bayesiana, MCMC.

SHIMIZU, T. K. O. **Bayesian Analysis of Compositional Data in Presence of Covariates.** Dissertação de mestrado - Faculdade de Ciências e Tecnologia - FCT, UNESP, Presidente Prudente - SP, 2014.

Abstract

Compositional data consist of known compositions vectors whose components are positive and defined in the interval $(0,1)$ representing proportions or fractions of a “whole”. The sum of these components must be equal to one. Compositional data is present in different areas, as in ecology, economy, medicine among many others. In this way, there is a great interest in new modeling approaches for compositional data. In this study we introduced additive log-ratio (alr) and Box-Cox transformations models used for compositional data, under uncorrelated normal errors. The main objective of this project is to apply Bayesian methods to these models using standard Markov Chain Monte Carlo (MCMC) methods to simulate samples of the joint posterior of interest. We apply the proposed methodology in two data sets, whereas one of them is about an experiment of repeated measures where we introduced a random effect variable to capture the dependence for the longitudinal data and also the introduction of two extra random effects in the model. These modeling results could be of great interest in the applied work dealing with compositional data sets.

Keywords: Compositional data, Bayesian Inference, MCMC.

Lista de Figuras

4.1	Box plots para os componentes dos pontos dos times da Superliga de vôlei.	16
4.2	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros de intercepto do modelo 1.	23
4.3	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros de inclinação do modelo 1.	24
4.4	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os desvios padrões do modelo 1.	25
4.5	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros de intercepto do modelo 2.	34
4.6	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros de inclinação do modelo 2.	35
4.7	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os desvios padrões do modelo 2.	36
4.8	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros λ do modelo 2.	37
4.9	Gráficos dos valores observados e ajustados dos componentes do vôlei de acordo com o modelo	40
5.1	Box plots dos componentes do leite para os grupos antes e depois da dieta.	43
5.2	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros β_0 's no modelo 3.	51
5.3	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros β_1 's no modelo 3.	52
5.4	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros de variância no modelo 3.	53

5.5	Densidades <i>a posteriori</i> e traços <i>a posteriori</i> para os parâmetros de variância no modelo 3.	54
5.6	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros β_0 's no modelo 4.	63
5.7	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros β_1 's no modelo 4.	64
5.8	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros de variância no modelo 4.	65
5.9	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros de variância no modelo 4.	66
5.10	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros λ 's no modelo 4.	67
5.11	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros β_0 's no modelo 5.	75
5.12	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros β_1 's no modelo 5.	76
5.13	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros de variância no modelo 5.	77
5.14	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros de variância no modelo 5.	78
5.15	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros β_0 's no modelo 6.	86
5.16	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros β_1 's no modelo 6.	87
5.17	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros λ 's no modelo 6.	88
5.18	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros de variância no modelo 6.	89
5.19	Densidades <i>a posteriori</i> via amostrador de Gibbs e traços <i>a posteriori</i> para os parâmetros de variância no modelo 6.	90

5.20 Gráficos dos valores observados e ajustados dos componentes do leite de acordo com o modelo	95
---	----

Lista de Tabelas

1.1	Transformações logísticas elementares de S^g para R^g	3
4.1	Resumos <i>a posteriori</i> e Estatística de Teste para o Diagnóstico de Geweke - Modelo 1.	22
4.2	Estimativas das proporções dos componentes - Modelo 1.	22
4.3	Resumos <i>a posteriori</i> e Estatística de Teste para o Diagnóstico de Geweke - Modelo 2.	32
4.4	Estimativas das proporções dos componentes - Modelo 2.	33
4.5	Estimativas dos parâmetros do modelo 1 - Método Clássico	38
4.6	Estimativas dos parâmetros do modelo de Regressão Dirichlet	39
4.7	Critério DIC - Modelos 1 e 2.	39
4.8	Soma do quadrado das diferenças entre os valores observados e os valores ajustados.	39
5.1	Resumos <i>a posteriori</i> e Estatística de Teste para o Diagnóstico de Geweke - Modelo 3.	50
5.2	Estimativas das proporções dos componentes - Modelo 3.	55
5.3	Resumos <i>a posteriori</i> e Estatística de Teste para o Diagnóstico de Geweke - Modelo 4.	62
5.4	Estimativas das proporções dos componentes - Modelo 4.	68
5.5	Resumos <i>a posteriori</i> e Estatística de Teste para o Diagnóstico de Geweke - Modelo 5.	74
5.6	Estimativas das proporções dos componentes - Modelo 5.	79
5.7	Resumos <i>a posteriori</i> e Estatística de Teste para o Diagnóstico de Geweke - Modelo 6.	85
5.8	Estimativas das proporções dos componentes - Modelo 6.	91

5.9	Estimativas dos parâmetros do modelo 3 - Método Clássico	92
5.10	Estimativas dos parâmetros do modelo de Regressão Dirichlet	93
5.11	Critério DIC - Modelos 3, 4, 5 e 6.	93
5.12	Soma do quadrado das diferenças entre os valores observados e os valores ajustados.	94
A.1	Conjunto de dados referente aos jogos da Superliga de vôlei masculina 2011/2012.	102
B.1	Composições alimentares do leite de 30 vacas (pr=proteína, mf=gordura do leite, ch=carboidrato, Ca=cálcio, Na=sódio, K=potássio) antes e depois de uma nova dieta.	103

Sumário

1	Introdução	1
2	Introdução aos Métodos Bayesianos	6
2.1	Teorema de Bayes	6
2.2	Distribuições a <i>Priori</i> e a <i>Posteriori</i>	7
2.3	Diagnóstico de Convergência	8
2.3.1	Diagnóstico de Geweke	8
2.4	Estimação	9
2.4.1	Estimação Pontual	9
2.4.2	Estimação por Intervalos	10
2.5	Critério de Seleção de Modelos	10
2.5.1	Critério DIC	11
3	Transformações e Modelo de Regressão Dirichlet para Dados Composicionais	12
3.1	Transformações Logaritmo da Razão (<i>alr</i>) e Box-Cox	12
3.2	Modelo de Regressão Dirichlet	13
4	Análise Bayesiana para os Dados da Superliga de Vôlei Masculina	15
4.1	Transformação <i>alr</i> - Modelo 1	16
4.1.1	Análise Bayesiana - Modelo 1	18
4.2	Transformação Box-Cox - Modelo 2	25
4.2.1	Análise Bayesiana - Modelo 2	28

4.3	Análise clássica	37
4.4	Discussão dos Resultados	39
5	Análise Bayesiana para Dados Longitudinais	42
5.1	Transformação <i>alr</i> Considerando um Efeito Aleatório - Modelo 3	43
5.1.1	Análise Bayesiana - Modelo 3	46
5.2	Transformação Box-Cox Considerando um Efeito Aleatório - Modelo 4	55
5.2.1	Análise Bayesiana - Modelo 4	58
5.3	Transformação <i>alr</i> Considerando Três Efeitos Aleatórios - Modelo 5	69
5.3.1	Análise Bayesiana - Modelo 5	70
5.4	Transformação Box-Cox Considerando Três Efeitos Aleatórios - Modelo 6	79
5.4.1	Análise Bayesiana - Modelo 6	80
5.5	Análise Clássica	91
5.6	Discussão dos Resultados	93
6	Considerações Finais	96
	Referências	98
	Apêndice A – Conjunto de Dados da Superliga de Vôlei Masculina	102
	Apêndice B – Conjunto de Dados Longitudinais	103
	Apêndice C – Programas	104
C.1	OPENBUGS	104
C.1.1	Programa - Modelo 1	104
C.1.2	Programa - Modelo 2	105
C.1.3	Programa - Modelo 3	106
C.1.4	Programa - Modelo 4	108
C.2	SAS - <i>Proc NLMIXED</i>	110

C.2.1	Programa - Dados Superliga de Vôlei	110
C.2.2	Programa - Dados Longitudinais (qualidade do leite)	110
C.3	R - Modelo de Regressão Dirichlet	111

1 Introdução

O estudo adequado da teoria de dados composicionais desenvolveu-se na década de 80 após os trabalhos de Aitchison e Shen [7] e Aitchison [2], que contribuíram com alguns dos princípios necessários para a análise de dados composicionais e com a natureza especial do seu espaço amostral.

Desde então, a metodologia de dados composicionais vem sendo aplicada em estudos de diversas áreas do conhecimento, com ênfase na geologia, ciências do solo, ciências sociais, medicina, genética, entre outras.

Segundo Aitchison [3], os dados composicionais são definidos como um vetor $\mathbf{x}$ com elementos positivos $x_1, \dots, x_G$ que representam proporções de um “todo”, tais que a soma de todos os elementos do vetor é igual a um, ou seja, $x_1 + \dots + x_G = 1$, em que G representa o número total de componentes dos dados composicionais.

Os métodos multivariados usuais raramente são adequados para a análise de dados composicionais e há uma carência relativa de técnicas alternativas adequadas para o mesmo. Além disso, a presença de covariáveis acrescenta a complexidade da situação, de acordo com Iyengar e Dey [28].

O espaço amostral denominado Simplex desempenha um papel importante quando se requer interpretações para esses dados. A definição do espaço amostral dos dados composicionais no Simplex, denotado por S^{G-1} é dado por,

$$S^{G-1} = \{(x_1, \dots, x_G) : x_1 > 0, \dots, x_G > 0; x_1 + \dots + x_G = 1\}.$$

Dessa forma, podemos denominar um vetor $\mathbf{x}$ no espaço Simplex como uma composição, os elementos desse vetor como componentes e o conjunto dos vetores são definidos como dados composicionais [2].

O tratamento estatístico para esse tipo de dados mostra-se complicado devido à au-

sência de conceitos de independência e de classes paramétricas no Simplex.

Com isso, várias distribuições têm sido sugeridas para a modelagem de dados composicionais, dentre elas a distribuição Dirichlet $D^g(\boldsymbol{\alpha})$ no espaço amostral S^g , em que $g = G - 1$ é definida por

$$\frac{\Gamma(\alpha_1 + \dots + \alpha_{g+1})}{\Gamma(\alpha_1) \dots \Gamma(\alpha_{g+1})} x_1^{\alpha_1-1} \dots x_g^{\alpha_g-1} (1 - x_1 - \dots - x_g)^{\alpha_{g+1}-1}$$

em que $\mathbf{x} \in S^g$, $\boldsymbol{\alpha} \in R_+^{g+1}$ e $\Gamma(\cdot)$ é a função Gama.

Porém, segundo Aitchison [3], a estrutura de correlação de uma composição Dirichlet é completamente negativa, com $\text{corr}(x_i, x_j) < 0$ para todo $i \neq j$. Assim, tornando-a inadequada para o padrão dos dados composicionais, os quais possuem algumas de suas correlações definitivamente positivas.

Nesse contexto, novas alternativas para satisfazer as restrições existentes na análise de dados composicionais é proposto em Aitchison [2] através de transformações adequadas que levam do espaço amostral restrito S^g para o espaço bem definido R^g .

A ideia sobre a indução de classes de distribuições bem estabelecidas em espaços amostrais complicados é antiga, sendo que em McAlister [35] considerou $y \in R$ com distribuição $N(\mu, \sigma^2)$ expressando a sua ideia em termos de inversa, ou seja, a transformação $y = \log(x)$ induziu a uma distribuição $\Lambda(\mu, \sigma^2)$ nos R_+ , denotando essa nova classe de distribuições lognormal.

Aitchison e Shen [7] desenvolveram a classe de distribuições Normal-logística a qual induziu de S^g para a classe de distribuições $N^g(\boldsymbol{\mu}, \Sigma)$ dos R^g através da transformação logaritmo da Razão (*alr*) e apresentaram a simplicidade da sua aplicação em vários problemas.

Na Tabela 1.1, apresentamos outras duas transformações elementares: a logística multiplicativa e a logística híbrida, além da transformação *alr* já citada.

Tabela 1.1: Transformações logísticas elementares de S^g para R^g .

Transformações	Inversas
logaritmo da razão <i>alr</i>	$y_i = \log \frac{x_i}{x_{g+1}}$
logística multiplicativa	$y_i = \log \frac{x_i}{1 - \sum_{k=1}^i x_k}$
logística híbrida	$y_1 = \log \frac{x_1}{1 - x_1};$ $y_i = \log \frac{x_i}{\left(1 - \sum_{k=1}^{i-1} x_k\right) \left(1 - \sum_{k=1}^i x_k\right)}, \quad i = 2, \dots, g$

Recentemente, alguns trabalhos sobre a aplicação da teoria de dados composicionais vêm sendo desenvolvidos. Por exemplo, podemos citar o trabalho de Aitchison e Greenacre [6] que desenvolveram técnicas de biplots composicionais sendo um avanço importante no estudo de dados composicionais, pois a sua adaptação é simples e representa uma ferramenta útil e exploratória.

Em Hijazi e Jernigan [27], considerou-se uma comparação entre os modelos de regressão Dirichlet e a transformação logaritmo da razão (*alr*) para averiguar qual deles se adequam melhor na presença de uma covariável observada, concluindo que o modelo de regressão Dirichlet é uma alternativa à transformação *alr* para o conjunto de dados analisado no artigo e, que além disso, parece controlar a variabilidade do componente de uma forma mais verdadeira. Em Gueorguieva et al. [24] aplicou-se o modelo de regressão Dirichlet em dados médicos, mais especificamente, em dados psiquiátricos.

Os métodos Bayesianos tornaram-se uma alternativa relevante para a análise de dados composicionais, considerando ainda a aplicação dos métodos de Monte Carlo via cadeias de Markov (MCMC). Sob esse enfoque, um modelo geoestatístico bivariado de dados composicionais foi proposto por Martins et al. [34]. Achcar e Obage [1] estudaram os dados composicionais utilizando métodos Bayesianos, considerando as transformações *alr* e Box-Cox, em que assumem erros correlacionados com distribuição Normal.

A transformação Box-Cox tem como objetivo contornar problemas relativos a dados com falta de normalidade, assim essa transformação foi sugerida para dados composicionais em Aitchison [3] e adotada em Rayens e Srinivasan [44, 45] como um meio de desenvolver modelos mais confiáveis, além de incorporar a transformação *alr* quando o

parâmetro $\lambda \rightarrow 0$ (transformação Box-Cox).

Aitchison e Egozcue [5] fizeram um levantamento de trabalhos realizados nessa teoria durante os últimos vinte anos e sugeriram possibilidades de pesquisas futuras, que venham a acrescentar o estudo de dados composicionais.

Dentre eles, estão os trabalhos relacionados a análise de dados composicionais com a presença de componentes iguais a zero, que representa um dos principais obstáculos enfrentados quando aplica-se a transformação *alr* e o modelo de regressão Dirichlet, de acordo com Hijazi [26], em que propôs uma nova técnica baseada no algoritmo EM para substituir os componentes de valor zero sob o modelo de regressão Dirichlet. Podemos citar outros trabalhos que abordaram a presença de componentes igual a zero, sendo eles Martín-Fernández et al. [33] e Neocleous et al. [39].

Em virtude do exposto, existe a necessidade de mais estudos acerca da teoria, principalmente em modelos de regressão para dados composicionais.

Deste modo, o objetivo principal do presente trabalho é a aplicação da transformação dos dados composicionais, sendo elas a transformação *alr* e Box-Cox, realizando uma comparação entre elas e aplicando a metodologia proposta em dois conjuntos de dados, em que vale ressaltar que a abordagem de dados composicionais na análise desses dados é inédita. Utilizamos os métodos Bayesianos considerando métodos de Monte Carlo via cadeias de Markov (MCMC), aplicando os algoritmos amostrador de Gibbs e Metropolis-Hastings nos modelos propostos. Além disso, os resultados obtidos dos modelos em que foram aplicados a transformação *alr* são comparados aos obtidos pelos métodos de inferência clássica e aos obtidos com o modelo de regressão Dirichlet clássico (conforme em Hijazi e Jernigan [27]).

O trabalho está organizado da seguinte maneira. No Capítulo 2, apresentamos uma síntese dos métodos Bayesianos, descrevendo alguns conceitos importantes como a estimação pontual e intervalar, o diagnóstico de convergência de Geweke e o critério de seleção de modelos DIC.

No Capítulo 3, apresentamos as definições das transformações *alr*, Box-Cox e o modelo de regressão Dirichlet aplicados em dados composicionais.

No Capítulo 4, aplicamos uma análise Bayesiana em modelos de regressão utilizando os métodos MCMC, mais especificamente, o amostrador de Gibbs e o algoritmo de Metropolis-Hastings, utilizando as transformações *alr* e Box-Cox para dados composicionais considerando erros não correlacionados com distribuição Normal, sendo que para

efeito de comparação, foram calculadas as estimativas dos parâmetros dos modelos propostos através da inferência clássica, e também para o modelo de regressão Dirichlet clássico. O conjunto de dados reais trata-se de 128 jogos da Superliga de Vôlei Masculina Brasileira 2011/2012.

No Capítulo 5, apresentamos uma análise Bayesiana em modelos de regressão utilizando os métodos MCMC, aplicando as transformações *alr* e Box-Cox ao conjunto de dados longitudinais referente a um experimento de tentativa de melhora na qualidade do leite de vaca, nas quais trinta vacas receberam composições de dieta diferentes antes e depois de uma dieta controlada. Inserimos um efeito aleatório no primeiro modelo proposto para capturar a dependência entre as medidas repetidas para cada indivíduo e no segundo modelo introduzimos dois efeitos aleatórios, sendo um efeito para capturar a dependência entre as medidas repetidas para o grupo 1 (antes de receber a dieta) e o segundo efeito para capturar a dependência entre as medidas repetidas para o grupo 2 (depois de receber a dieta). Da mesma forma que no Capítulo 4, apresentamos as estimativas dos parâmetros através da inferência clássica e do modelo de regressão Dirichlet clássico.

Por fim, no Capítulo 6, apresentamos as principais considerações dos resultados obtidos no trabalho e algumas sugestões para pesquisas futuras.

2 Introdução aos Métodos Bayesianos

Os métodos Bayesianos tornaram-se uma ferramenta importante em modelagem estatística e análise de dados, pois ao contrário da inferência clássica, não fazem uso da teoria assintótica [19]. Dessa forma, podemos considerar que a inferência Bayesiana é uma alternativa interessante em relação à inferência clássica. A principal diferença entre as teorias é que a inferência Bayesiana trata um parâmetro qualquer, denotado por θ como uma variável aleatória, assumindo que possui uma distribuição de probabilidade, que é caracterizada com uma distribuição *a priori*, denotada por $p(\theta)$. Esta distribuição deve representar (probabilisticamente) o conhecimento que se tem sobre o parâmetro θ antes da realização do experimento, ou seja, através de *prioris* informativas, em que se conhecem previamente os parâmetros de interesse; e de *prioris* não-informativas, onde há pouco ou nenhum conhecimento acerca dos parâmetros de interesse.

A fundamentação da teoria de inferência Bayesiana é baseada no teorema de Bayes, que associa a função de verossimilhança (informação oriunda dos dados) e a distribuição *a priori* de θ (informação prévia de θ). Logo, estas duas fontes de informações são combinadas resultando na distribuição denominada distribuição *a posteriori* de θ . Como a distribuição *a posteriori* contém toda informação de θ , partindo dela podemos realizar processos inferenciais para a obtenção de informações referente à θ .

2.1 Teorema de Bayes

Considere uma quantidade de interesse desconhecida θ (tipicamente não observável). A informação de que dispomos sobre θ , resumida probabilisticamente através de $p(\theta)$, pode ser aumentada observando-se uma quantidade aleatória X relacionada com θ . A distribuição amostral $p(x|\theta)$ define esta relação. A ideia de que após observar $X = x$ a quantidade de informação sobre θ aumenta é bastante intuitiva e o teorema de Bayes é a

regra de atualização utilizada para quantificar este aumento de informação,

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} = \frac{p(x|\theta)p(\theta)}{p(x)} = \frac{p(x|\theta)p(\theta)}{\int p(\theta, x)d\theta}$$

Podemos notar que $1/p(x)$ não depende de θ , podendo ser considerado como uma constante normalizadora de $p(\theta|x)$. Para um valor fixo de x , a função $L(\theta; x) = p(x|\theta)$ fornece a verossimilhança de cada um dos possíveis valores de θ enquanto que $p(\theta)$ é chamada de distribuição *a priori* de θ . Associando estas duas fontes de informação, obtemos a distribuição *a posteriori* de θ , $p(\theta|x)$. Assim, podemos reescrever o teorema de Bayes, sob o ponto de vista Bayesiano, da seguinte forma

$$p(\theta|y) \propto L(\theta)p(\theta),$$

em que y refere-se ao conjunto de observações de x .

2.2 Distribuições *a Priori* e *a Posteriori*

A utilização de informação *a priori* em inferência Bayesiana requer a especificação de uma distribuição *a priori* para a quantidade de interesse θ . Esta distribuição deve representar (probabilisticamente) o conhecimento que se tem sobre θ antes da realização do experimento.

Dessa forma, obtendo informação *a priori* do pesquisador, podemos aplicar distribuições *a priori* informativas. Entretanto, podemos nos deparar com situações em que não possuímos informações suficientes *a priori* sobre θ , assim especificamos distribuições *a priori* não informativas. Nesse contexto, é natural que diferentes pesquisadores possam ter diferentes graus de incerteza sobre θ (especificando modelos distintos).

Por outro lado, a distribuição *a posteriori* contém toda a informação do parâmetro a ser estimado θ , pois é resultado da combinação da distribuição *a priori* e a função de verossimilhança, podendo assim realizar processos inferenciais para obter informações a respeito de θ . Estes processos inferenciais envolvem a integração de funções, muitas vezes complexas, em que o núcleo da densidade *a posteriori* não apresenta forma conhecida e, portanto, tornando a utilização de métodos numéricos importantes para a aproximação

da distribuição *a posteriori* de interesse.

Nas últimas décadas, as técnicas de simulação estocástica mais utilizadas foram as técnicas de Monte Carlo via cadeias de Markov (MCMC) que contribuíram para a propagação e desenvolvimento da teoria Bayesiana.

Na inferência Bayesiana, a metodologia MCMC é uma alternativa aos métodos não iterativos em problemas complexos, sendo que possui a vantagem de obter a amostra da densidade desejada e apresentando grande importância devido a alta complexidade das densidades *a posteriori*. Entre os métodos MCMC mais utilizados temos: o algoritmo de Metropolis-Hastings e o amostrador de Gibbs.

O algoritmo de Metropolis-Hastings foi inicialmente proposto por Metropolis et al. [37] e generalizado por Hastings [25]. O propósito do método é obter amostras das distribuições condicionais que não apresentam formas conhecidas. Caso tenhamos distribuições condicionais com formas conhecidas, podemos utilizar o método amostrador de Gibbs, que é um caso especial do algoritmo de Metropolis-Hastings. O amostrador de Gibbs foi introduzido por Geman e Geman [22] e tornou-se popular através de Gelfand e Smith [20].

2.3 Diagnóstico de Convergência

O diagnóstico da convergência das cadeias das amostras das distribuições *a posteriori* de interesse é um importante procedimento para investigar quando aplica-se os algoritmos MCMC. Existem métodos formais e informais de identificação e monitoração de convergência (ver por exemplo, Gamerman e Lopes [18]).

Os métodos informais são baseados nas técnicas gráficas em que analisa-se a trajetória das cadeias geradas e verifica se houve convergência para o mesmo ponto de estabilidade. Porém, essas técnicas devem ser utilizadas com cautela, e devem ser acompanhadas de alguma fundamentação teórica.

Os métodos formais diagnosticam a convergência baseados na exploração de propriedades estatísticas da cadeia observada. Muitos testes de diagnóstico formais têm sido desenvolvidos na literatura, dentre eles o diagnóstico proposto por Geweke [23].

2.3.1 Diagnóstico de Geweke

Geweke [23] propôs um diagnóstico de convergência baseado na aplicação de técnicas usuais em séries temporais para checar convergência da cadeia gerada. Seja um número

n suficientemente grande de iterações. O objetivo é testar a igualdade das médias $\bar{\psi}_a$ e $\bar{\psi}_b$ calculadas através das n_a primeiras iterações e n_b últimas iterações. Se a cadeia é estacionária, logo a média $\bar{\psi}_a$ deverá ser similar à média $\bar{\psi}_b$.

Considerando que $\widehat{Var}(\psi_a)$ e $\widehat{Var}(\psi_b)$ são os respectivos estimadores das variâncias assintóticas de $\bar{\psi}_a$ e $\bar{\psi}_b$, assumindo que as razões n_a/n e n_b/n são fixas e $n \rightarrow \infty$, pode-se mostrar que,

$$z_G = \frac{\bar{\psi}_a - \bar{\psi}_b}{\sqrt{\widehat{Var}(\psi_a) + \widehat{Var}(\psi_b)}} \xrightarrow{d} N(0, 1).$$

Assim, valores extremos para a diferença padronizada entre as médias ergódicas indicam falta de convergência. Ou seja, mais especificamente, parâmetros com $|z_G| > 1,96$ indicam não convergência da cadeia. Geweke [23] sugeriu o uso dos valores $n_a = 0,1n$ e $n_b = 0,5n$ para tentar fornecer um diagnóstico poderoso.

2.4 Estimação

Há a necessidade de resumir a informação contida na distribuição *a posteriori* através de valores numéricos. O caso mais simples é a estimação pontual de θ na qual resume toda a distribuição *a posteriori* em um único valor, denotado por $\hat{\theta}$.

Outra forma de sumarizar a informação contida é através de intervalos que fornecem a região de credibilidade de valores de θ . A seguir são apresentados os conceitos de estimação pontual e por intervalos sob o ponto de vista Bayesiano.

2.4.1 Estimação Pontual

Um estimador Bayesiano para θ dado por $\hat{\theta} = d(\mathbf{x})$ é obtido ao minimizar o erro esperado (função de risco) com respeito à distribuição *a posteriori* para θ . Vamos considerar uma função de perda quadrática dada por

$$L(d; \theta) = (d(\mathbf{x}) - \theta)^2.$$

Assim, devemos encontrar $d(\mathbf{x})$ que minimiza o risco Bayesiano definido como

$$\begin{aligned}
R(d(\mathbf{x}); \theta) &= E_{\theta|\mathbf{x}}[L(d; \theta)] \\
&= \int_{\theta} (d(\mathbf{x}) - \theta)^2 \pi(\theta|\mathbf{x}) d\theta.
\end{aligned}$$

Sendo $d(\mathbf{x})$ uma função diferenciável, o estimador de Bayes com respeito à função de perda quadrática é dada por

$$\frac{dR(d(\mathbf{x}); \theta)}{d(d)} = 2 \int (d(\mathbf{x}) - \theta) \pi(\theta|\mathbf{x}) d\theta = 0,$$

ou seja, $\hat{\theta} = d(\mathbf{x}) = E(\theta|\mathbf{x})$, que é a média *a posteriori* de θ .

2.4.2 Estimação por Intervalos

A definição de intervalos de credibilidade *a posteriori* para θ é dada por:

Seja $C \in \Theta$, C caracteriza uma região de $100(1 - \alpha)\%$ de credibilidade para θ se $P(\theta \in C|\mathbf{x}) \geq 1 - \alpha$. Neste caso, $1 - \alpha$ é chamado de nível de credibilidade.

Assim, quanto menor for o tamanho do intervalo mais concentrada é a distribuição do parâmetro, ou seja, o tamanho do intervalo informa sobre a dispersão de θ .

Outro aspecto importante é que os intervalos de credibilidade são invariantes a transformações 1 a 1, $\phi(\theta)$. Ou seja, se $C = [a, b]$ é um intervalo de credibilidade $100(1 - \alpha)\%$ para θ , então $[\phi(a), \phi(b)]$ é um intervalo de credibilidade $100(1 - \alpha)\%$ para $\phi(\theta)$.

2.5 Critério de Seleção de Modelos

O avanço dos métodos MCMC possibilitou a construção de modelos com maior complexidade, fazendo com que a escolha entre os modelos tornasse fundamental na análise estatística. Vários critérios de seleção de modelos são propostos sob o enfoque Bayesiano. A seguir destacaremos o critério *Deviance Information Criterion* (DIC).

2.5.1 Critério DIC

Spiegelhalter et al. [46] propuseram o critério DIC considerado como uma generalização do critério AIC (*Akaike's Information Criterion*), sendo aplicado especialmente em métodos de Monte Carlo via cadeias de Markov. O critério DIC é baseado na distribuição *a posteriori*, denominado desvio, dada por

$$D(\boldsymbol{\theta}) = -2\ln L(\boldsymbol{\theta}),$$

em que $\boldsymbol{\theta}$ é um vetor de parâmetros desconhecidos do modelo e $L(\boldsymbol{\theta})$ é a função de verossimilhança. Dessa forma, o DIC é definido como

$$\begin{aligned} DIC_k &= D_k(\hat{\boldsymbol{\theta}}) + 2p_{D_k} \\ &= \bar{D}_k + p_{D_k}, \end{aligned}$$

em que $D_k(\hat{\boldsymbol{\theta}})$ é o desvio calculado na média *a posteriori* $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{x})$, $p_{D_k} = \bar{D}_k - D_k(\hat{\boldsymbol{\theta}})$ é o número efetivo de parâmetros no k -ésimo modelo e também é associado a complexidade do modelo, em que $\bar{D}_k = E[D_k(\boldsymbol{\theta})|\mathbf{x}]$ é a média *a posteriori* do desvio que corresponde a qualidade do ajuste dos dados ao modelo.

Assim, menores valores do DIC indicam o melhor ajuste do modelo e esses valores podem ser negativos.

3 Transformações e Modelo de Regressão Dirichlet para Dados Composicionais

Neste capítulo, apresentamos as transformações logaritmo da razão (*alr*) e Box-Cox em dados composicionais para modelos de regressão na presença de uma covariável proposto em Iyengar e Dey [28]. Além disso, também descrevemos o modelo de regressão Dirichlet proposto em Hijazi e Jernigan [27], em que aplicamos a transformação *alr* para os dados composicionais e os parâmetros para esse modelo foram estimados através da inferência clássica nos dois conjuntos de dados apresentados com o objetivo de compararmos os resultados obtidos com o enfoque Bayesiano.

3.1 Transformações Logaritmo da Razão (*alr*) e Box-Cox

Para este trabalho, utilizamos a aplicação de modelos de regressão em dados composicionais, na qual a variável resposta do modelo é o vetor com as proporções dos componentes e $\mathbf{z}_i$ é o vetor das variáveis explicativas (covariável). A restrição existente da soma dos componentes do vetor da variável resposta ser igual a um é o que difere dos outros modelos de regressão.

Dessa forma, podemos considerar o modelo de regressão (ver por exemplo, Iyengar e Dey [28]) dado por

$$\mathbf{y}_i = \beta_0 + \mathbf{z}_i \beta_1 + \epsilon_i, \quad \text{para } i = 1, \dots, n, \quad (3.1)$$

em que $\mathbf{z}_i$ é um vetor $(1 \times p)$ de covariáveis associadas à *i-ésima* amostra; β_0 é um vetor $(1 \times g)$ de interceptos; β_1 é um vetor $(p \times g)$ de coeficientes de regressão; ϵ_i é o vetor

de erros, $\mathbf{y}_i = (y_{i1}, \dots, y_{ig})$ é um vetor $(1 \times g)$ em que $g = G - 1$, sendo G o número de componentes dos dados composicionais.

De acordo com Achcar e Obage [1], podemos considerar $y_{ij} = H(x_{ij}/x_{iG})$, $i = 1, \dots, n$ e $j = 1, \dots, g$, sendo $H(\bullet)$ a função de transformação escolhida para assegurar que o vetor resultante tenha componentes reais, em que x_{ij} representa a i -ésima observação para o j -ésimo componente, tal que $x_{i1} > 0, \dots, x_{iG} > 0$ e $\sum_{j=1}^G x_{ij} = 1$, para $i = 1, \dots, n$.

A transformação logaritmo da razão (*alr*) para a análise de dados composicionais é dada por

$$y_{ij} = H\left(\frac{x_{ij}}{x_{iG}}\right) = \log\left(\frac{x_{ij}}{x_{iG}}\right). \quad (3.2)$$

Alternativamente, podemos utilizar a transformação Box-Cox em dados composicionais, que é dada por

$$y_{ij} = H\left(\frac{x_{ij}}{x_{iG}}\right) = \begin{cases} \frac{\left(\frac{x_{ij}}{x_{iG}}\right)^{\lambda_j} - 1}{\lambda_j} & \text{se } \lambda_j \neq 0, \\ \log\left(\frac{x_{ij}}{x_{iG}}\right) & \text{se } \lambda_j = 0. \end{cases} \quad (3.3)$$

Nota-se que a transformação *alr* é um caso especial da transformação Box-Cox quando $\lambda = 0$.

Deve-se enfatizar que para o modelo de regressão composicional, a permutação é invariante, ou seja, uma diferente escolha do componente no denominador da transformação *alr* ou Box-Cox conduzirá à resultados compatíveis. Neste trabalho, utilizamos como denominador o último componente de ambos os conjuntos de dados aplicados.

3.2 Modelo de Regressão Dirichlet

Campbell e Mosimann [12] abordaram uma extensão da distribuição Dirichlet para uma classe de modelos de regressão Dirichlet. Eles mostraram que essa classe de modelos pode ser desenvolvida através de reparametrização dos parâmetros da distribuição Dirichlet em termos da covariável associada e, além disso, que a estrutura de covariância para

os modelos de regressão Dirichlet não são necessariamente negativos, como no caso da distribuição Dirichlet [27].

Assim, podemos assumir $\mathbf{x} = (x_{i1}, \dots, x_{iG})$ um vetor positivo ($1 \times G$) com distribuição Dirichlet com parâmetros positivos $(\lambda_1, \dots, \lambda_G)$ em que sua função densidade de probabilidade é dada por

$$f(\mathbf{x}) = \left(\Gamma(\lambda) / \prod_{j=1}^G \Gamma(\lambda_j) \right) \prod_{j=1}^G x_j^{\lambda_j-1},$$

em que $\sum_{j=1}^G x_j = 1$ e $\lambda = \sum_{j=1}^G \lambda_j$.

O modelo de regressão Dirichlet é facilmente obtido permitindo que os parâmetros de uma distribuição Dirichlet mudem com covariáveis associadas [11]. Para um dado vetor de covariável $\mathbf{z}_i$, $i = 1, \dots, n$, cada parâmetro λ_j pode ser escrito como uma combinação linear $\lambda(\mathbf{z}_i)$ da covariável $\mathbf{z}_i$, ou seja,

$$\lambda_j(\mathbf{z}_i) = z_{i,1}\beta_{1,j} + z_{i,2}\beta_{2,j} + \dots + z_{i,C}\beta_{C,j} = \mathbf{z}_i\boldsymbol{\beta}_j \quad (3.4)$$

em que os parâmetros a serem estimados são $\boldsymbol{\beta} = (\beta_{k,j}, j = 1, \dots, G, k = 1, \dots, C)$.

Para o presente estudo, podemos reescrever (3.4) da seguinte forma

$$\lambda_j(\mathbf{z}_i) = \sum_i \sum_j z_{i,1}\beta_{1,j} = \mathbf{z}_i\boldsymbol{\beta}_j, \quad (3.5)$$

em que $i = 1, \dots, n$ e $j = 1, \dots, G$.

4 Análise Bayesiana para os Dados da Superliga de Vôlei Masculina

Neste capítulo apresentamos a análise Bayesiana aplicada em modelos de regressão com dados composicionais que foram ajustadas à um conjunto de dados referente aos jogos da Superliga de Vôlei Masculina Brasileira 2011/2012 (Apêndice A, disponível em [15]). Mais especificamente, aos times que jogaram e venceram nos 1º e 2º turnos, nos quais os pontos do time vencedor de cada jogo foram definidos como composição e os fundamentos ataque, bloqueio, saque e erro do adversário como proporções de cada composição.

Os pontos do time vencedor em cada jogo é formado por quatro componentes, sendo que denotamos x_{i1} como a proporção dos pontos em ataque, x_{i2} a proporção dos pontos em bloqueio, x_{i3} a proporção dos pontos em saque e x_{i4} a proporção dos pontos em erros do adversário. A covariável associada ao i -ésimo jogo é definida por

$$z_i = \begin{cases} 1, & \text{se o time ganhou pelo menos uma vez a Superliga nos últimos 12 anos} \\ 0, & \text{caso contrário.} \end{cases}$$

O objetivo é verificar se há dependência entre as proporções dos componentes (ataque, bloqueio, saque, erro adversário) e se o time que venceu o i -ésimo jogo já venceu pelo menos uma vez a Superliga nos últimos 12 anos.

Na Figura 4.1 apresentamos os componentes dos pontos dos jogos de vôlei da Superliga de acordo com a covariável associada, sendo que 0 representa que o time não ganhou nenhuma vez a Superliga de vôlei e 1 representa que o time já ganhou pelo menos uma vez nos últimos 12 anos. Observando-a, percebe-se que apenas para o componente bloqueio há evidência de uma pequena diferença entre os times que não venceram e os times que já venceram a Superliga.

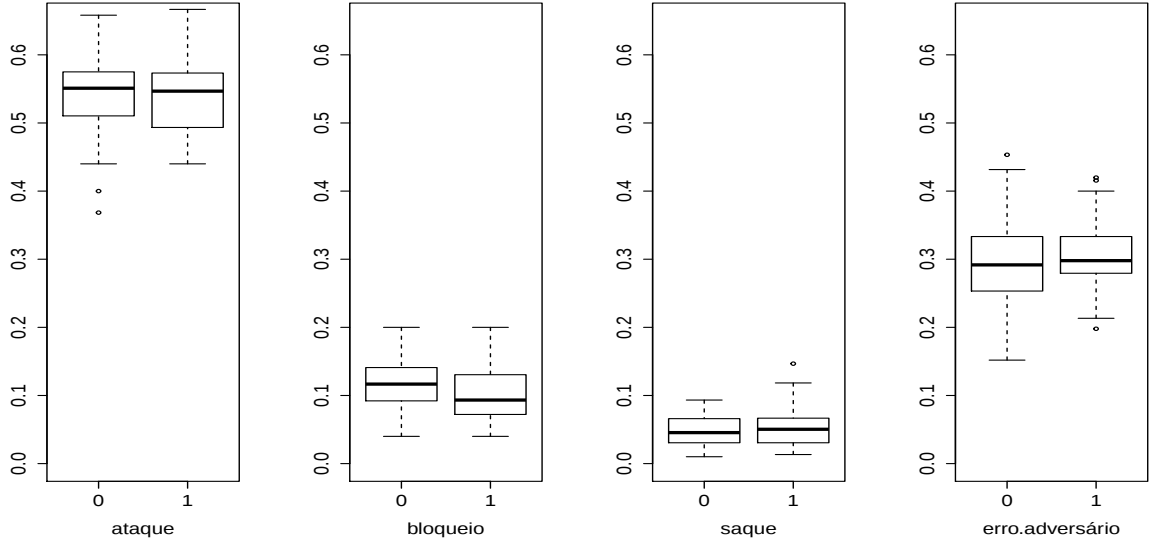


Figura 4.1: Box plots para os componentes dos pontos dos times da Superliga de vôlei.

4.1 Transformação *alr* - Modelo 1

Para análise dos dados composicionais apresentados no Apêndice A, x_{i1} , x_{i2} , x_{i3} e x_{i4} , para $i = 1, \dots, n$, consideramos a transformação *alr* apresentada em (3.2). Com isso, obtemos

$$y_{i1} = \log\left(\frac{x_{i1}}{x_{i4}}\right), \quad y_{i2} = \log\left(\frac{x_{i2}}{x_{i4}}\right) \quad \text{e} \quad y_{i3} = \log\left(\frac{x_{i3}}{x_{i4}}\right).$$

Conforme apresentado em (3.1), o modelo de regressão para os dados transformados y_{i1} , y_{i2} e y_{i3} , denominado por modelo 1, é dado por

$$\begin{aligned} y_{i1} &= \beta_{01} + \beta_{11}z_i + \epsilon_{i1}, \\ y_{i2} &= \beta_{02} + \beta_{12}z_i + \epsilon_{i2} \quad \text{e} \\ y_{i3} &= \beta_{03} + \beta_{13}z_i + \epsilon_{i3}, \end{aligned} \tag{4.1}$$

em que $\mathbf{y}_i$ representa a proporção transformada dos pontos do j -ésimo componente (ataque, bloqueio, saque, erro adversário) no i -ésimo jogo, z_i é uma variável *dummy*, em que $z_i = 0$ indica que o time não venceu a Superliga e $z_i = 1$ indica que o time já ganhou pelo menos uma vez a Superliga nos últimos 12 anos. Além disso, β_{0j} representa a média

da proporção de pontos no j -ésimo componente em relação ao componente x_{i4} erro do adversário para o time que não venceu a Superliga, β_{1j} indica se há ou não efeito da covariável associada ao i -ésimo jogo e ϵ_i representa o vetor de erros.

Assumiremos $\epsilon_i \sim N(\mathbf{0}, \Sigma)$, em que $\mathbf{0}$ é o vetor de zeros e Σ é a matriz de variâncias e covariâncias especificada por

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}.$$

Considerando o modelo apresentado em (4.1), assumiremos que os erros ϵ_{i1} , ϵ_{i2} e ϵ_{i3} são mutuamente independentes. Desta forma, obtém-se

$$\begin{aligned} \epsilon_{i1} &= y_{i1} - \beta_{01} - \beta_{11}z_i, \\ \epsilon_{i2} &= y_{i2} - \beta_{02} - \beta_{12}z_i \quad \text{e} \\ \epsilon_{i3} &= y_{i3} - \beta_{03} - \beta_{13}z_i. \end{aligned}$$

Além disso, podemos estimar as proporções dos componentes $\alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i4}$, em que $\alpha_{i1} + \alpha_{i2} + \alpha_{i3} + \alpha_{i4} = 1$, com $\alpha_{i1} > 0, \alpha_{i2} > 0, \alpha_{i3} > 0$ e $\alpha_{i4} > 0$, quando utiliza a transformação *alr*, obtendo a seguinte relação a partir de (4.1)

$$\begin{aligned} \log\left(\frac{\alpha_{i1}}{\alpha_{i4}}\right) &= \beta_{01} + \beta_{11}z_i, \\ \log\left(\frac{\alpha_{i2}}{\alpha_{i4}}\right) &= \beta_{02} + \beta_{12}z_i \quad \text{e} \\ \log\left(\frac{\alpha_{i3}}{\alpha_{i4}}\right) &= \beta_{03} + \beta_{13}z_i. \end{aligned} \tag{4.2}$$

Aplicando a função exponencial em (4.2), temos que

$$\frac{\alpha_{i1}}{\alpha_{i4}} = e^{\beta_{01} + \beta_{11}z_i}, \quad \frac{\alpha_{i2}}{\alpha_{i4}} = e^{\beta_{02} + \beta_{12}z_i} \quad \text{e} \quad \frac{\alpha_{i3}}{\alpha_{i4}} = e^{\beta_{03} + \beta_{13}z_i},$$

assim

$$\begin{cases} \alpha_{i1} &= \alpha_{i4}(e^{\beta_{01}+\beta_{11}z_i}) \\ \alpha_{i2} &= \alpha_{i4}(e^{\beta_{02}+\beta_{12}z_i}) \\ \alpha_{i3} &= \alpha_{i4}(e^{\beta_{03}+\beta_{13}z_i}). \end{cases} \quad (4.3)$$

Sabemos que $\alpha_{i1} + \alpha_{i2} + \alpha_{i3} + \alpha_{i4} = 1$, ou seja, $\alpha_{i4} = 1 - (\alpha_{i1} + \alpha_{i2} + \alpha_{i3}) \Rightarrow \alpha_{i1} + \alpha_{i2} + \alpha_{i3} = 1 - \alpha_{i4}$.

Então, a partir de (4.3), obtemos a seguinte relação

$$\alpha_{i1} + \alpha_{i2} + \alpha_{i3} = \alpha_{i4} (e^{\beta_{01}+\beta_{11}z_i} + e^{\beta_{02}+\beta_{12}z_i} + e^{\beta_{03}+\beta_{13}z_i}).$$

Dessa forma, podemos reescrever a expressão $1 - \alpha_{i4}$ como

$$\begin{aligned} 1 - \alpha_{i4} &= \alpha_{i4} (e^{\beta_{01}+\beta_{11}z_i} + e^{\beta_{02}+\beta_{12}z_i} + e^{\beta_{03}+\beta_{13}z_i}) \Rightarrow \\ 1 &= \alpha_{i4} (1 + e^{\beta_{01}+\beta_{11}z_i} + e^{\beta_{02}+\beta_{12}z_i} + e^{\beta_{03}+\beta_{13}z_i}) \Rightarrow \\ \alpha_{i4} &= \frac{1}{(1 + e^{\beta_{01}+\beta_{11}z_i} + e^{\beta_{02}+\beta_{12}z_i} + e^{\beta_{03}+\beta_{13}z_i})}. \end{aligned} \quad (4.4)$$

Logo, substituindo (4.4) em (4.3), obtemos as proporções verdadeiras dos componentes para a transformação *alr*,

$$\begin{aligned} \alpha_{ij} &= \frac{e^{\beta_{0j}+\beta_{1j}z_i}}{(1 + e^{\beta_{01}+\beta_{11}z_i} + e^{\beta_{02}+\beta_{12}z_i} + e^{\beta_{03}+\beta_{13}z_i})} \quad \text{e} \\ \alpha_{i4} &= \frac{1}{(1 + e^{\beta_{01}+\beta_{11}z_i} + e^{\beta_{02}+\beta_{12}z_i} + e^{\beta_{03}+\beta_{13}z_i})}, \end{aligned} \quad (4.5)$$

em que $i = 1, \dots, 128$ e $j = 1, 2, 3$. De acordo com Achcar e Obage [1], a parametrização $\alpha_{i1} > 0$, $\alpha_{i2} > 0$, $\alpha_{i3} > 0$ e $\alpha_{i4} > 0$ pode ser usada para obter inferências sobre as composições em cada valor da covariável.

4.1.1 Análise Bayesiana - Modelo 1

A função densidade de probabilidade conjunta de $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ dado o vetor de parâmetros $\mathbf{v}_1 = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{11}, \beta_{12}, \beta_{13}, \sigma_1^2, \sigma_2^2, \sigma_3^2)$ é dada por

$$f(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 | \mathbf{v}_1) = \prod_{j=1}^3 \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left[-\frac{1}{2\sigma_j^2} (y_{ij} - \beta_{0j} - \beta_{1j}z_i)^2 \right],$$

pois, as suposições de normalidade e independência foram consideradas, e ainda, $y_{ij} \sim N(\beta_{0j} + \beta_{1j}z_i, \sigma_j^2)$, para $j = 1, 2, 3$ e $i = 1, \dots, n$.

Assim, assumindo o modelo (4.1), a função de verossimilhança para os parâmetros $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \beta_{03})$, $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13})$ e $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ é dada por

$$L(\mathbf{v}_1) = \prod_{j=1}^3 (\sigma_j^2)^{-n/2} \exp \left(-\frac{1}{2\sigma_j^2} \sum_{i=1}^n \epsilon_{ij}^2 \right), \quad (4.6)$$

em que

$$\sum_{i=1}^n \epsilon_{ij}^2 = \sum_{i=1}^n (y_{ij} - \beta_{0j} - \beta_{1j}z_i)^2, \text{ para } j = 1, 2, 3.$$

Para a estimação dos parâmetros do modelo (4.1), as seguintes distribuições *a priori* para $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \beta_{03})$, $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13})$ e $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ são consideradas

$$\begin{aligned} \beta_{0j} &\sim N(a_{0j}, b_{0j}^2), \\ \beta_{1j} &\sim N(a_{1j}, b_{1j}^2), \\ \sigma_j^2 &\sim IG(c_j, d_j), \end{aligned} \quad (4.7)$$

em que $N(a, b^2)$ denota uma distribuição Normal com média a e variância b^2 , $IG(c, d)$ denota uma distribuição Gama inversa com média $d/(c-1)$ e variância $d^2/[(c-1)^2(c-2)]$, $c > 2$; e $a_{0j}, b_{0j}, a_{1j}, b_{1j}, c_j$ e d_j são hiperparâmetros conhecidos, para $j = 1, 2, 3$. Além disso, assumimos independência *a priori* entre os parâmetros.

Desta forma, a partir do teorema de Bayes que associa a função de verossimilhança (4.6) com as distribuições *a priori* (4.7), obtemos a distribuição *a posteriori* conjunta para os parâmetros $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02}, \beta_{03})$, $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13})$ e $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ dada por

$$\begin{aligned} \pi(\beta_0, \beta_1, \sigma^2 | \mathbf{y}) &\propto \prod_{j=1}^3 \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \times \prod_{j=1}^3 \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \\ &\times \prod_{j=1}^3 (\sigma_j^2)^{-(c_j+1)} \exp \left(-\frac{d_j}{\sigma_j^2} \right) \times \prod_{j=1}^3 (\sigma_j^2)^{-n/2} \exp \left(-\frac{1}{2\sigma_j^2} \sum_{i=1}^n \epsilon_{ij}^2 \right). \end{aligned}$$

Consequentemente, as distribuições *a posteriori* condicionais utilizadas para o amos-
trador de Gibbs (Gelfand e Smith [20]) são dadas por:

$$i) \quad \pi(\beta_{0j} | \beta_1, \sigma^2, \mathbf{y}) \propto \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \times \exp \left[-\frac{1}{2\sigma_j^2} \sum_{i=1}^n (\beta_{0j} - \mu_i^{(j)})^2 \right],$$

em que $\mu_i^{(j)} = y_{ij} - \beta_{1j} z_i$; $i = 1, \dots, n$ e $j = 1, 2, 3$.

Logo,

$$\pi(\beta_{0j} | \beta_1, \sigma^2, \mathbf{y}) \sim N \left[\frac{a_{0j}\sigma_j^2 + b_{0j} \sum_{i=1}^n \mu_i^{(j)}}{\sigma_j^2 + nb_{0j}^2}, \frac{b_{0j}^2\sigma_j^2}{\sigma_j^2 + nb_{0j}^2} \right]. \quad (4.8)$$

$$ii) \quad \pi(\beta_{1j} | \beta_0, \sigma^2, \mathbf{y}) \propto \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \times \exp \left[-\frac{1}{2\sigma_j^2} \sum_{i=1}^n (\beta_{1j} z_i - \theta_i^{(j)})^2 \right],$$

em que $\theta_i^{(j)} = y_{ij} - \beta_{0j}$; $i = 1, \dots, n$ e $j = 1, 2, 3$.

Logo,

$$\pi(\beta_{1j} | \beta_0, \sigma^2, \mathbf{y}) \sim N \left[\frac{a_{1j}\sigma_j^2 + b_{1j} \sum_{i=1}^n z_i \theta_i^{(j)}}{\sigma_j^2 + b_{1j}^2 \sum_{i=1}^n z_i^2}, \frac{b_{1j}^2\sigma_j^2}{\sigma_j^2 + b_{1j}^2 \sum_{i=1}^n z_i^2} \right]. \quad (4.9)$$

$$\begin{aligned}
iii) \quad \pi(\sigma_j^2 | \beta_0, \beta_1, \mathbf{y}) &\propto (\sigma_j^2)^{-(c_j+1)} \exp\left(-\frac{d_j}{\sigma_j^2}\right) \times (\sigma_j^2)^{-n/2} \exp\left(-\frac{1}{2\sigma_j^2} \sum_{i=1}^n \epsilon_{ij}^2\right) \\
&\propto (\sigma_j^2)^{-(c_j+n/2+1)} \exp\left[-\frac{1}{\sigma_j^2} \left(d_j + \frac{1}{2} \sum_{i=1}^n \epsilon_{ij}^2\right)\right],
\end{aligned}$$

em que $\epsilon_{ij} = y_{ij} - \beta_{0j} - \beta_{1j}z_i$; $i = 1, \dots, n$ e $j = 1, 2, 3$.

Logo,

$$\pi(\sigma_j^2 | \beta_0, \beta_1, \mathbf{y}) \sim IG\left[c_j + \frac{n}{2}, d_j + \frac{1}{2} \sum_{i=1}^n \epsilon_{ij}^2\right]. \quad (4.10)$$

Podemos observar que as distribuições *a posteriori* condicionais (4.8), (4.9) e (4.10) apresentaram formas fechadas e, portanto, obtemos as estimativas dos parâmetros do modelo (4.1) através do algoritmo amostrador de Gibbs. A implementação do algoritmo foi realizada no *software* OPENBUGS (Lunn et al., [32]).

Para a análise Bayesiana dos dados apresentados no Apêndice A, considera-se a transformação *alr* com erros normais não correlacionados para o modelo proposto (4.1).

Assumindo distribuições *a priori* não informativas para $\beta_{0j}, \beta_{1j}, \sigma_j^2$ com hiperparâmetros cujos valores foram $a_{0j} = 0, b_{0j} = b_{1j} = 100, c_j = 0,001$ e $d_j = 0,001$, $j = 1, 2, 3$, geramos no *software* OPENBUGS 210.000 iterações, nas quais foram descartadas as 10.000 iterações iniciais (“burn-in”). Após o período de “burn-in”, consideramos saltos de tamanho 20 para obtermos amostras não correlacionadas, na qual resultou uma amostra final de tamanho 10.000 para cada parâmetro. A convergência do algoritmo foi verificada pelos gráficos dos traços *a posteriori* das amostras de Gibbs simuladas e pelo diagnóstico de Geweke (Geweke, [23]) implementado no *software* R através do pacote CODA.

Na Tabela 4.1 apresentamos os sumários *a posteriori* dos parâmetros do modelo (4.1) e os valores correspondentes da estatística de teste para o diagnóstico de convergência Geweke.

Podemos verificar na Tabela 4.1 que houve efeito significativo do time ter vencido pelo menos uma vez a Superliga para a proporção de pontos de bloqueio (intervalo de credibilidade não inclui o zero no parâmetro β_{12}), sendo que o valor de β_{12} é negativo, indicando que os times que não venceram nenhuma vez a Superliga fizeram mais pontos no bloqueio.

Tabela 4.1: Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke - Modelo 1.

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	0,622	0,033	(0,557; 0,685)	0,151
β_{02}	-0,989	0,053	(-1,092; -0,886)	0,345
β_{03}	-1,902	0,073	(-2,047; -1,758)	1,925
β_{11}	-0,045	0,047	(-0,139; 0,047)	-0,312
β_{12}	-0,171	0,074	(-0,318; -0,024)	0,067
β_{13}	0,013	0,104	(-0,189; 0,219)	-1,463
σ_1	0,070	0,009	(0,054; 0,089)	-1,389
σ_2	0,184	0,024	(0,143; 0,237)	-0,536
σ_3	0,346	0,044	(0,269; 0,440)	-0,883

Analisando os interceptos do modelo (Tabela 4.1), observamos que β_{01} é positivo, indicando que a média da proporção do componente ataque em relação ao componente erro do adversário tem uma tendência crescente, o que não ocorre com os interceptos dos componentes bloqueio e saque que são negativos, apresentando uma tendência decrescente em relação ao componente erro do adversário x_{i4} .

A Tabela 4.2 apresenta as estimativas das proporções verdadeiras para os componentes ataque, bloqueio, saque e erros do adversário para cada valor da covariável z_i considerando a transformação *alr*. Observamos que as estimativas das proporções para todos os componentes (ataque, bloqueio, saque e erros do adversário) não houve diferença em relação aos times terem vencido ou não a Superliga de vôlei, pois apresentaram o mesmo comportamento, ou seja, os valores das estimativas foram bem semelhantes, conforme mostra a Tabela 4.2.

Tabela 4.2: Estimativas das proporções dos componentes - Modelo 1.

Proporções	$z = 0$	$z = 1$
α_1	0,550	0,548
α_2	0,110	0,097
α_3	0,044	0,047
α_4	0,296	0,308

As Figuras 4.2, 4.3 e 4.4 apresentam as densidades *a posteriori* e o comportamento da cadeia gerada para os parâmetros do modelo proposto. Através dessas figuras, percebe-se que há uma certa uniformidade nos traços *a posteriori* dos parâmetros estimados, indicando possível convergência. Com a aplicação do diagnóstico de Geweke (Tabela 4.1), verificamos que as cadeias geradas apresentaram convergência, pois os valores da estatística do teste z_G encontram-se no intervalo $(-1,96; 1,96)$.

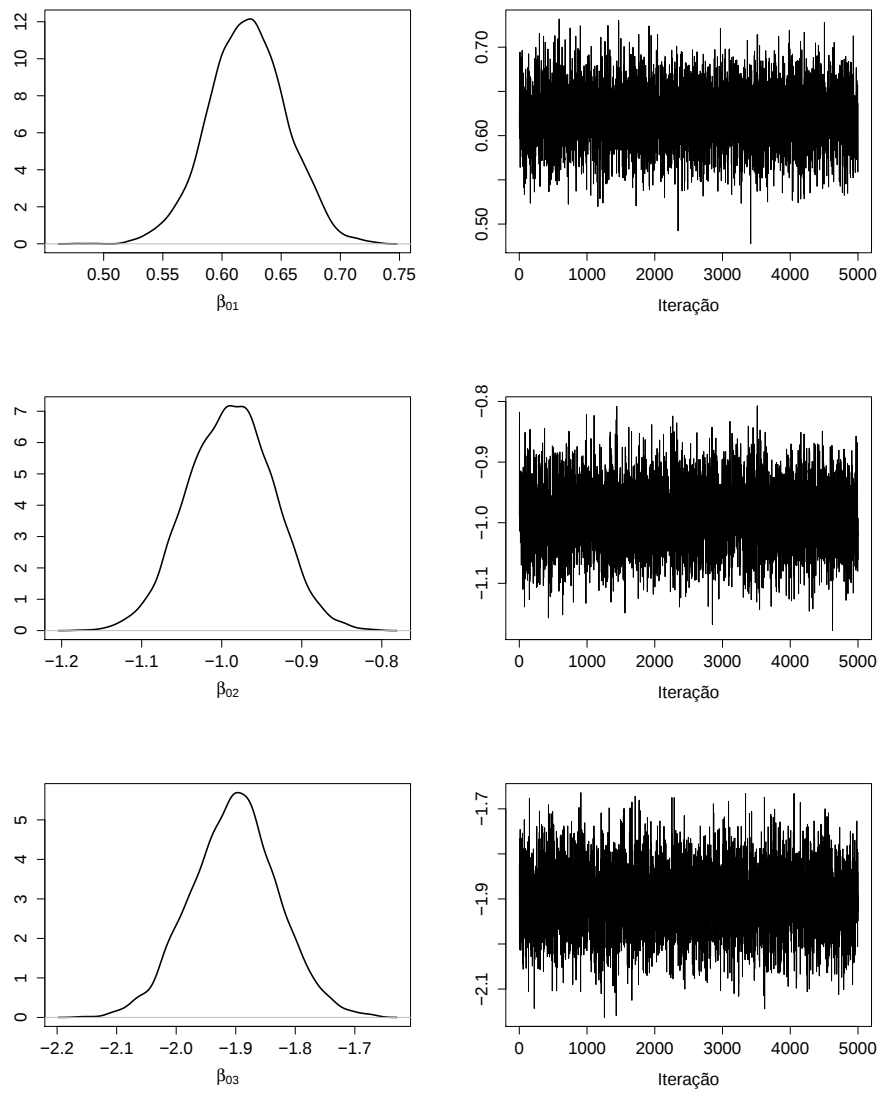


Figura 4.2: Densidades *a posteriori* e traços *a posteriori* para os parâmetros de intercepto do modelo 1.

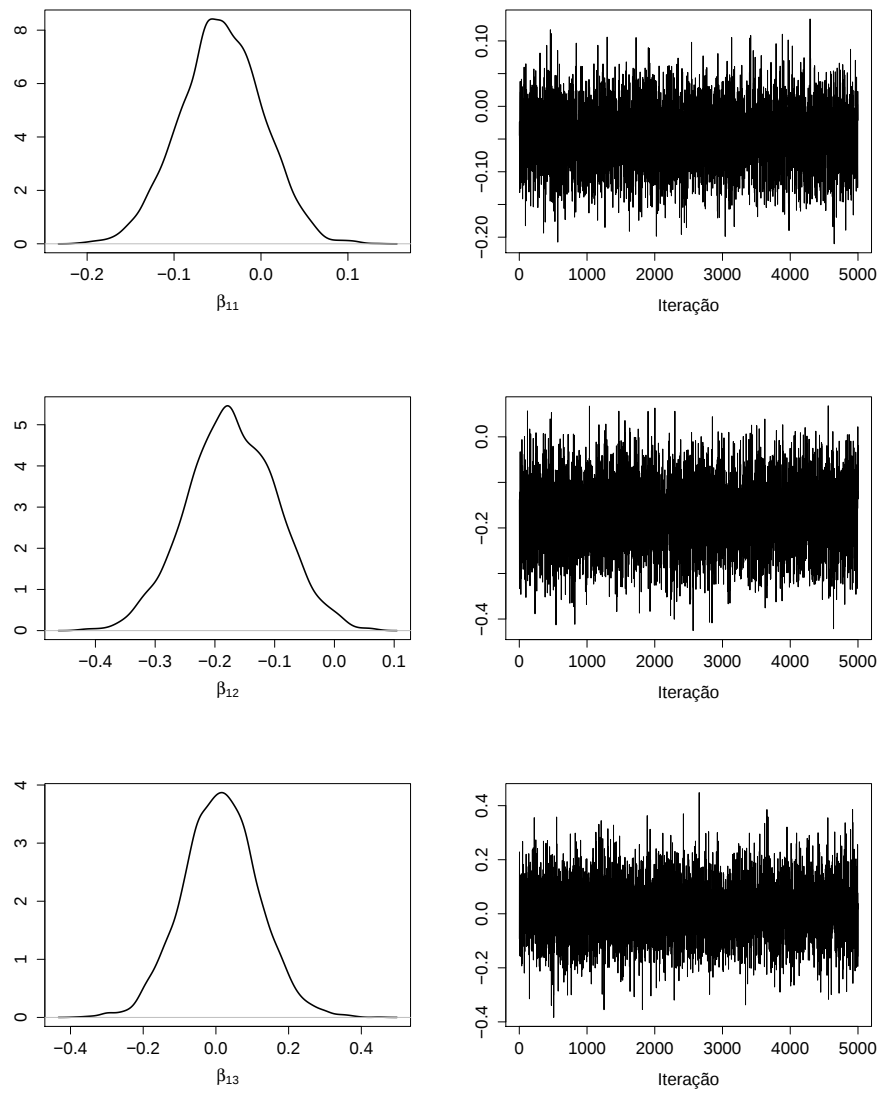


Figura 4.3: Densidades *a posteriori* e traços *a posteriori* para os parâmetros de inclinação do modelo 1.

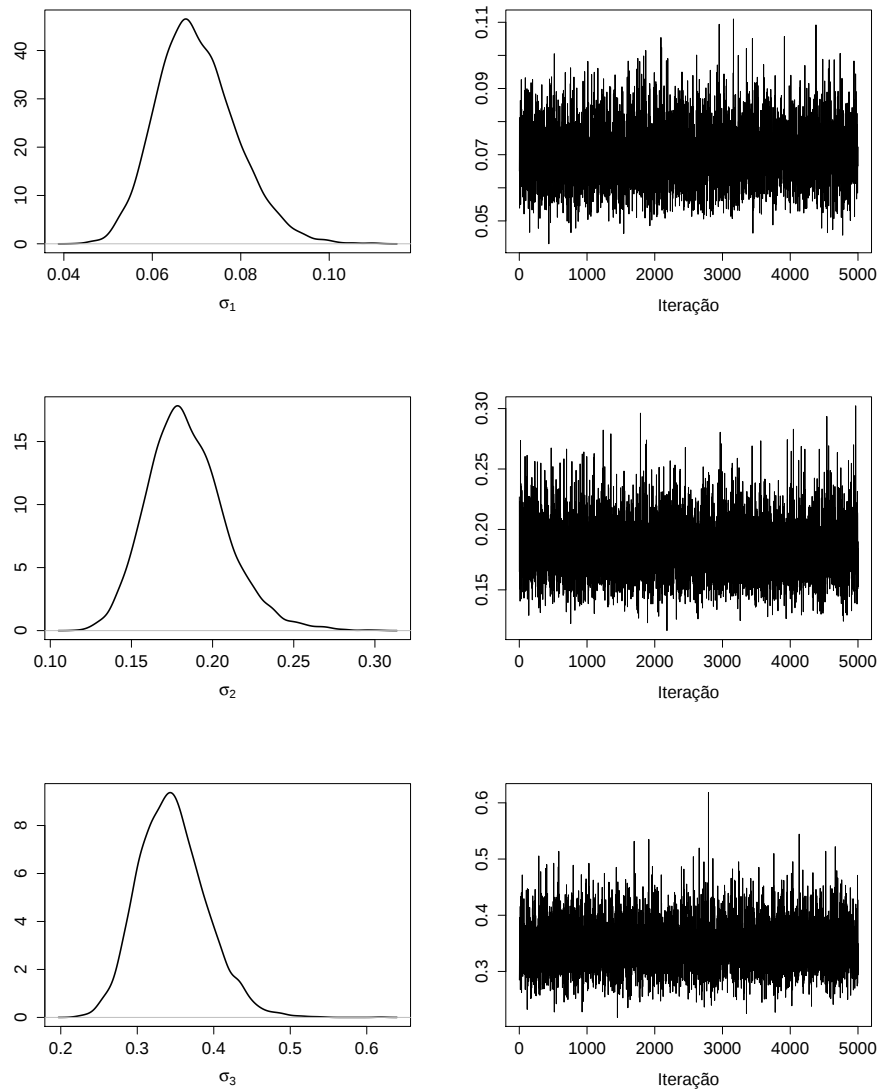


Figura 4.4: Densidades *a posteriori* e traços *a posteriori* para os desvios padrões do modelo 1.

4.2 Transformação Box-Cox - Modelo 2

Nesta seção, é desenvolvido uma aplicação da transformação Box-Cox (3.3) nos dados composicionais apresentados no Apêndice A, x_{i1} , x_{i2} , x_{i3} e x_{i4} , para $i = 1, \dots, n$.

Para a obtenção das amostras simuladas *a posteriori* foram utilizados os métodos Bayesianos baseados no amostrador de Gibbs e no algoritmo de Metropolis-Hastings.

Considerando a transformação Box-Cox para os dados da Superliga de Vôlei, obtemos

$$y_{i1}^{(\lambda_1)} = \begin{cases} \frac{\left(\frac{x_{i1}}{x_{i4}}\right)^{\lambda_1} - 1}{\lambda_1} & \text{se } \lambda_1 \neq 0, \\ \log\left(\frac{x_{i1}}{x_{i4}}\right) & \text{se } \lambda_1 = 0 \end{cases}$$

$$y_{i2}^{(\lambda_2)} = \begin{cases} \frac{\left(\frac{x_{i2}}{x_{i4}}\right)^{\lambda_2} - 1}{\lambda_2} & \text{se } \lambda_2 \neq 0, \\ \log\left(\frac{x_{i2}}{x_{i4}}\right) & \text{se } \lambda_2 = 0 \end{cases}$$

$$y_{i3}^{(\lambda_3)} = \begin{cases} \frac{\left(\frac{x_{i3}}{x_{i4}}\right)^{\lambda_3} - 1}{\lambda_3} & \text{se } \lambda_3 \neq 0, \\ \log\left(\frac{x_{i3}}{x_{i4}}\right) & \text{se } \lambda_3 = 0. \end{cases}$$

Vale ressaltar que a transformação *alr* é um caso especial da transformação Box-Cox quando $\lambda_1 = \lambda_2 = \lambda_3 = 0$.

Utilizando os dados transformados $y_{i1}^{(\lambda_1)}$, $y_{i2}^{(\lambda_2)}$ e $y_{i3}^{(\lambda_3)}$, e assumindo o modelo (4.1), podemos especificar

$$\begin{aligned} y_{i1}^{(\lambda_1)} &= \beta_{01} + \beta_{11}z_i + \epsilon_{i1}, \\ y_{i2}^{(\lambda_2)} &= \beta_{02} + \beta_{12}z_i + \epsilon_{i2} \quad \text{e} \\ y_{i3}^{(\lambda_3)} &= \beta_{03} + \beta_{13}z_i + \epsilon_{i3}. \end{aligned} \tag{4.11}$$

em que $i = 1, \dots, 128$ e ϵ_i é o vetor de erros. Da mesma forma, que em (4.2), assumiremos $\epsilon_i \sim N(\mathbf{0}, \Sigma)$, em que $\mathbf{0}$ é o vetor de zeros e Σ é a matriz de variâncias e covariâncias.

Considerando o modelo apresentado em (4.11), assumiremos que os erros ϵ_{i1} , ϵ_{i2} e ϵ_{i3} são mutuamente independentes. Assim, obtém-se

$$\begin{aligned}\epsilon_{i1} &= y_{i1}^{(\lambda_1)} - \beta_{01} - \beta_{11}z_i, \\ \epsilon_{i2} &= y_{i2}^{(\lambda_2)} - \beta_{02} - \beta_{12}z_i \quad \text{e} \\ \epsilon_{i3} &= y_{i3}^{(\lambda_3)} - \beta_{03} - \beta_{13}z_i.\end{aligned}$$

As suposições de normalidade e independência foram consideradas, e ainda, $y_{ij}^{(\lambda_j)} \sim N(\beta_{0j} + \beta_{1j}z_i, \sigma_j^2)$, para $j = 1, 2, 3$ e $i = 1, \dots, 128$.

Além disso, podemos considerar $\alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i4}$ as estimativas das proporções dos componentes, em que $\alpha_{i1} + \alpha_{i2} + \alpha_{i3} + \alpha_{i4} = 1$, com $\alpha_{i1} > 0, \alpha_{i2} > 0, \alpha_{i3} > 0$ e $\alpha_{i4} > 0$, quando utiliza a transformação Box-Cox, obtendo a seguinte relação a partir de (4.11),

$$\begin{aligned}\frac{(\alpha_{i1}/\alpha_{i4})^{\lambda_1} - 1}{\lambda_1} &= \beta_{01} + \beta_{11}z_i, \\ \frac{(\alpha_{i2}/\alpha_{i4})^{\lambda_2} - 1}{\lambda_2} &= \beta_{02} + \beta_{12}z_i \quad \text{e} \\ \frac{(\alpha_{i3}/\alpha_{i4})^{\lambda_3} - 1}{\lambda_3} &= \beta_{03} + \beta_{13}z_i.\end{aligned} \tag{4.12}$$

Reescrevendo (4.12), temos que

$$\begin{aligned}\left(\frac{\alpha_{i1}}{\alpha_{i4}}\right)^{\lambda_1} &= \lambda_1(\beta_{01} + \beta_{11}z_i) + 1 \Rightarrow \left(\frac{\alpha_{i1}}{\alpha_{i4}}\right)^{\lambda_1 \cdot \frac{1}{\lambda_1}} = [\lambda_1(\beta_{01} + \beta_{11}z_i) + 1]^{\frac{1}{\lambda_1}}, \\ \left(\frac{\alpha_{i2}}{\alpha_{i4}}\right)^{\lambda_2} &= \lambda_2(\beta_{02} + \beta_{12}z_i) + 1 \Rightarrow \left(\frac{\alpha_{i2}}{\alpha_{i4}}\right)^{\lambda_2 \cdot \frac{1}{\lambda_2}} = [\lambda_2(\beta_{02} + \beta_{12}z_i) + 1]^{\frac{1}{\lambda_2}}, \\ \left(\frac{\alpha_{i3}}{\alpha_{i4}}\right)^{\lambda_3} &= \lambda_3(\beta_{03} + \beta_{13}z_i) + 1 \Rightarrow \left(\frac{\alpha_{i3}}{\alpha_{i4}}\right)^{\lambda_3 \cdot \frac{1}{\lambda_3}} = [\lambda_3(\beta_{03} + \beta_{13}z_i) + 1]^{\frac{1}{\lambda_3}},\end{aligned}$$

ou seja,

$$\begin{aligned}\alpha_{i1} &= \alpha_{i4} [\lambda_1(\beta_{01} + \beta_{11}z_i) + 1]^{\frac{1}{\lambda_1}}, \\ \alpha_{i2} &= \alpha_{i4} [\lambda_2(\beta_{02} + \beta_{12}z_i) + 1]^{\frac{1}{\lambda_2}} \quad \text{e} \\ \alpha_{i3} &= \alpha_{i4} [\lambda_3(\beta_{03} + \beta_{13}z_i) + 1]^{\frac{1}{\lambda_3}}.\end{aligned} \tag{4.13}$$

Sabemos que $\alpha_{i1} + \alpha_{i2} + \alpha_{i3} + \alpha_{i4} = 1$, ou seja, $\alpha_{i4} = 1 - (\alpha_{i1} + \alpha_{i2} + \alpha_{i3}) \Rightarrow$

$$\alpha_{i1} + \alpha_{i2} + \alpha_{i3} = 1 - \alpha_{i4}.$$

Assim, de (4.13), obtemos a seguinte relação

$$\alpha_{i1} + \alpha_{i2} + \alpha_{i3} = \alpha_{i4} \left\{ [\lambda_1(\beta_{01} + \beta_{11}z_i) + 1]^{\frac{1}{\lambda_1}} + [\lambda_2(\beta_{02} + \beta_{12}z_i) + 1]^{\frac{1}{\lambda_2}} + [\lambda_3(\beta_{03} + \beta_{13}z_i) + 1]^{\frac{1}{\lambda_3}} \right\},$$

Dessa forma, podemos reescrever a expressão $1 - \alpha_{i4}$ como

$$\begin{aligned} 1 - \alpha_{i4} &= \alpha_{i4} \left\{ [\lambda_1(\beta_{01} + \beta_{11}z_i) + 1]^{\frac{1}{\lambda_1}} + [\lambda_2(\beta_{02} + \beta_{12}z_i) + 1]^{\frac{1}{\lambda_2}} + [\lambda_3(\beta_{03} + \beta_{13}z_i) + 1]^{\frac{1}{\lambda_3}} \right\} \Rightarrow \\ 1 &= \alpha_{i4} \left\{ 1 + [\lambda_1(\beta_{01} + \beta_{11}z_i) + 1]^{\frac{1}{\lambda_1}} + [\lambda_2(\beta_{02} + \beta_{12}z_i) + 1]^{\frac{1}{\lambda_2}} + [\lambda_3(\beta_{03} + \beta_{13}z_i) + 1]^{\frac{1}{\lambda_3}} \right\} \Rightarrow \\ \alpha_{i4} &= \frac{1}{\left\{ 1 + [\lambda_1(\beta_{01} + \beta_{11}z_i) + 1]^{\frac{1}{\lambda_1}} + [\lambda_2(\beta_{02} + \beta_{12}z_i) + 1]^{\frac{1}{\lambda_2}} + [\lambda_3(\beta_{03} + \beta_{13}z_i) + 1]^{\frac{1}{\lambda_3}} \right\}}. \end{aligned} \quad (4.14)$$

Logo, substituindo (4.14) em (4.13), obtemos os estimadores das proporções verdadeiras dos componentes para a transformação Box-Cox,

$$\begin{aligned} \alpha_{ij} &= \frac{[\lambda_j(\beta_{0j} + \beta_{1j}z_i) + 1]^{\frac{1}{\lambda_j}}}{\left\{ 1 + [\lambda_1(\beta_{01} + \beta_{11}z_i) + 1]^{\frac{1}{\lambda_1}} + [\lambda_2(\beta_{02} + \beta_{12}z_i) + 1]^{\frac{1}{\lambda_2}} + [\lambda_3(\beta_{03} + \beta_{13}z_i) + 1]^{\frac{1}{\lambda_3}} \right\}} \quad \text{e} \\ \alpha_{i4} &= \frac{1}{\left\{ 1 + [\lambda_1(\beta_{01} + \beta_{11}z_i) + 1]^{\frac{1}{\lambda_1}} + [\lambda_2(\beta_{02} + \beta_{12}z_i) + 1]^{\frac{1}{\lambda_2}} + [\lambda_3(\beta_{03} + \beta_{13}z_i) + 1]^{\frac{1}{\lambda_3}} \right\}}, \end{aligned} \quad (4.15)$$

em que $i = 1, \dots, 128$ e $j = 1, 2, 3$.

4.2.1 Análise Bayesiana - Modelo 2

A função densidade de probabilidade conjunta de $\mathbf{y}_1^{(\lambda_1)}, \mathbf{y}_2^{(\lambda_2)}, \mathbf{y}_3^{(\lambda_3)}$ dado o vetor de parâmetros $\mathbf{v}_2 = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{11}, \beta_{12}, \beta_{13}, \sigma_1^2, \sigma_2^2, \sigma_3^2, \lambda_1, \lambda_2, \lambda_3)$ é dada por

$$f(\mathbf{y}_1^{(\lambda_1)}, \mathbf{y}_2^{(\lambda_2)}, \mathbf{y}_3^{(\lambda_3)} | \mathbf{v}_2) = \prod_{j=1}^3 \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left[-\frac{1}{2\sigma_j^2} \left(y_{ij}^{(\lambda_j)} - \beta_{0j} - \beta_{1j}z_i \right)^2 \right].$$

Por outro lado, a função densidade de probabilidade das observações não transformadas (variáveis originais) [Box & Tiao [9]], $y_{i1} = x_{i1}/x_{i4}$, $y_{i2} = x_{i2}/x_{i4}$ e $y_{i3} = x_{i3}/x_{i4}$ é dada por

$$f(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 | \mathbf{v}_2) = \prod_{j=1}^3 \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp \left[-\frac{1}{2\sigma_j^2} \left(y_{ij}^{(\lambda_j)} - \beta_{0j} - \beta_{1j}z_i \right)^2 \right] \left(\frac{dy_{ij}^{(\lambda_j)}}{dy_{ij}} \right),$$

em que $\frac{dy_{ij}^{(\lambda_j)}}{dy_{ij}} = \frac{\lambda_j y_{ij}^{\lambda_j-1}}{\lambda_j} = y_{ij}^{\lambda_j-1}$ são os elementos da diagonal da matriz Jacobiano, para $j = 1, 2, 3$.

Assim, assumindo o modelo (4.11), a função de verossimilhança para os parâmetros $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})$, $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})$, $\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ e $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ é dada por

$$L(\beta_0, \beta_1, \sigma^2, \lambda) = \prod_{j=1}^3 \left[(\sigma_j^2)^{-n/2} \left(\prod_{i=1}^n y_{ij}^{\lambda_j-1} \right) \exp \left(-\frac{1}{2\sigma_j^2} \sum_{i=1}^n \epsilon_{ij}^2 \right) \right], \quad (4.16)$$

em que $\left(\prod_{i=1}^n y_{ij}^{\lambda_j-1} \right)$ são os produtos do Jacobiano, e

$$\epsilon_{ij} = y_{ij}^{(\lambda_j)} - \beta_{0j} - \beta_{1j}z_i \quad \text{em que } y_{ij}^{(\lambda_j)} = \frac{(x_{ij}/x_{i4})^{\lambda_j} - 1}{\lambda_j} \quad \text{se } \lambda_j \neq 0, \text{ para } j = 1, 2, 3.$$

Para a estimação dos parâmetros do modelo (4.11), as seguintes distribuições *a priori* para $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})$, $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})$, $\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ e $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ são consideradas,

$$\begin{aligned} \beta_{0j} &\sim N(a_{0j}, b_{0j}^2), \\ \beta_{1j} &\sim N(a_{1j}, b_{1j}^2), \\ \sigma_j^2 &\sim IG(c_j, d_j) \quad \text{e} \\ \lambda_j &\sim N(e_j, f_j^2), \end{aligned} \quad (4.17)$$

em que $N(a, b^2)$ denota uma distribuição Normal com média a e variância b^2 , $IG(c, d)$ denota uma distribuição Gama inversa com média $d/(c-1)$ e variância $d^2/[(c-1)^2(c-2)]$, $c > 2$; e $a_{0j}, b_{0j}, a_{1j}, b_{1j}, c_j, d_j, e_j$ e f_j são hiperparâmetros conhecidos, para $j = 1, 2, 3$. Para este caso, também assumimos independência *a priori* entre os parâmetros.

Desta forma, a partir do teorema de Bayes que associa a função de verossimilhança (4.16) com as distribuições *a priori* (4.17), obtemos a distribuição *a posteriori* conjunta para os parâmetros $\beta_0 = (\beta_{01}, \beta_{02}, \beta_{03})$, $\beta_1 = (\beta_{11}, \beta_{12}, \beta_{13})$, $\sigma^2 = (\sigma_1^2, \sigma_2^2, \sigma_3^2)$ e $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ que é dada por

$$\begin{aligned} \pi(\beta_0, \beta_1, \sigma^2, \lambda | \mathbf{y}^{(\lambda)}) &\propto \prod_{j=1}^3 \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \times \prod_{j=1}^3 \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \\ &\times \prod_{j=1}^3 (\sigma_j^2)^{-(c_j+1)} \exp \left(-\frac{d_j}{\sigma_j^2} \right) \times \prod_{j=1}^3 \exp \left[-\frac{1}{2f_j^2} (\lambda_j - e_j)^2 \right] \\ &\times \left(\prod_{i=1}^n y_{i1}^{\lambda_1-1} \right) \left(\prod_{i=1}^n y_{i2}^{\lambda_2-1} \right) \left(\prod_{i=1}^n y_{i3}^{\lambda_3-1} \right) \prod_{j=1}^3 (\sigma_j^2)^{-n/2} \exp \left(-\frac{1}{2\sigma_j^2} \sum_{i=1}^n \epsilon_{ij}^2 \right). \end{aligned} \quad (4.18)$$

Considerando a distribuição *a posteriori* conjunta (4.18), as distribuições *a posteriori* condicionais utilizadas para o amostrador de Gibbs (Gelfand e Smith [20]) e para o algoritmo de Metropolis-Hastings são dadas por:

$$i) \quad \pi(\beta_{0j} | \beta_1, \sigma^2, \lambda, \mathbf{y}^{(\lambda)}) \propto \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \times \exp \left[-\frac{1}{2\sigma_j^2} \sum_{i=1}^n (\beta_{0j} - \mu_i^{(j)})^2 \right],$$

em que $\mu_i^{(j)} = y_{ij}^{(\lambda_j)} - \beta_{1j} z_i$; $i = 1, \dots, n$ e $j = 1, 2, 3$.

Logo,

$$\pi(\beta_{0j} | \beta_1, \sigma^2, \lambda, \mathbf{y}^{(\lambda)}) \sim N \left[\frac{a_{0j}\sigma_j^2 + b_{0j} \sum_{i=1}^n \mu_i^{(j)}}{\sigma_j^2 + nb_{0j}^2}, \frac{b_{0j}^2 \sigma_j^2}{\sigma_j^2 + nb_{0j}^2} \right]. \quad (4.19)$$

$$ii) \quad \pi(\beta_{1j} | \beta_0, \sigma^2, \lambda, \mathbf{y}^{(\lambda)}) \propto \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \times \exp \left[-\frac{1}{2\sigma_j^2} \sum_{i=1}^n (\beta_{1j} z_i - \theta_i^{(j)})^2 \right],$$

em que $\theta_i^{(j)} = y_{ij}^{(\lambda_j)} - \beta_{0j}$; $i = 1, \dots, n$ e $j = 1, 2, 3$.

Logo,

$$\pi(\beta_{1j}|\beta_0, \sigma^2, \lambda, \mathbf{y}^{(\lambda)}) \sim N \left[\frac{a_{1j}\sigma_j^2 + b_{1j} \sum_{i=1}^n z_i \theta_i^{(j)}}{\sigma_j^2 + b_{1j}^2 \sum_{i=1}^n z_i^2}, \frac{b_{1j}^2 \sigma_j^2}{\sigma_j^2 + b_{1j}^2 \sum_{i=1}^n z_i^2} \right]. \quad (4.20)$$

$$\begin{aligned} iii) \quad \pi(\sigma_j^2|\beta_0, \beta_1, \lambda, \mathbf{y}^{(\lambda)}) &\propto (\sigma_j^2)^{-(c_j+1)} \exp\left(-\frac{d_j}{\sigma_j^2}\right) \times (\sigma_j^2)^{-n/2} \times \exp\left(-\frac{1}{2\sigma_j^2} \sum_{i=1}^n \epsilon_{ij}^2\right) \\ &\propto (\sigma_j^2)^{-(c_j+n/2+1)} \exp\left[-\frac{1}{\sigma_j^2} \left(d_j + \frac{1}{2} \sum_{i=1}^n \epsilon_{ij}^2\right)\right], \end{aligned}$$

em que $\epsilon_{ij} = y_{ij}^{(\lambda_j)} - \beta_{0j} - \beta_{1j}z_i$; $i = 1, \dots, n$ e $j = 1, 2, 3$.

Logo,

$$\pi(\sigma_j^2|\beta_0, \beta_1, \lambda, \mathbf{y}^{(\lambda)}) \sim IG \left[c_j + \frac{n}{2}, d_j + \frac{1}{2} \sum_{i=1}^n \epsilon_{ij}^2 \right]. \quad (4.21)$$

$$iv) \quad \pi(\lambda_j|\beta_0, \beta_1, \sigma^2, \mathbf{y}^{(\lambda)}) \propto \exp \left[-\frac{1}{2f_j^2} (\lambda_j - e_j)^2 \right] \psi_j(\beta_0, \beta_1, \sigma^2, \lambda), \quad (4.22)$$

em que

$$\begin{aligned} \psi_j(\beta_0, \beta_1, \sigma^2, \lambda) &= \exp \left[(\lambda_j - 1) \sum_{i=1}^n \ln y_{ij} + \frac{1}{\sigma_j^2} \sum_{i=1}^n \left(y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j}) \right)^2 \right] \\ &= \exp \left[(\lambda_j - 1) \sum_{i=1}^n \ln y_{ij} + \frac{1}{\sigma_j^2} \left(\sum_{i=1}^n y_{ij}^{(\lambda_j)^2} - 2\beta_{0j} \sum_{i=1}^n y_{ij}^{(\lambda_j)} - 2\beta_{1j} \sum_{i=1}^n y_{ij}^{(\lambda_j)} z_i \right) \right] \end{aligned}$$

e $j = 1, 2, 3$.

Podemos observar que as distribuições *a posteriori* condicionais (4.19), (4.20) e (4.21) apresentaram formas fechadas e, portanto, obtemos as estimativas dos parâmetros do modelo (4.11) através do algoritmo amostrador de Gibbs. Por outro lado, para a distribuição *a posteriori* condicional (4.22) utilizamos o algoritmo de Metropolis-Hastings, por não apresentar forma fechada.

Assim, considerando a transformação Box-Cox com erros normais não correlacionados para o modelo (4.11), assumimos as distribuições *a priori* não informativas para

$\beta_{0j}, \beta_{1j}, \sigma_j^2, \lambda_j$ com hiperparâmetros cujos valores foram $a_{0j} = a_{1j} = e_j = 0, b_{0j} = b_{1j} = 1.000, c_j = 0,01, d_j = 0,01$ e $f_j = 100, j = 1, 2, 3$, geramos no *software* OPENBUGS 210.000 iterações, em que foram descartadas as 10.000 iterações iniciais (“burn-in”). Após o período de “burn-in”, consideramos saltos de tamanho 20 para obtermos amostras não correlacionadas, na qual resultou uma amostra final de tamanho 10.000 para cada parâmetro. A convergência do algoritmo foi verificada pelos gráficos dos traços *a posteriori* das amostras de Gibbs simuladas e pelo diagnóstico de Geweke implementado no *software* R através do pacote *CODA*.

Na Tabela 4.3 apresentamos os sumários *a posteriori* dos parâmetros do modelo (4.11) e os valores correspondentes da estatística de teste para o diagnóstico de convergência Geweke.

Tabela 4.3: Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke - Modelo 2.

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	0,632	0,068	(0,515; 0,780)	0,716
β_{02}	-1,014	0,128	(-1,297; -0,798)	0,210
β_{03}	-1,830	0,266	(-2,428; -1,380)	-1,449
β_{11}	-0,034	0,050	(-0,136; 0,065)	-1,118
β_{12}	0,006	0,072	(-0,135; 0,153)	0,327
β_{13}	0,133	0,103	(-0,044; 0,363)	-0,304
σ_1	0,078	0,026	(0,040; 0,143)	0,434
σ_2	0,164	0,073	(0,067; 0,349)	0,533
σ_3	0,275	0,157	(0,088; 0,680)	1,241
λ_1	0,043	0,244	(-0,441; 0,523)	0,570
λ_2	0,113	0,183	(-0,246; 0,466)	-0,180
λ_3	0,089	0,131	(-0,167; 0,350)	-1,706

Os resultados da Tabela 4.3 apresentam os sumários *a posteriori* dos parâmetros estimados do modelo (4.11). Podemos verificar que não houve efeito significativo para nenhuma proporção de pontos de ataque, bloqueio e saque (o intervalo de credibilidade inclui o zero nos parâmetros β_{11}, β_{12} e β_{13}). Também, os intervalos de credibilidade para os parâmetros $\lambda_j, j = 1, 2, 3$ está incluído o valor zero, ou seja, como a aplicação da transformação Box-Cox é um caso especial da transformação *alr* quando $\lambda = 0$, indica que a transformação *alr* é mais adequada para esses dados do que a transformação Box-Cox.

A Tabela 4.4 apresenta as estimativas das proporções verdadeiras para os componentes ataque, bloqueio, saque e erros do adversário para cada valor da covariável z_i considerando a transformação Box-Cox. Da mesma forma que na transformação *alr*, não houve diferença

em relação aos times terem vencido ou não a Superliga de vôlei para as estimativas das proporções para todos os componentes (ataque, bloqueio, saque e erros do adversário).

Tabela 4.4: Estimativas das proporções dos componentes - Modelo 2.

Proporções	$z = 0$	$z = 1$
α_1	0,558	0,546
α_2	0,102	0,104
α_3	0,041	0,048
α_4	0,299	0,302

As Figuras 4.5, 4.6, 4.7 e 4.8 apresentam as densidades *a posteriori* e o comportamento da cadeia gerada para os parâmetros do modelo (4.11). Podemos observar que os traços *a posteriori* apresentaram certa uniformidade, indicando possível convergência, sendo confirmado através do diagnóstico de Geweke (Tabela 4.3) em que para todos os parâmetros os valores do teste z_G encontram-se no intervalo $(-1,96; 1,96)$.

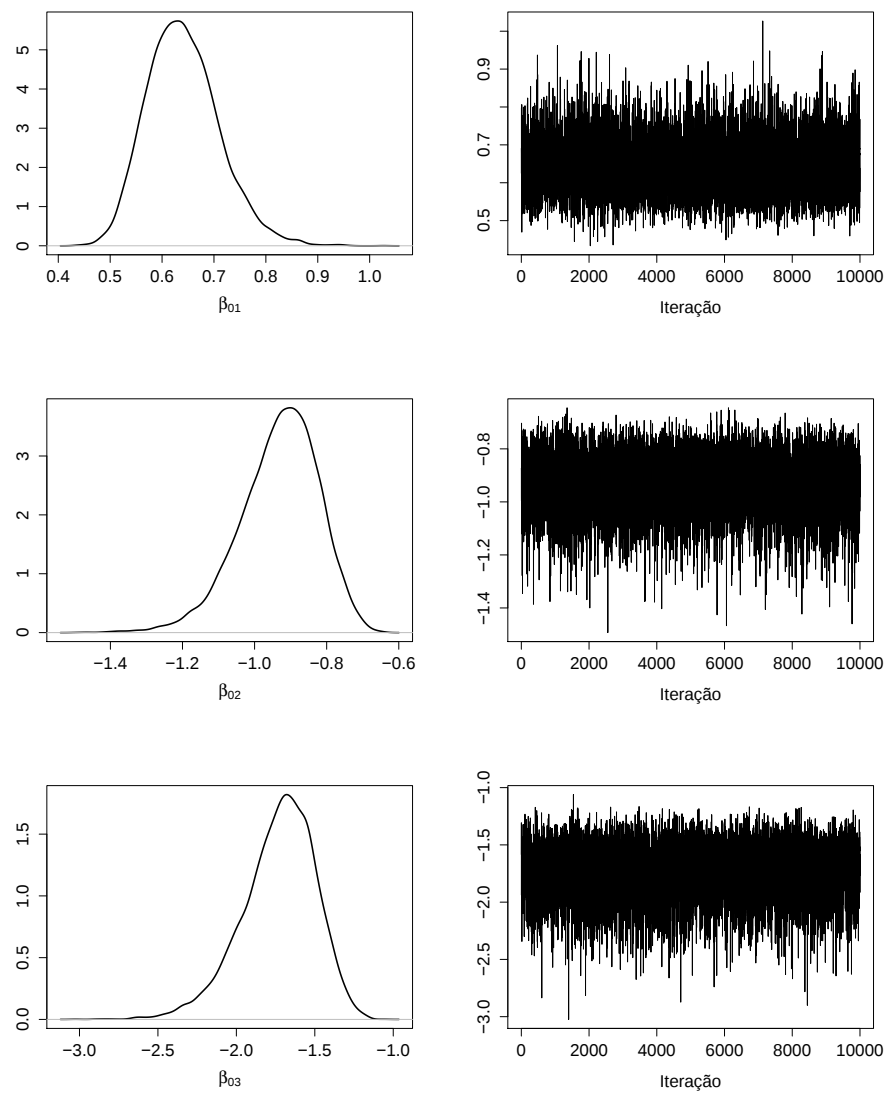


Figura 4.5: Densidades *a posteriori* e traços *a posteriori* para os parâmetros de intercepto do modelo 2.

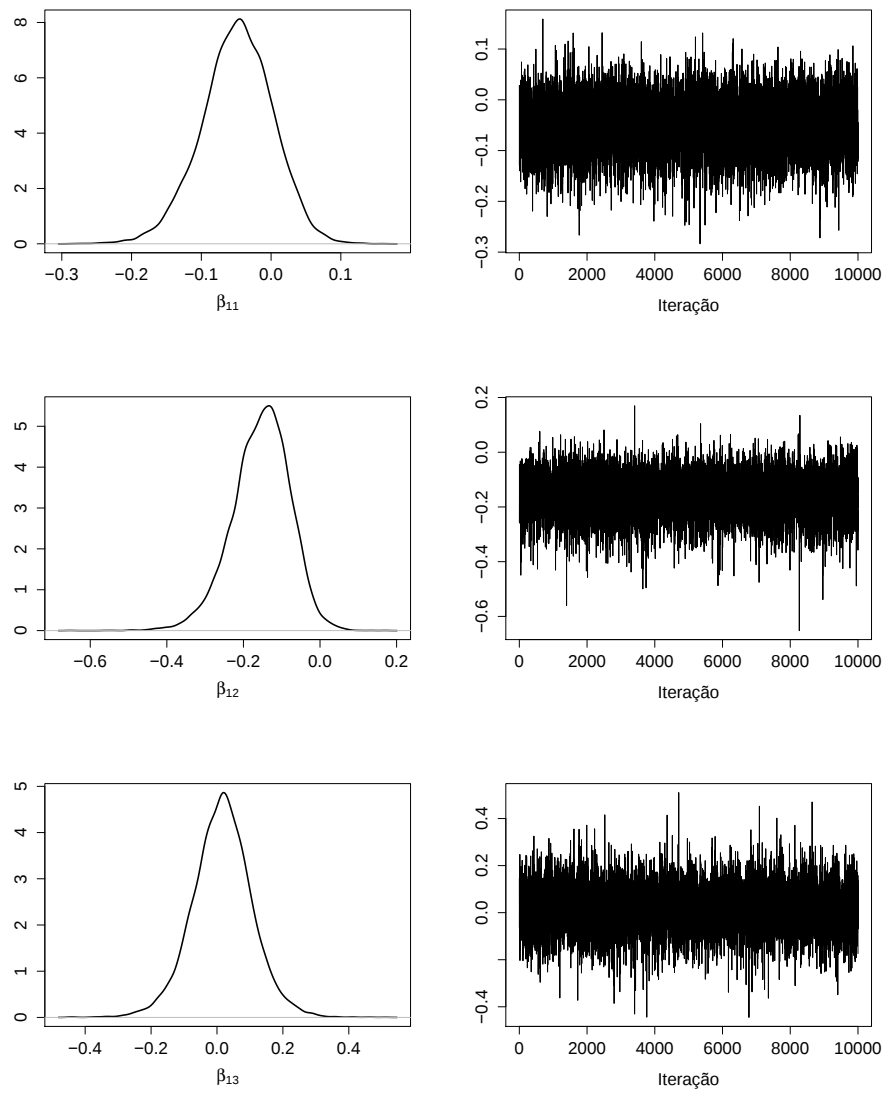


Figura 4.6: Densidades *a posteriori* e traços *a posteriori* para os parâmetros de inclinação do modelo 2.

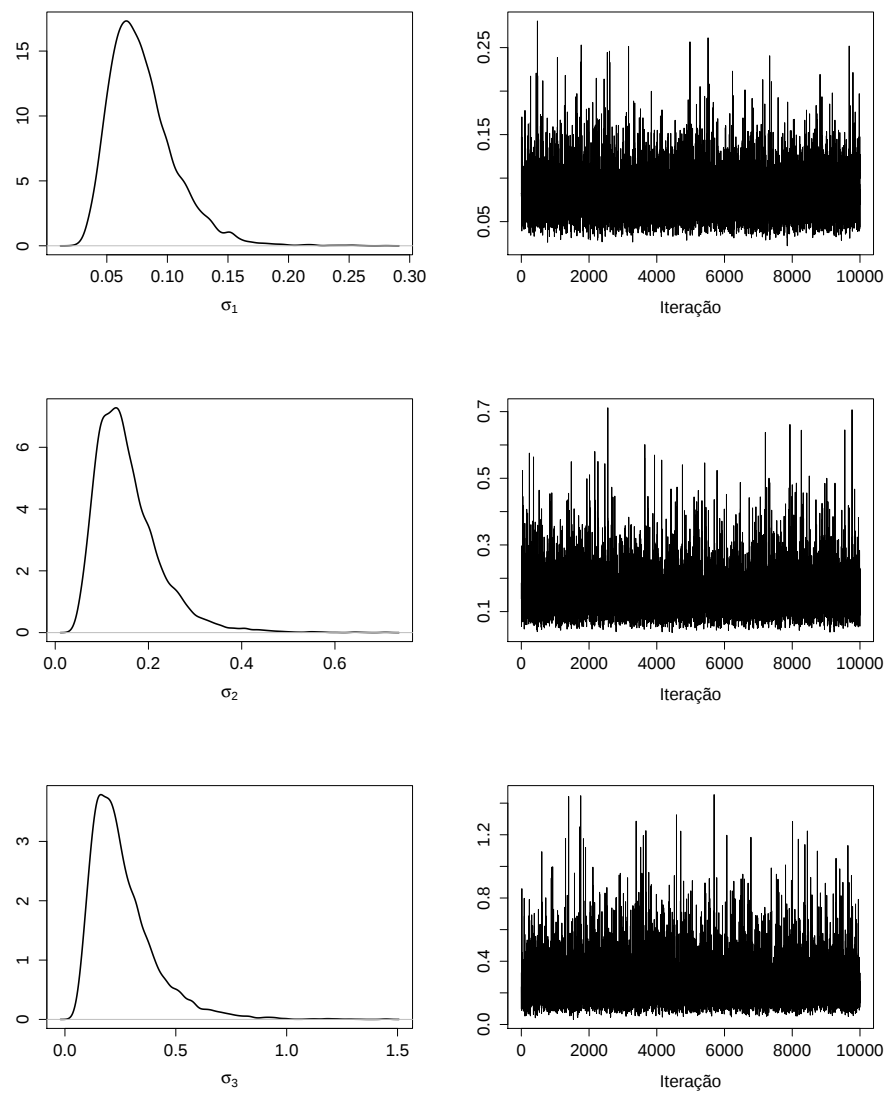


Figura 4.7: Densidades *a posteriori* e traços *a posteriori* para os desvios padrões do modelo 2.

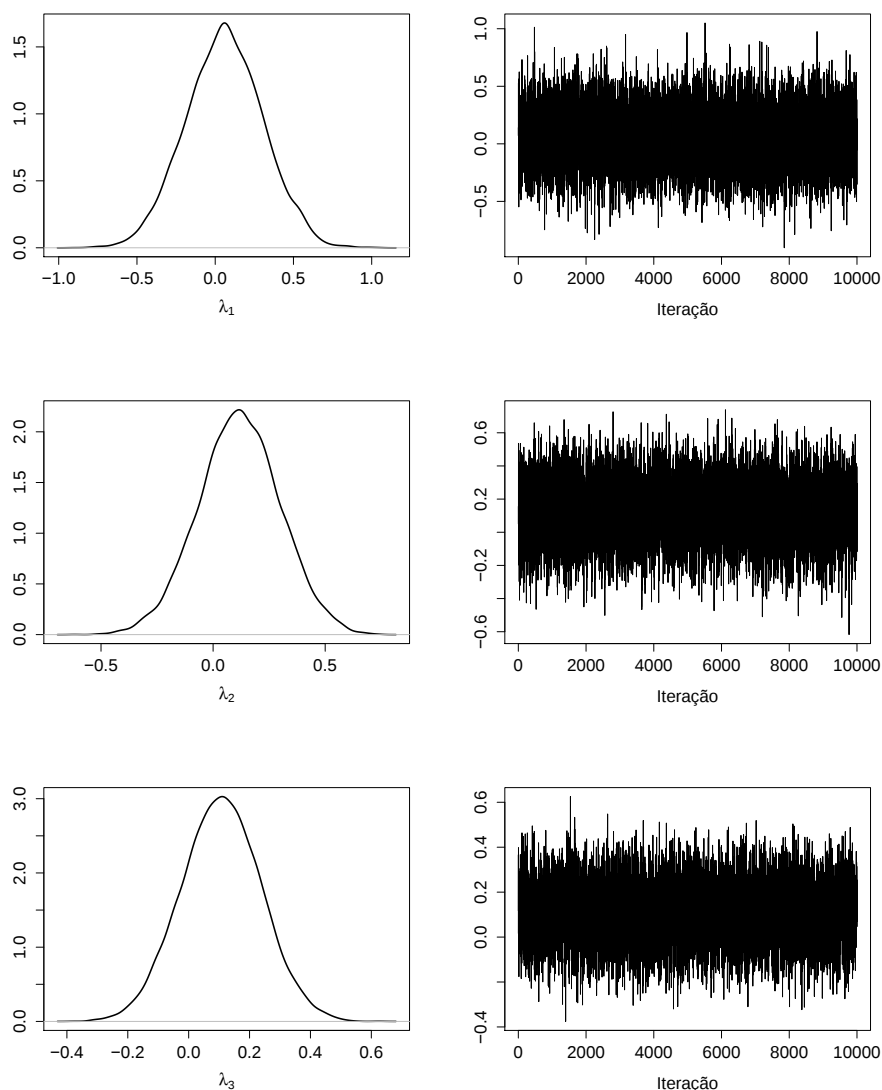


Figura 4.8: Densidades *a posteriori* e traços *a posteriori* para os parâmetros λ do modelo 2.

4.3 Análise clássica

Com o objetivo de comparação, os dados composicionais do Apêndice A foram também analisados através de métodos clássicos aplicando a transformação *alr* em que utilizou-se a rotina *PROC NLMIXED* do *software* SAS para a estimação dos parâmetros do modelo (4.1).

As estimativas dos parâmetros e seus respectivos intervalos de confiança são apresentados na Tabela 4.5.

Tabela 4.5: Estimativas dos parâmetros do modelo 1 - Método Clássico

Parâmetro	Estimativa	Desvio Padrão	Intervalo de Confiança (95%)
β_{01}	0,270	0,014	(0,243; 0,298)
β_{02}	-0,429	0,023	(-0,474; -0,385)
β_{03}	-0,827	0,031	(-0,888; -0,765)
β_{11}	-0,020	0,020	(-0,059; 0,020)
β_{12}	-0,074	0,032	(-0,138; -0,010)
β_{13}	0,006	0,044	(-0,081; 0,093)
σ_1	0,113	0,007	(0,099; 0,127)
σ_2	0,183	0,011	(0,161; 0,206)
σ_3	0,251	0,016	(0,220; 0,282)

Como foram utilizadas distribuições *a priori* não-informativas para a modelagem Bayesiana, percebe-se que as estimativas da Tabela 4.5 apresentaram resultados semelhantes às estimativas Bayesianas (Tabela 4.1), ou seja, houve efeito significativo do time que venceu pelo menos uma vez a Superliga de vôlei (z) em relação à proporção de pontos de bloqueio (β_{12}). Porém, a sua estimativa é negativa, indicando que o time que não venceu nenhuma vez a Superliga fez mais pontos de bloqueio no jogo.

A vantagem do método Bayesiano deve-se ao fato da possibilidade de inserir informações adicionais, representadas pela distribuição *a priori* informativa, juntamente com a verossimilhança para a estimação dos parâmetros. Por outro lado, os métodos clássicos necessitam de resultados assintóticos e algumas vezes sem boa precisão, devido à situações com amostras pequenas.

Prosseguindo com a análise dos dados da Superliga de vôlei, ajustamos um modelo de regressão Dirichlet sob o enfoque clássico em que as estimativas de máxima verossimilhança para os parâmetros do modelo proposto (3.4) foram obtidas no *software* R, através do pacote *DirichletReg*. Os resultados da estimação dos parâmetros do modelo de regressão Dirichlet estão dispostos na Tabela 4.6.

Na Tabela 4.6, podemos verificar que nenhum parâmetro do modelo de regressão Dirichlet foi significativo em relação à covariável do time ter ganho pelo menos uma vez a Superliga de vôlei, com exceção dos interceptos, que apresentaram estimativas positivas, diferente dos outros modelos ajustados anteriormente.

Tabela 4.6: Estimativas dos parâmetros do modelo de Regressão Dirichlet

Parâmetro	Estimativa	Desvio Padrão	Intervalo de Confiança (95%)
β_{01}	3,752	0,103	(3,550; 3,950)
β_{02}	2,186	0,105	(1,980; 2,392)
β_{03}	1,350	0,108	(1,139; 1,561)
β_{04}	3,140	0,104	(2,937; 3,343)
β_{11}	-0,068	0,147	(-0,356; 0,220)
β_{12}	-0,183	0,150	(-0,477; 0,111)
β_{13}	-0,009	0,153	(-0,309; 0,292)
β_{14}	-0,023	0,148	(-0,313; 0,266)

4.4 Discussão dos Resultados

Para a seleção de modelos, pode-se utilizar alguns critérios de discriminação de modelos, sendo que destacamos o critério DIC (Spiegelhalter et al., [46]). Os valores de DIC foram obtidos através do *software* OPENBUGS e menores valores do DIC indicam o modelo mais adequado. Na tabela 4.7, temos os valores de DIC para os modelos 1 (transformação *alr*) e 2 (transformação Box-Cox), modelos que abordaram os métodos Bayesianos.

Tabela 4.7: Critério DIC - Modelos 1 e 2.

Modelo	DIC
Modelo 1	402,90
Modelo 2	-197,70

Tabela 4.8: Soma do quadrado das diferenças entre os valores observados e os valores ajustados.

Modelo	$\sum_{i=1}^n \sum_{j=1}^3 (y_{ij} - \hat{y}_{ij})^2$
Modelo 1 (transformação <i>alr</i> - Bayesiano)	0,995
Modelo 2 (transformação Box-Cox - Bayesiano)	1,014
Modelo (transformação <i>alr</i> - Análise Clássica)	5,741
Modelo de Regressão Dirichlet clássico	0,983

Podemos observar na Tabela 4.7 que o melhor ajuste dos dados da Superliga de vôlei foi para o modelo 2 (transformação Box-Cox), que apresentou menor valor de DIC. Porém, vale ressaltar que os intervalos de credibilidade de λ_1 , λ_2 , λ_3 incluem o valor zero, indicando que para esse conjunto de dados a transformação *alr* seria a mais apropriada do que a transformação Box-Cox.

Além disso, calculou-se a soma do quadrado das diferenças entre os valores observados e os valores ajustados dos modelos e são apresentados na Tabela 4.8. Analisando os resultados obtidos na Tabela 4.8, percebe-se que o modelo de regressão Dirichlet apresentou menor valor da soma do quadrado das diferenças entre os valores observados e os valores ajustados, ou seja, é o melhor que se ajusta aos dados. Em termos de predição, apesar da estimativa do DIC ser maior para o modelo 1 (transformação *alr*), observa-se que seu valor na Tabela 4.8 foi muito similar ao valor do modelo de regressão Dirichlet.

A Figura 4.9 apresenta as estimativas de cada componente do vôlei para cada valor da covariável associada aos modelos ajustados.

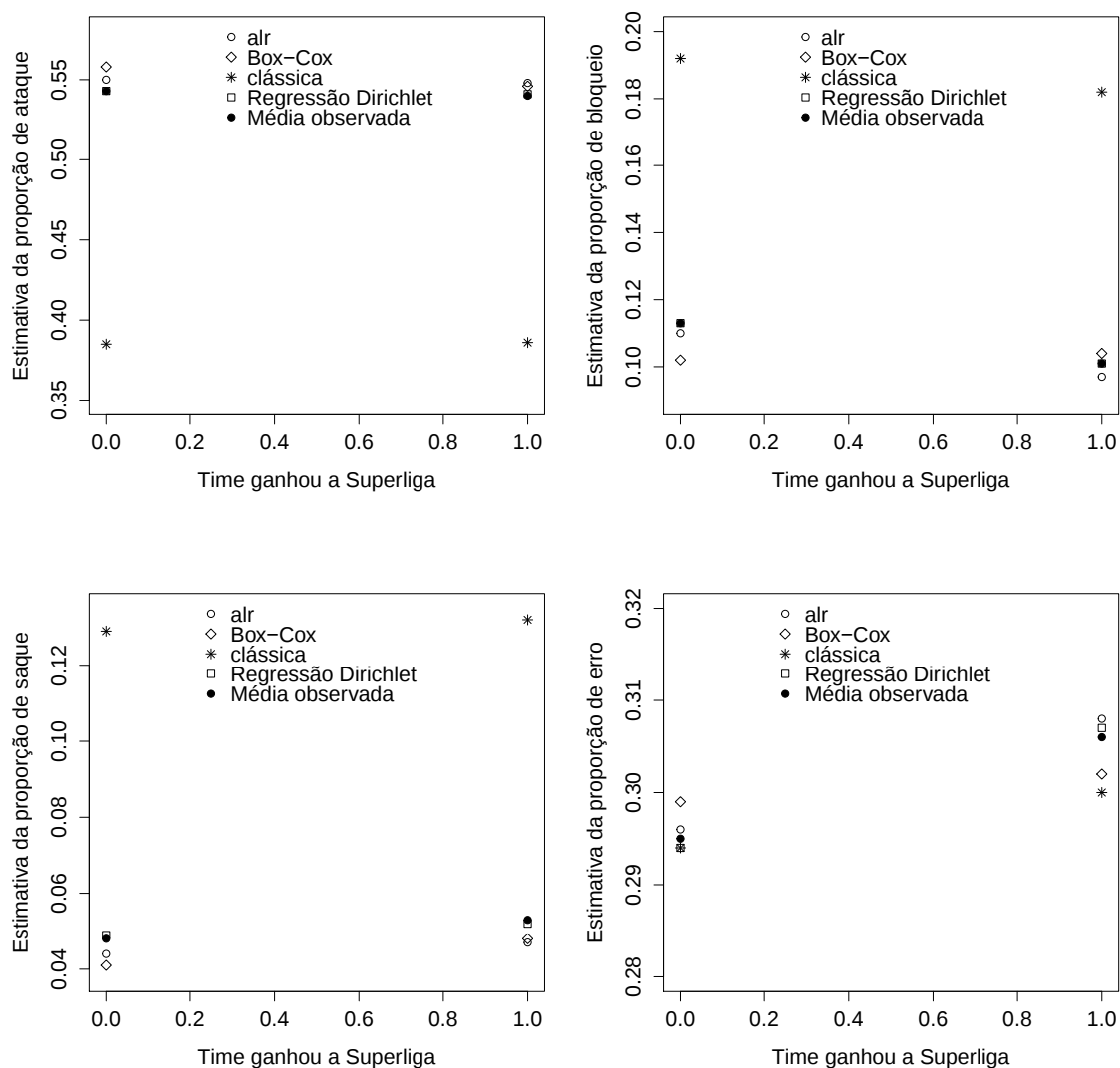


Figura 4.9: Gráficos dos valores observados e ajustados dos componentes do vôlei de acordo com o modelo

Observa-se que para todos os componentes, as estimativas das proporções dos componentes do método clássico está mais distante da média observada.

5 Análise Bayesiana para Dados Longitudinais

O segundo conjunto de dados (Apêndice B) é referente a uma tentativa de melhorar a qualidade do leite de vaca, nas quais trinta vacas foram avaliadas antes e depois de uma dieta rigorosamente controlada durante um período de oito semanas. Apesar de variações sazonais na qualidade do leite terem sido insignificantes neste período, decidiu-se por um grupo controle de trinta vacas nas mesmas condições, mas em um regime regular estabelecido.

O propósito do experimento foi determinar se o novo regime produziu alguma mudança significativa na composição do leite, que é composto por seis componentes, sendo que denotamos por: x_{i1} a proporção de proteína (pr), x_{i2} a proporção da gordura do leite (mf), x_{i3} a proporção de carboidrato (ch), x_{i4} a proporção de cálcio (Ca), x_{i5} a proporção de sódio (Na) e x_{i6} a proporção de potássio (K).

A covariável associada à i -ésima vaca é definida por,

$$z_i = \begin{cases} 0, & \text{pertence ao grupo 1 (antes da dieta)} \\ 1, & \text{pertence ao grupo 2 (depois da dieta)}. \end{cases}$$

Estes dados foram introduzidos em Aitchison [4](pág. 73 e 74).

Como uma análise preliminar, verificamos se há uma indicação de diferenças para as respostas entre os dois grupos (antes e depois). Na Figura 5.1 apresentamos os box-plots dos dados originais denotando grupo 1 por 0 e grupo 2 por 1.

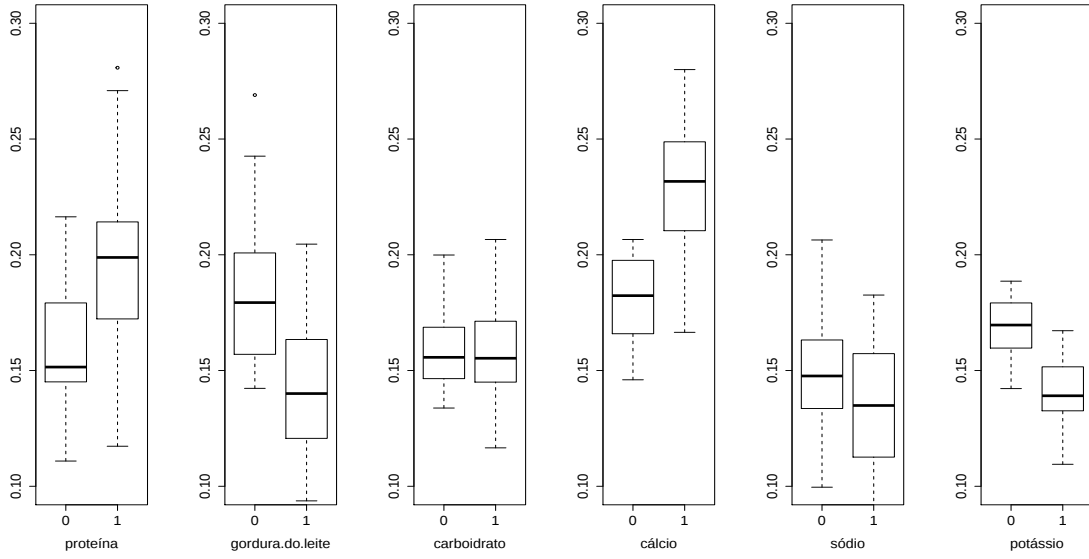


Figura 5.1: Box plots dos componentes do leite para os grupos antes e depois da dieta.

Analisando a Figura 5.1, temos uma indicação preliminar que há diferenças nos componentes proteína x_{i1} , gordura x_{i2} , cálcio x_{i4} e potássio x_{i6} em relação aos dois grupos.

5.1 Transformação *alr* Considerando um Efeito Aleatório - Modelo 3

Para a análise dos dados composicionais do Apêndice B, x_{i1} , x_{i2} , x_{i3} , x_{i4} , x_{i5} e x_{i6} para $i = 1, \dots, 60$, consideramos a transformação *alr* apresentada em (3.2). Com isso, obtemos

$$y_{i1} = \log\left(\frac{x_{i1}}{x_{i6}}\right), y_{i2} = \log\left(\frac{x_{i2}}{x_{i6}}\right), y_{i3} = \log\left(\frac{x_{i3}}{x_{i6}}\right), y_{i4} = \log\left(\frac{x_{i4}}{x_{i6}}\right), y_{i5} = \log\left(\frac{x_{i5}}{x_{i6}}\right).$$

O modelo de regressão para os dados longitudinais transformados y_{ij} , para $j = 1, \dots, 5$, denominado por modelo 3, é dado por

$$\mathbf{y}_i = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_i + w_i + \boldsymbol{\epsilon}_i, \quad (5.1)$$

em que $\mathbf{y}_i$ é o vetor de variáveis resposta, para $j = 1, \dots, 5$ e $i = 1, \dots, 60$ ($n = n_1 + n_2$, em que $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2), z_i representa a covariável

associada ao i -ésimo indivíduo (variável indicadora em que $z_i = 0$ para o grupo 1 e $z_i = 1$ para o grupo 2), w_i é um efeito aleatório (variável latente não observável) que captura a dependência entre as medidas repetidas para cada indivíduo assumindo $w_i \sim N(0, \sigma_w^2)$ e ϵ_i é o vetor de erros. Assumiremos $\epsilon_i \sim N(\mathbf{0}, \Sigma)$, em que $\mathbf{0}$ é o vetor de zeros e Σ é a matriz de variâncias e covariâncias especificada por

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_5^2 \end{pmatrix}.$$

Considerando o modelo apresentado em (5.1), assumiremos que os erros ϵ_i são mutuamente independentes. Desta forma, obtém-se

$$\begin{aligned} \epsilon_{i1} &= y_{i1} - \beta_{01} - \beta_{11}z_i - w_i, \\ &\vdots \\ \epsilon_{i5} &= y_{i5} - \beta_{05} - \beta_{15}z_i - w_i. \end{aligned}$$

Além disso, podemos estimar as proporções verdadeiras dos componentes do leite $\alpha_{i1}, \dots, \alpha_{i6}$, em que $\alpha_{i1} + \dots + \alpha_{i6} = 1$, com $\alpha_{i1} > 0, \dots, \alpha_{i6} > 0$ e a transformação *alr*, obtendo a seguinte relação a partir de (5.1),

$$\begin{aligned} \log \left(\frac{\alpha_{i1}}{\alpha_{i6}} \right) &= \beta_{01} + \beta_{11}z_i + w_i, \\ &\vdots \\ \log \left(\frac{\alpha_{i5}}{\alpha_{i6}} \right) &= \beta_{05} + \beta_{15}z_i + w_i. \end{aligned} \tag{5.2}$$

Aplicando a função exponencial em (5.2), temos que

$$\frac{\alpha_{i1}}{\alpha_{i6}} = e^{\beta_{01} + \beta_{11}z_i + w_i}; \dots; \frac{\alpha_{i5}}{\alpha_{i6}} = e^{\beta_{05} + \beta_{15}z_i + w_i}.$$

Assim,

$$\begin{cases} \alpha_{i1} = \alpha_{i6}(e^{\beta_{01}+\beta_{11}z_i+w_i}), \\ \vdots \\ \alpha_{i5} = \alpha_{i6}(e^{\beta_{05}+\beta_{15}z_i+w_i}). \end{cases} \quad (5.3)$$

Sabemos que $\alpha_{i1} + \dots + \alpha_{i6} = 1$, ou seja,

$$\alpha_{i6} = 1 - (\alpha_{i1} + \dots + \alpha_{i5}) \Rightarrow \alpha_{i1} + \dots + \alpha_{i5} = 1 - \alpha_{i6}.$$

De (5.3), obtemos a seguinte relação,

$$\alpha_{i1} + \dots + \alpha_{i5} = \alpha_{i6} (e^{\beta_{01}+\beta_{11}z_i+w_i} + \dots + e^{\beta_{05}+\beta_{15}z_i+w_i}),$$

ou seja,

$$\begin{aligned} 1 - \alpha_{i6} &= \alpha_{i6} (e^{\beta_{01}+\beta_{11}z_i+w_i} + \dots + e^{\beta_{05}+\beta_{15}z_i+w_i}) \Rightarrow \\ 1 &= \alpha_{i6} (1 + e^{\beta_{01}+\beta_{11}z_i+w_i} + \dots + e^{\beta_{05}+\beta_{15}z_i+w_i}) \Rightarrow \\ \alpha_{i6} &= \frac{1}{(1 + e^{\beta_{01}+\beta_{11}z_i+w_i} + \dots + e^{\beta_{05}+\beta_{15}z_i+w_i})}. \end{aligned} \quad (5.4)$$

Logo, substituindo (5.4) em (5.3), obtemos os estimadores das proporções verdadeiras dos componentes do grupo 1 (antes da dieta) quando $z_i = 0$ e do grupo 2 (depois da dieta) quando $z_i = 1$, para a transformação *alr*,

$$\begin{aligned} \alpha_{ij} &= \frac{e^{\beta_{0j}+\beta_{1j}z_i+w_i}}{(1 + e^{\beta_{01}+\beta_{11}z_i+w_i} + e^{\beta_{02}+\beta_{12}z_i+w_i} + \dots + e^{\beta_{05}+\beta_{15}z_i+w_i})}, \\ &\vdots \\ \alpha_{i6} &= \frac{1}{(1 + e^{\beta_{01}+\beta_{11}z_i+w_i} + e^{\beta_{02}+\beta_{12}z_i+w_i} + \dots + e^{\beta_{05}+\beta_{15}z_i+w_i})}, \end{aligned} \quad (5.5)$$

em que $i = 1, \dots, 60$ ($n = n_1 + n_2$, onde $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2) e $j = 1, \dots, 5$.

A parametrização $\alpha_{i1} > 0, \dots, \alpha_{i6} > 0$ pode ser usada para obter inferências sobre as composições em cada valor da covariável.

5.1.1 Análise Bayesiana - Modelo 3

A função densidade de probabilidade conjunta de $\mathbf{y}_1, \dots, \mathbf{y}_5$ dado o vetor de parâmetros $\mathbf{v}_3 = (\beta_{01}, \dots, \beta_{05}, \beta_{11}, \dots, \beta_{15}, \sigma_1^2, \dots, \sigma_5^2, \sigma_w^2)$ é dada por

$$f(\mathbf{y}_1, \dots, \mathbf{y}_5 | \mathbf{v}_3) = \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2)}} \exp \left\{ -\frac{1}{2(\sigma_j^2 + \sigma_w^2)} [y_{ij} - (\beta_{0j} + \beta_{1j}z_i)]^2 \right\},$$

pois, as suposições de normalidade e independência foram consideradas, e ainda, se os parâmetros de regressão β_{1j} são diferentes de zero para $j = 1, \dots, 5$ e $i = 1, \dots, 60$, $y_{ij} \sim N(\beta_{0j} + \beta_{1j}z_i; \sigma_j^2 + \sigma_w^2)$, ou seja,

$$\begin{aligned} E(y_{ij}|z_i) &= E(E(y_{ij}|z_i, w_i)) \\ &= E(\beta_{0j} + \beta_{1j}z_i + w_i) \\ &= \beta_{0j} + \beta_{1j}z_i, \end{aligned}$$

e

$$\begin{aligned} \text{var}(y_{ij}|z_i) &= \text{var}[E(y_{ij}|z_i, w_i)] + E[\text{var}(y_{ij}|z_i, w_i)] \\ &= \text{var}(\beta_{0j} + \beta_{1j}z_i + w_i) + E(\sigma_j^2) \\ &= \sigma_w^2 + \sigma_j^2. \end{aligned}$$

A função de verossimilhança para os parâmetros $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{05})$, $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{15})$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_5^2)$ e σ_w^2 do modelo (5.1) é dada por

$$L(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\sigma}^2, \sigma_w^2) = \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2)^{-n/2} \exp \left(-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^{60} \epsilon_{ij}^2 \right). \quad (5.6)$$

em que

$$\sum_{i=1}^{60} \epsilon_{ij}^2 = \sum_{i=1}^{60} (y_{ij} - (\beta_{0j} + \beta_{1j}z_i))^2, j = 1, \dots, 5.$$

Também poderíamos considerar a função de verossimilhança sob uma análise Bayesiana com as quantidades aleatórias, $\beta_{0j}, \beta_{1j}, \sigma_j$ e σ_w , $i = 1, \dots, n$. Dessa forma, obteríamos a distribuição *a posteriori* conjunta via métodos Bayesianos hierárquicos para $\beta_{0j}, \beta_{1j}, \sigma_j, \sigma_w$, $i = 1, \dots, n$.

Para uma análise Bayesiana hierárquica, assumimos as seguintes distribuições *a priori* para $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$ para a estimação dos parâmetros do modelo (5.1),

$$\begin{aligned} \beta_{0j} &\sim N(a_{0j}, b_{0j}^2), \\ \beta_{1j} &\sim N(a_{1j}, b_{1j}^2), \\ \sigma_j^2 &\sim IG(c_j, d_j), \end{aligned} \tag{5.7}$$

em que $N(a, b^2)$ denota uma distribuição Normal com média a e variância b^2 , $IG(c, d)$ denota uma distribuição Gama inversa com média $d/(c-1)$ e variância $d^2/[(c-1)^2(c-2)]$, $c > 2$; e $a_{0j}, b_{0j}, a_{1j}, b_{1j}, c_j$ e d_j são hiperparâmetros conhecidos, para $j = 1, \dots, 5$.

Para o segundo estágio da análise Bayesiana hierárquica, assumimos a seguinte distribuição *a priori* para σ_w^2 da variável latente w_i ,

$$\sigma_w^2 \sim IG(c_w, d_w), \tag{5.8}$$

em que c_w e d_w são hiperparâmetros conhecidos. Além disso, assumimos independência *a priori* entre os parâmetros.

Desta forma, a partir do teorema de Bayes que associa a função de verossimilhança (5.6) com as distribuições *a priori* (5.7) e (5.8), obtemos a distribuição *a posteriori* conjunta para os parâmetros $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$ e σ_w^2 dada por

$$\begin{aligned}
\pi(\beta_0, \beta_1, \sigma^2, \sigma_w^2 | \mathbf{y}) &\propto \prod_{j=1}^5 \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \times \prod_{j=1}^5 \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \\
&\times \prod_{j=1}^5 (\sigma_j^2)^{-(c_j+1)} \exp \left(-\frac{d_j}{\sigma_j^2} \right) \times \prod_{j=1}^5 (\sigma_w^2)^{-(c_w+1)} \exp \left(-\frac{d_w}{\sigma_w^2} \right) \\
&\times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2)^{-n/2} \exp \left(-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right).
\end{aligned}$$

Consequentemente, as distribuições *a posteriori* condicionais utilizadas para o amostrador de Gibbs (Gelfand e Smith, [20]) e para o Metropolis-Hastings são dadas por:

$$i) \quad \pi(\beta_{0j} | \beta_1, \sigma^2, \sigma_w^2, \mathbf{y}) \propto \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n (\beta_{0j} - \mu_i^{(j)})^2 \right],$$

em que $\mu_i^{(j)} = y_{ij} - \beta_{1j} z_i$; $i = 1, \dots, n$ e $j = 1, \dots, 5$.

Logo,

$$\pi(\beta_{0j} | \beta_1, \sigma^2, \mathbf{y}) \sim N \left[\frac{a_{0j}(\sigma_j^2 + \sigma_w^2) + b_{0j} \sum_{i=1}^n \mu_i^{(j)}}{(\sigma_j^2 + \sigma_w^2) + nb_{0j}^2}, \frac{b_{0j}^2(\sigma_j^2 + \sigma_w^2)}{(\sigma_j^2 + \sigma_w^2) + nb_{0j}^2} \right]. \quad (5.9)$$

$$ii) \quad \pi(\beta_{1j} | \beta_0, \sigma^2, \sigma_w^2, \mathbf{y}) \propto \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n (\beta_{1j} z_i - \theta_i^{(j)})^2 \right],$$

em que $\theta_i^{(j)} = y_{ij} - \beta_{0j}$; $i = 1, \dots, n$ e $j = 1, \dots, 5$.

Logo,

$$\pi(\beta_{1j} | \beta_0, \sigma^2, \sigma_w^2, \mathbf{y}) \sim N \left[\frac{a_{1j}(\sigma_j^2 + \sigma_w^2) + b_{1j} \sum_{i=1}^n z_i \theta_i^{(j)}}{(\sigma_j^2 + \sigma_w^2) + b_{1j}^2 \sum_{i=1}^n z_i^2}, \frac{b_{1j}^2(\sigma_j^2 + \sigma_w^2)}{(\sigma_j^2 + \sigma_w^2) + b_{1j}^2 \sum_{i=1}^n z_i^2} \right]. \quad (5.10)$$

$$\begin{aligned}
iii) \quad \pi(\sigma_j^2 | \beta_0, \beta_1, \sigma_w^2, \mathbf{y}) &\propto (\sigma_j^2)^{-(c_j+1)} \exp\left(-\frac{d_j}{\sigma_j^2}\right) \\
&\times (\sigma_j^2 + \sigma_w^2)^{-n/2} \exp\left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2\right] \\
&\propto (\sigma_j^2)^{-(c_j+1)} (\sigma_j^2 + \sigma_w^2)^{-n/2} \psi_1(\beta_0, \beta_1, \sigma^2, \sigma_w^2), \quad (5.11)
\end{aligned}$$

em que $\psi_1(\beta_0, \beta_1, \sigma^2, \sigma_w^2) = \exp\left[-\frac{d_j}{\sigma_j^2} - \frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2\right]$; $i = 1, \dots, n$ e $j = 1, \dots, 5$.

$$\begin{aligned}
iv) \quad \pi(\sigma_w^2 | \beta_0, \beta_1, \sigma^2, \mathbf{y}) &\propto (\sigma_w^2)^{-(c_w+1)} \exp\left(-\frac{d_w}{\sigma_w^2}\right) \\
&\times (\sigma_j^2 + \sigma_w^2)^{-n/2} \exp\left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2\right] \\
&\propto (\sigma_w^2)^{-(c_w+1)} (\sigma_j^2 + \sigma_w^2)^{-n/2} \psi_2(\beta_0, \beta_1, \sigma^2, \sigma_w^2), \quad (5.12)
\end{aligned}$$

em que $\psi_2(\beta_0, \beta_1, \sigma^2, \sigma_w^2) = \exp\left[-\frac{d_w}{\sigma_w^2} - \frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2\right]$; $i = 1, \dots, n$ e $j = 1, \dots, 5$.

Podemos observar que as distribuições *a posteriori* condicionais (5.9) e (5.10) apresentaram formas fechadas e, portanto, obtemos as estimativas dos parâmetros do modelo (5.1) através do algoritmo amostrador de Gibbs. Para a distribuição *a posteriori* condicional (5.11) e (5.12) utilizamos o algoritmo de Metropolis-Hastings, por não apresentar forma fechada.

Para a análise Bayesiana dos dados apresentados no Apêndice B, considera-se a transformação *alr* com erros normais não correlacionados para o modelo proposto (5.1). Assumindo distribuições *a priori* para $\beta_{0j}, \beta_{1j}, \sigma_j^2, \sigma_w^2$ com hiperparâmetros cujos valores foram $a_{0j} = 1, a_{1j} = 0, b_{0j} = b_{1j} = 1.000.000, c_j = d_j = a_w = b_w = 1.000$, geramos no *software* OPENBUGS 205.000 iterações, nas quais foram descartadas as 5.000 iterações iniciais (“burn-in”). Após o período de “burn-in”, consideramos saltos de tamanho 20 para obtermos amostras não correlacionadas, na qual resultou uma amostra final de tamanho 10.000 para cada parâmetro. A convergência do algoritmo foi verificada pelos gráficos dos traços *a posteriori* das amostras de Gibbs simuladas e pelo diagnóstico de Geweke (Geweke, [23]).

Na Tabela 5.1 apresentamos os sumários *a posteriori* dos parâmetros do modelo (5.1) e os valores correspondentes da estatística de teste para o diagnóstico de convergência

Geweke. Dos resultados da Tabela 5.1, observamos que há evidências de diferenças significativas nos grupos antes e depois da dieta nos seguintes componentes do leite: proteína, carboidrato e cálcio, pois os intervalos de credibilidade para β_{11} , β_{13} e β_{14} não incluem o valor zero.

Tabela 5.1: Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke - Modelo 3.

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	-0,060	0,042	(-0,143; 0,022)	-0,374
β_{02}	0,079	0,042	(-0,003; 0,162)	-0,979
β_{03}	-0,058	0,031	(-0,119; 0,003)	-0,473
β_{04}	0,070	0,022	(0,027; 0,114)	0,593
β_{05}	-0,150	0,034	(-0,150; -0,082)	0,452
β_{11}	0,365	0,059	(0,247; 0,482)	0,111
β_{12}	-0,083	0,060	(-0,206; 0,035)	0,946
β_{13}	0,170	0,044	(0,082; 0,258)	-1,006
β_{14}	0,415	0,032	(0,353; 0,478)	0,405
β_{15}	0,090	0,048	(-0,003; 0,182)	-0,267
σ_1	0,050	0,010	(0,034; 0,073)	0,163
σ_2	0,052	0,011	(0,035; 0,077)	0,623
σ_3	0,027	0,006	(0,018; 0,041)	-0,760
σ_4	0,012	0,003	(0,008; 0,019)	0,426
σ_5	0,033	0,007	(0,022; 0,048)	-0,785
σ_w	0,002	0,001	(0,001; 0,006)	-0,156

As densidades *a posteriori* e o comportamento da cadeia gerada para os parâmetros do modelo (5.1) são apresentadas nas Figuras 5.2 a 5.5. Podemos observar que aparentemente o comportamento das cadeias para todos os parâmetros apresentam certa uniformidade.

Além disso, a convergência dos algoritmos amostrador de Gibbs e Metropolis-Hastings foram verificadas utilizando o diagnóstico de Geweke para todos os parâmetros. Assim, observa-se que os valores da estatística do teste z_G estão no intervalo $(-1,96; 1,96)$ indicando convergência para todos os parâmetros do modelo (5.1).

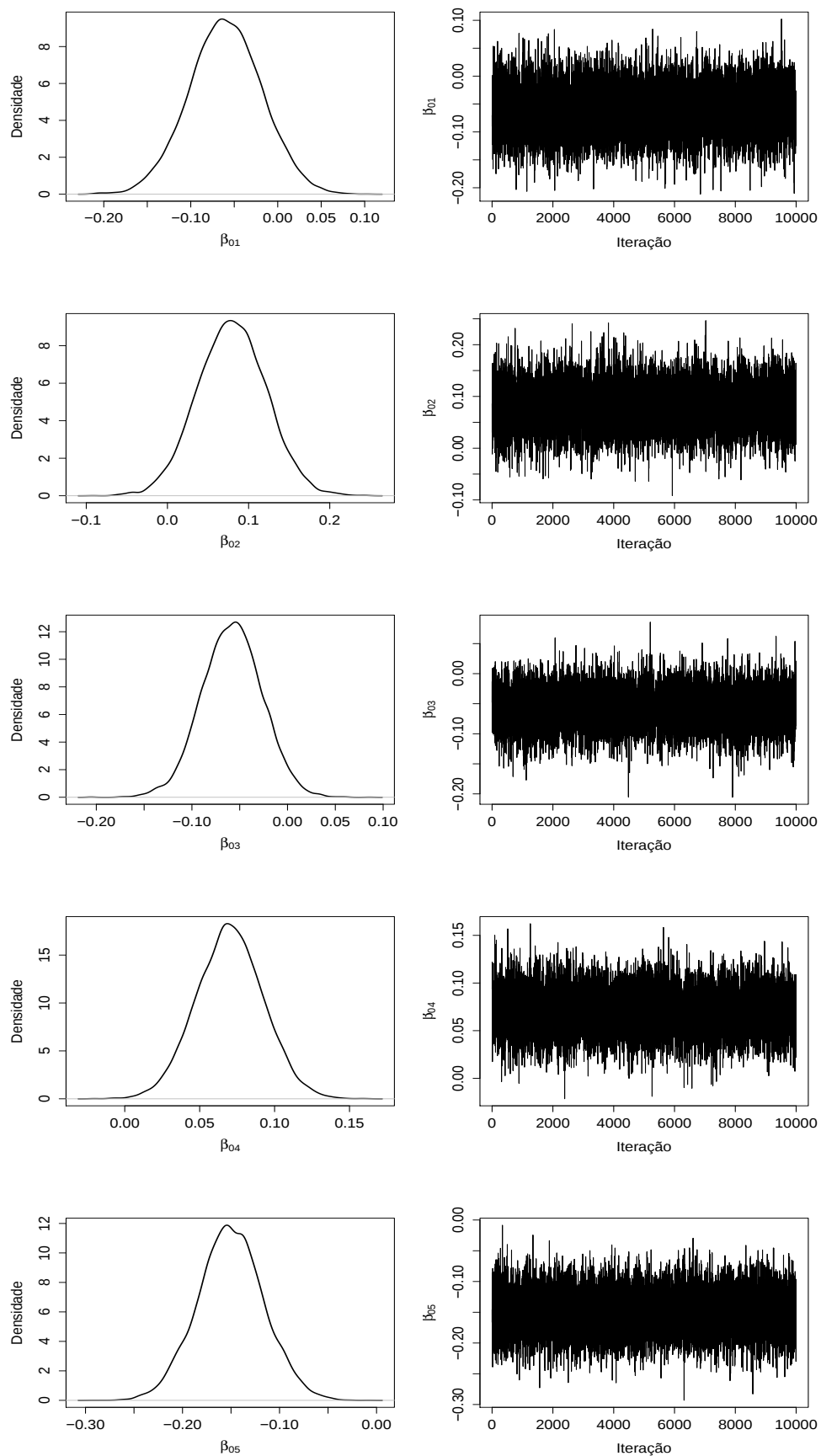


Figura 5.2: Densidades *a posteriori* e traços *a posteriori* para os parâmetros β_0 's no modelo 3.

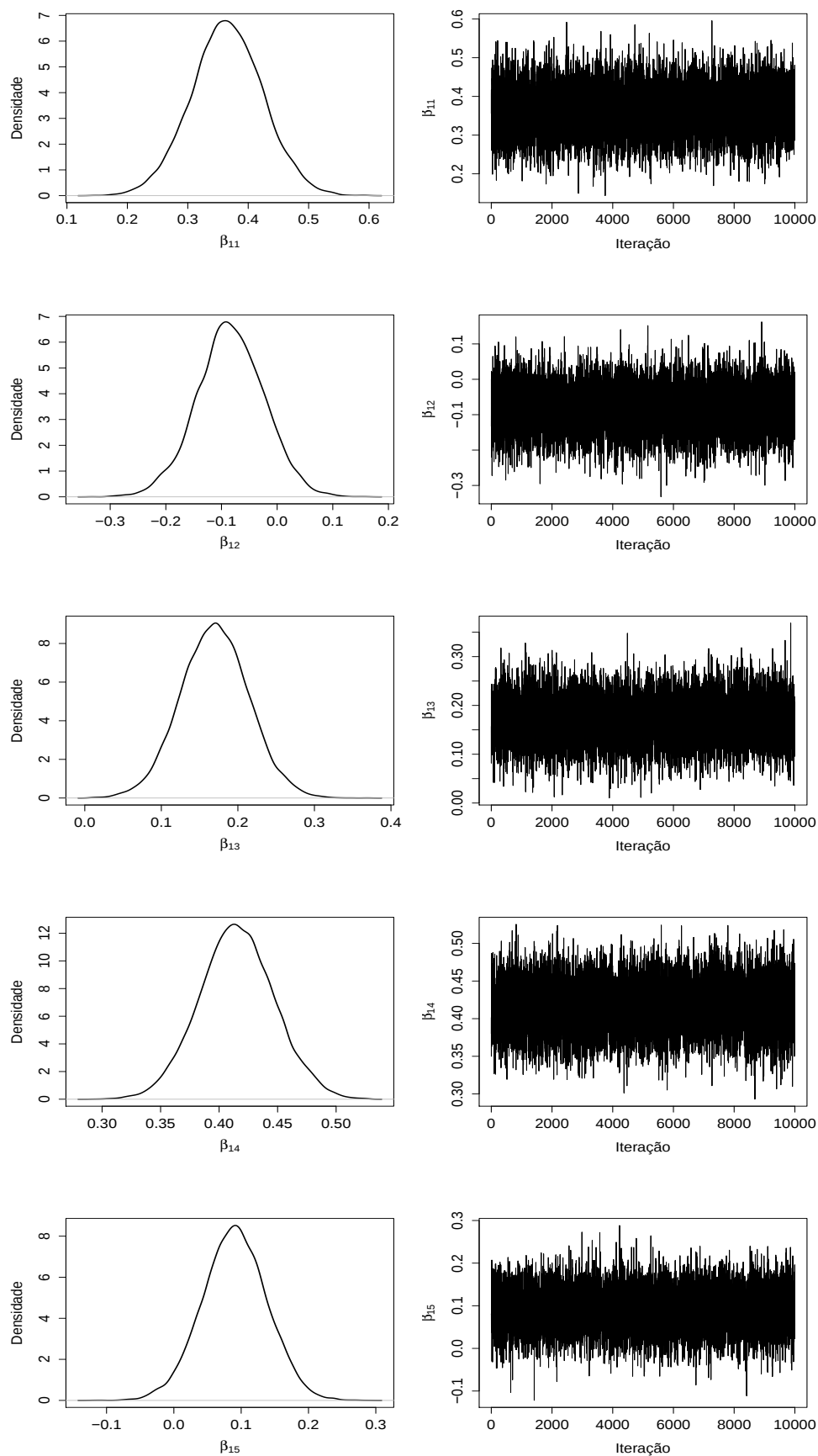


Figura 5.3: Densidades *a posteriori* e traços *a posteriori* para os parâmetros β_1 's no modelo 3.

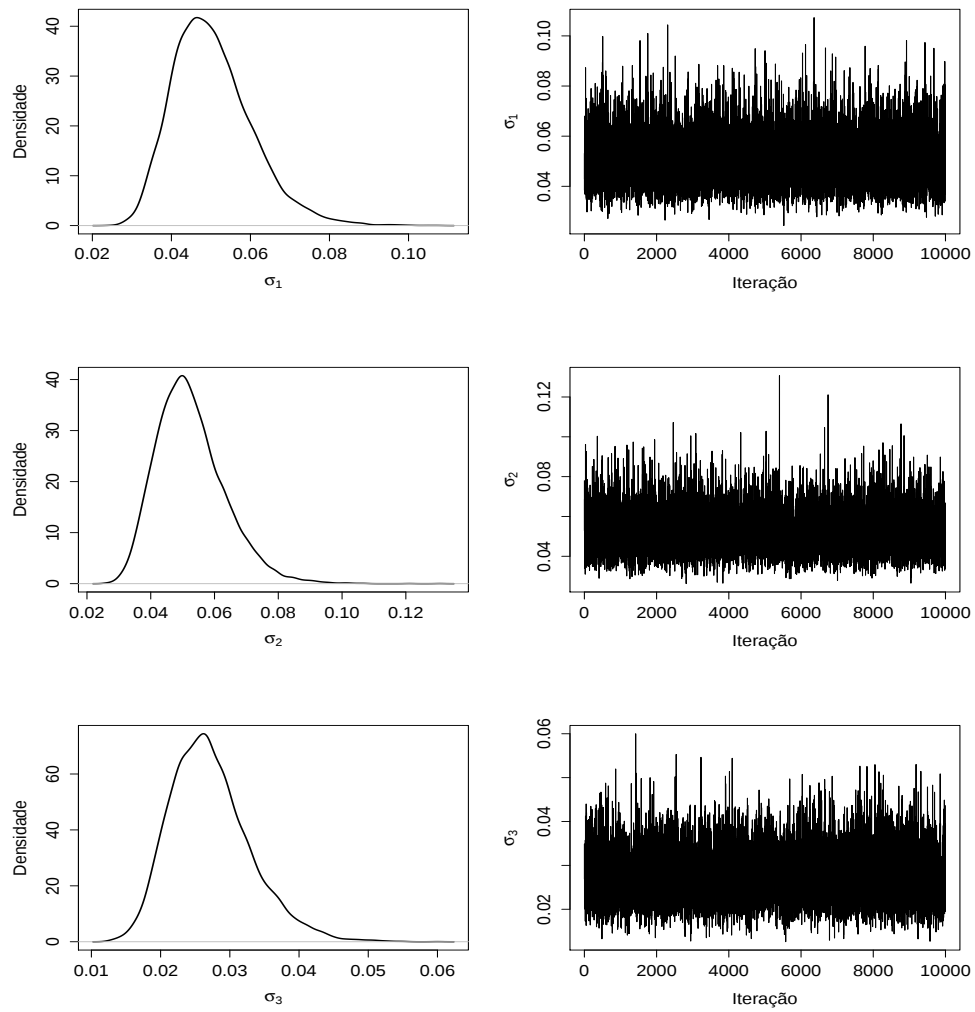


Figura 5.4: Densidades *a posteriori* e traços *a posteriori* para os parâmetros de variância no modelo 3.

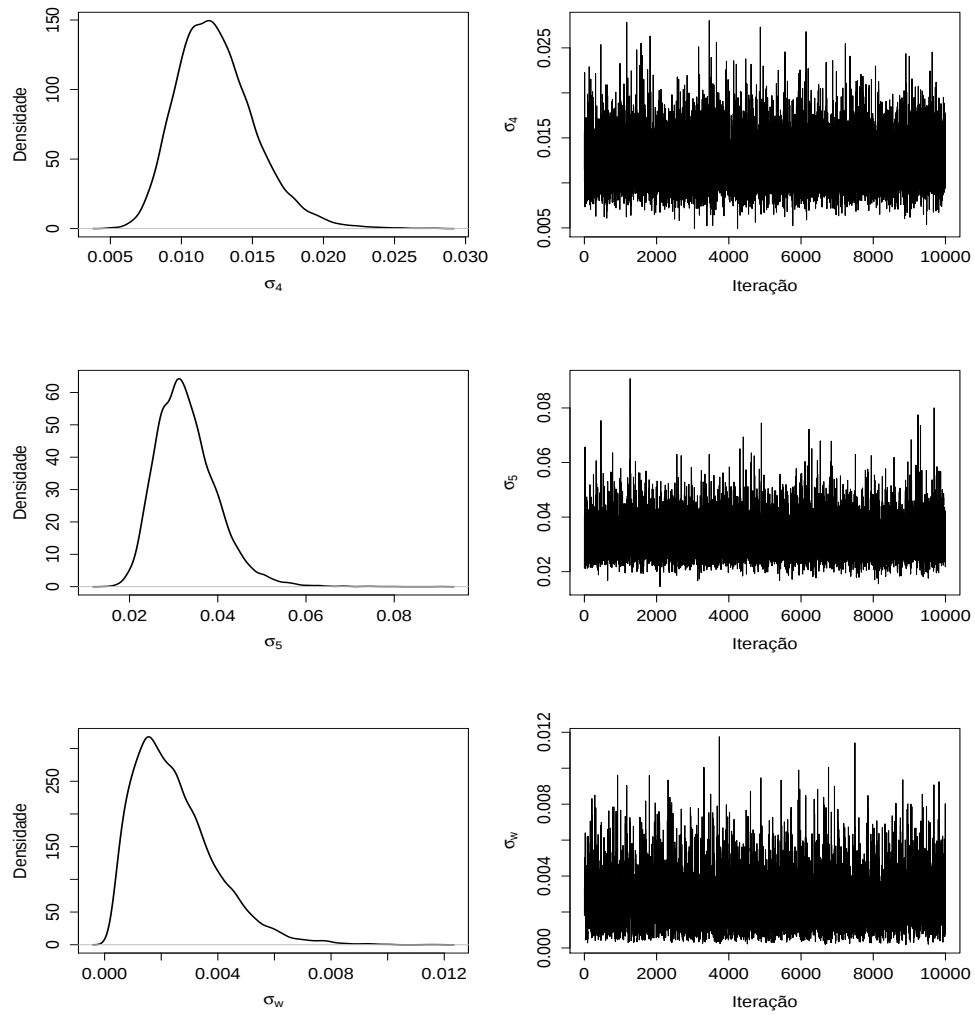


Figura 5.5: Densidades *a posteriori* e traços *a posteriori* para os parâmetros de variância no modelo 3.

A Tabela 5.2 apresenta as estimativas das proporções verdadeiras para os componentes de proteína, gordura do leite, carboidrato, cálcio, sódio e potássio para $z_i = 0$ (Grupo 1 - antes da dieta) e para $z_i = 1$ (Grupo 2 - depois da dieta), considerando a transformação *alr* e um efeito aleatório no modelo.

Tabela 5.2: Estimativas das proporções dos componentes - Modelo 3.

Componentes do leite	Proporções	$z = 0$ (Grupo 1)	Proporções	$z = 1$ (Grupo 2)
proteína	α_1	0,159	α_7	0,193
gordura do leite	α_2	0,183	α_8	0,141
carboidrato	α_3	0,160	α_9	0,159
cálcio	α_4	0,182	α_{10}	0,231
sódio	α_5	0,146	α_{11}	0,134
potássio	α_6	0,169	α_{12}	0,142

Os resultados da Tabela 5.2 mostram que após a dieta balanceada houve um aumento para as proporções estimadas dos seguintes componentes do leite: proteína e cálcio. Por outro lado, para os componentes gordura do leite e potássio as proporções estimadas diminuíram. E finalmente, para os componentes carboidrato e sódio, aparentemente não houve diferença depois da dieta estabelecida nas proporções estimadas.

5.2 Transformação Box-Cox Considerando um Efeito Aleatório - Modelo 4

Nesta seção, é desenvolvido uma aplicação da transformação Box-Cox (3.3) nos dados composicionais do Apêndice B, x_{i1} , x_{i2} , x_{i3} , x_{i4} , x_{i5} e x_{i6} para $i = 1, \dots, 60$.

Para a obtenção das amostras simuladas *a posteriori* foram utilizados os métodos Bayesianos baseados no amostrador de Gibbs e no algoritmo de Metropolis-Hastings.

Considerando a transformação Box-Cox, obtemos

$$y_{i1}^{(\lambda_1)} = \begin{cases} \frac{(\frac{x_{i1}}{x_{i6}})^{\lambda_1} - 1}{\lambda_1} & \text{se } \lambda_1 \neq 0, \\ \log\left(\frac{x_{i1}}{x_{i6}}\right) & \text{se } \lambda_1 = 0 \end{cases}, \quad y_{i2}^{(\lambda_2)} = \begin{cases} \frac{(\frac{x_{i2}}{x_{i6}})^{\lambda_2} - 1}{\lambda_2} & \text{se } \lambda_2 \neq 0, \\ \log\left(\frac{x_{i2}}{x_{i6}}\right) & \text{se } \lambda_2 = 0 \end{cases}$$

$$y_{i3}^{(\lambda_3)} = \begin{cases} \frac{(\frac{x_{i3}}{x_{i6}})^{\lambda_3} - 1}{\lambda_3} & \text{se } \lambda_3 \neq 0, \\ \log\left(\frac{x_{i3}}{x_{i6}}\right) & \text{se } \lambda_3 = 0 \end{cases}, \quad y_{i4}^{(\lambda_4)} = \begin{cases} \frac{(\frac{x_{i4}}{x_{i6}})^{\lambda_4} - 1}{\lambda_4} & \text{se } \lambda_4 \neq 0, \\ \log\left(\frac{x_{i4}}{x_{i6}}\right) & \text{se } \lambda_4 = 0 \end{cases}$$

$$y_{i5}^{(\lambda_5)} = \begin{cases} \frac{(\frac{x_{i5}}{x_{i6}})^{\lambda_5} - 1}{\lambda_5} & \text{se } \lambda_5 \neq 0, \\ \log\left(\frac{x_{i5}}{x_{i6}}\right) & \text{se } \lambda_5 = 0 \end{cases}$$

Como já descrito na Seção 4.2, a transformação Box-Cox é um caso especial da transformação *alr* quando $\lambda_j = 0$ para $j = 1, \dots, 5$.

Utilizando os dados transformados $y_{i1}^{(\lambda_1)}, \dots, y_{i5}^{(\lambda_5)}$, $i = 1, \dots, 60$, e assumindo o modelo (5.1), o modelo 4 é dado por

$$\mathbf{y}_i^{(\boldsymbol{\lambda})} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_i + w_i + \boldsymbol{\epsilon}_i, \quad (5.13)$$

em que $\mathbf{y}_i$ é o vetor de variáveis resposta, para $j = 1, \dots, 5$ e $i = 1, \dots, 60$ ($n = n_1 + n_2$, em que $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2), z_i representa a covariável associada ao i -ésimo indivíduo (variável indicadora em que $z_i = 0$ para o grupo 1 e $z_i = 1$ para o grupo 2), w_i é um efeito aleatório (variável latente não observável) que captura a dependência entre as medidas repetidas para cada indivíduo e $\boldsymbol{\epsilon}_i$ é o vetor de erros.

Da mesma forma, que em (5.2), assumiremos para a variável latente não observável, $w_i \sim N(0, \sigma_w^2)$ e $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \Sigma)$, em que $\mathbf{0}$ é o vetor de zeros e Σ é a matriz de variâncias e covariâncias.

Considerando o modelo apresentado em (5.13), assumiremos que os erros $\epsilon_{i1}, \dots, \epsilon_{i5}$ são mutuamente independentes. Assim, obtém-se,

$$\begin{aligned} \epsilon_{i1} &= y_{i1}^{(\lambda_1)} - \beta_{01} - \beta_{11} z_i - w_i, \\ &\vdots \\ \epsilon_{i5} &= y_{i5}^{(\lambda_5)} - \beta_{05} - \beta_{15} z_i - w_i. \end{aligned}$$

Consideramos as suposições de normalidade e independência, e ainda, $y_{ij}^{(\lambda_j)} \sim N(\beta_{0j} + \beta_{1j} z_i, \sigma_j^2 + \sigma_w^2)$, para $j = 1, \dots, 5$ e $i = 1, \dots, 60$.

Além disso, podemos considerar $\alpha_{i1}, \dots, \alpha_{i6}$ as proporções verdadeiras dos componentes, em que $\alpha_{i1} + \dots + \alpha_{i6} = 1$, com $\alpha_{i1} > 0, \dots, \alpha_{i6} > 0$ e a transformação Box-Cox, obtendo a seguinte relação a partir de (5.13),

$$\begin{aligned}
\frac{(\alpha_{i1}/\alpha_{i6})^{\lambda_1} - 1}{\lambda_1} &= \beta_{01} + \beta_{11}z_i + w_i, \\
&\vdots \\
\frac{(\alpha_{i5}/\alpha_{i6})^{\lambda_5} - 1}{\lambda_5} &= \beta_{05} + \beta_{15}z_i + w_i.
\end{aligned} \tag{5.14}$$

Reescrevendo (5.14), temos que

$$\begin{aligned}
\left(\frac{\alpha_{i1}}{\alpha_{i6}}\right)^{\lambda_1} &= \lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1 \Rightarrow \left(\frac{\alpha_{i1}}{\alpha_{i6}}\right)^{\lambda_1 \cdot \frac{1}{\lambda_1}} = [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} \\
&\vdots \\
\left(\frac{\alpha_{i5}}{\alpha_{i6}}\right)^{\lambda_5} &= \lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1 \Rightarrow \left(\frac{\alpha_{i5}}{\alpha_{i6}}\right)^{\lambda_5 \cdot \frac{1}{\lambda_5}} = [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}},
\end{aligned}$$

ou seja,

$$\begin{cases} \alpha_{i1} &= \alpha_{i6} [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} \\ &\vdots \\ \alpha_{i5} &= \alpha_{i6} [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}}. \end{cases} \tag{5.15}$$

Sabemos que $\alpha_{i1} + \dots + \alpha_{i6} = 1$, ou seja,

$$\alpha_{i6} = 1 - (\alpha_{i1} + \dots + \alpha_{i5}) \Rightarrow \alpha_{i1} + \dots + \alpha_{i5} = 1 - \alpha_{i6}.$$

De (5.15), obtemos a seguinte relação,

$$\alpha_{i1} + \dots + \alpha_{i5} = \alpha_{i6} \left\{ [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}} \right\},$$

ou seja,

$$\begin{aligned}
1 - \alpha_{i6} &= \alpha_{i6} \left\{ [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}} \right\} \Rightarrow \\
1 &= \alpha_{i6} \left\{ 1 + [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}} \right\} \Rightarrow \\
\alpha_{i6} &= \frac{1}{\left\{ 1 + [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}} \right\}}.
\end{aligned} \tag{5.16}$$

Logo, substituindo (5.16) em (5.15), obtemos os estimadores das proporções verdadeiras dos componentes do grupo 1 (antes da dieta) quando $z_i = 0$ e do grupo 2 (depois da dieta) quando $z_i = 1$, para a transformação Box-Cox,

$$\begin{aligned}
\alpha_{ij} &= \frac{[\lambda_j(\beta_{0j} + \beta_{1j}z_i + w_i) + 1]^{\frac{1}{\lambda_j}}}{\left\{ 1 + [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}} \right\}}, \\
&\vdots \\
\alpha_{i6} &= \frac{1}{\left\{ 1 + [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i) + 1]^{\frac{1}{\lambda_5}} \right\}}, \tag{5.17}
\end{aligned}$$

em que $i = 1, \dots, 60$ ($n = n_1 + n_2$, em que $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2) e $j = 1, \dots, 5$.

5.2.1 Análise Bayesiana - Modelo 4

A função densidade de probabilidade conjunta de $\mathbf{y}_1^{(\lambda_1)}, \dots, \mathbf{y}_5^{(\lambda_5)}$ dado o vetor de parâmetros $\mathbf{v}_4 = (\beta_{01}, \dots, \beta_{05}, \beta_{11}, \dots, \beta_{15}, \sigma_1^2, \dots, \sigma_5^2, \sigma_w^2, \lambda_1, \dots, \lambda_5)$ é dada por

$$f(\mathbf{y}_1^{(\lambda_1)}, \dots, \mathbf{y}_5^{(\lambda_5)} | \mathbf{v}_4) = \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2)}} \exp \left\{ -\frac{1}{2(\sigma_j^2 + \sigma_w^2)} [y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j}z_i)]^2 \right\}.$$

Por outro lado, a função densidade de probabilidade das observações não transformadas (variáveis originais) [Box & Tiao (1973)], $y_{i1} = x_{i1}/x_{i6}$, $y_{i2} = x_{i2}/x_{i6}$, $y_{i3} = x_{i3}/x_{i6}$, $y_{i4} = x_{i4}/x_{i6}$, $y_{i5} = x_{i5}/x_{i6}$ é dada por

$$f(\mathbf{y}_1, \dots, \mathbf{y}_5 | \mathbf{v}_4) = \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2)}} \prod_{i=1}^n \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \left(y_{ij}^{(\lambda_j)} - \beta_{0j} - \beta_{1j} z_i \right)^2 \right] \left(\frac{dy_{ij}^{(\lambda_j)}}{dy_{ij}} \right),$$

em que para $j = 1, \dots, 5$, $\frac{dy_{ij}^{(\lambda_j)}}{dy_{ij}} = \frac{\lambda_j y_{ij}^{\lambda_j-1}}{\lambda_j} = y_{ij}^{\lambda_j-1}$ são os elementos da diagonal da matriz Jacobiano.

A função de verossimilhança para $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{05})$, $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{15})$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_5^2)$, σ_w^2 e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_5)$ do modelo (5.13) é dada por

$$L(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\sigma}^2, \sigma_w^2, \boldsymbol{\lambda}) = \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2)^{-n/2} \left(\prod_{i=1}^n y_{ij}^{\lambda_j-1} \right) \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^{60} \epsilon_{ij}^2 \right], \quad (5.18)$$

em que $\left(\prod_{i=1}^n y_{i1}^{\lambda_1-1} \right), \dots, \left(\prod_{i=1}^n y_{i5}^{\lambda_5-1} \right)$ são os produtos do Jacobiano, e

$$\sum_{i=1}^{60} \epsilon_{ij}^2 = \sum_{i=1}^{60} \left[y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j} z_i) \right]^2, \quad j = 1, \dots, 5.$$

Para uma análise Bayesiana hierárquica com os dados transformados através da transformação Box-Cox, assumimos as mesmas distribuições *a priori* para $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{05})$, $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{15})$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_5^2)$ e σ_w^2 dadas em (5.7) e (5.8) para a estimação dos parâmetros do modelo (5.13).

Para $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_5)$, assumimos a distribuição *a priori*,

$$\lambda_j \sim N(e_j, f_j^2), \quad (5.19)$$

em que $j = 1, \dots, 5$ e e_j e f_j são hiperparâmetros conhecidos.

Desta forma, a partir do teorema de Bayes que associa a função de verossimilhança (5.18) com as distribuições *a priori* (5.7), (5.8) e (5.19), obtemos a distribuição *a posteriori* conjunta para os parâmetros $\boldsymbol{\beta}_0 = (\beta_{01}, \dots, \beta_{05})$, $\boldsymbol{\beta}_1 = (\beta_{11}, \dots, \beta_{15})$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_5^2)$, σ_w^2 e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_5)$ dada por

$$\begin{aligned}
\pi(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \lambda | \mathbf{y}) &\propto \prod_{j=1}^5 \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \times \prod_{j=1}^5 \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \\
&\times \prod_{j=1}^5 (\sigma_j^2)^{-(c_j+1)} \exp \left(-\frac{d_j}{\sigma_j^2} \right) \times \prod_{j=1}^5 (\sigma_w^2)^{-(c_w+1)} \exp \left(-\frac{d_w}{\sigma_w^2} \right) \\
&\times \prod_{j=1}^5 \exp \left[-\frac{1}{2f_j^2} (\lambda_j - e_j)^2 \right] \times \left(\prod_{j=1}^5 \prod_{i=1}^n y_{ij}^{\lambda_j-1} \right) \\
&\times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right]. \quad (5.20)
\end{aligned}$$

Considerando a distribuição *a posteriori* conjunta (5.20), as distribuições *a posteriori* condicionais utilizadas para o amostrador de Gibbs (Gelfand e Smith, [20]) e para o algoritmo de Metropolis-Hastings são dadas por:

$$\begin{aligned}
i) \quad \pi(\beta_{0j} | \beta_1, \sigma^2, \sigma_w^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \\
&\times \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n (\beta_{0j} - \mu_i^{(j)})^2 \right],
\end{aligned}$$

em que $\mu_i^{(j)} = y_{ij}^{(\lambda_j)} - \beta_{1j} z_i$; $i = 1, \dots, n$ e $j = 1, \dots, 5$.

Logo,

$$\pi(\beta_{0j} | \beta_1, \sigma^2, \sigma_w^2, \lambda, \mathbf{y}^{(\lambda)}) \sim N \left[\frac{a_{0j}(\sigma_j^2 + \sigma_w^2) + b_{0j} \sum_{i=1}^n \mu_i^{(j)}}{(\sigma_j^2 + \sigma_w^2) + nb_{0j}^2}, \frac{b_{0j}^2(\sigma_j^2 + \sigma_w^2)}{(\sigma_j^2 + \sigma_w^2) + nb_{0j}^2} \right]. \quad (5.21)$$

$$\begin{aligned}
ii) \quad \pi(\beta_{1j} | \beta_0, \sigma^2, \sigma_w^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \\
&\times \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n (\beta_{1j} z_i - \theta_i^{(j)})^2 \right],
\end{aligned}$$

em que $\theta_i^{(j)} = y_{ij}^{(\lambda_j)} - \beta_{0j}$; $i = 1, \dots, n$ e $j = 1, \dots, 5$.

Logo,

$$\pi(\beta_{1j}|\beta_0, \sigma^2, \sigma_w^2, \lambda, \mathbf{y}^{(\lambda)}) \sim N \left[\frac{a_{1j}(\sigma_j^2 + \sigma_w^2) + b_{1j} \sum_{i=1}^n z_i \theta_i^{(j)}}{(\sigma_j^2 + \sigma_w^2) + b_{1j}^2 \sum_{i=1}^n z_i^2}, \frac{b_{1j}^2(\sigma_j^2 + \sigma_w^2)}{(\sigma_j^2 + \sigma_w^2) + b_{1j}^2 \sum_{i=1}^n z_i^2} \right]. \quad (5.22)$$

$$\begin{aligned} iii) \quad \pi(\sigma_j^2|\beta_0, \beta_1, \sigma_w^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto (\sigma_j^2)^{-(c_j+1)} \exp \left(-\frac{d_j}{\sigma_j^2} \right) \\ &\times (\sigma_j^2 + \sigma_w^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right] \\ &\propto (\sigma_j^2)^{-(c_j+1)} (\sigma_j^2 + \sigma_w^2)^{-n/2} \psi_1(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \lambda), \end{aligned} \quad (5.23)$$

em que $\psi_1(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \lambda) = \exp \left[-\frac{d_j}{\sigma_j^2} - \frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right] e$
 $\epsilon_{ij}^2 = [y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j} z_i)]^2; i = 1, \dots, n$ e $j = 1, \dots, 5$.

$$\begin{aligned} iv) \quad \pi(\sigma_w^2|\beta_0, \beta_1, \sigma^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto (\sigma_w^2)^{-(c_w+1)} \exp \left(-\frac{d_w}{\sigma_w^2} \right) \\ &\times (\sigma_j^2 + \sigma_w^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right] \\ &\propto (\sigma_w^2)^{-(c_w+1)} (\sigma_j^2 + \sigma_w^2)^{-n/2} \psi_2(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \lambda), \end{aligned} \quad (5.24)$$

em que $\psi_2(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \lambda) = \exp \left[-\frac{d_w}{\sigma_w^2} - \frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right]; i = 1, \dots, n$ e $j = 1, \dots, 5$.

$$v) \quad \pi(\lambda_j|\beta_0, \beta_1, \sigma^2, \sigma_w^2, \mathbf{y}^{(\lambda)}) \propto \exp \left[-\frac{1}{2f_j^2} (\lambda_j - e_j)^2 \right] \psi_j(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \lambda), \quad (5.25)$$

em que

$$\begin{aligned} \psi_j(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \lambda) &= \exp \left\{ (\lambda_j - 1) \sum_{i=1}^n \ln y_{ij} + \frac{1}{(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n [y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j})]^2 \right\} \\ &= \exp \left[(\lambda_j - 1) \sum_{i=1}^n \ln y_{ij} + \frac{1}{(\sigma_j^2 + \sigma_w^2)} \left(\sum_{i=1}^n y_{ij}^{(\lambda_j)^2} - 2\beta_{0j} \sum_{i=1}^n y_{ij}^{(\lambda_j)} - 2\beta_{1j} \sum_{i=1}^n y_{ij}^{(\lambda_j)} z_i \right) \right], \end{aligned}$$

para $i = 1, \dots, n$ e $j = 1, \dots, 5$.

Podemos observar que as distribuições *a posteriori* condicionais (5.21) e (5.22) apresentaram formas fechadas e, portanto, obtemos as estimativas dos parâmetros do modelo (5.13) através do algoritmo amostrador de Gibbs. Por outro lado, para as distribuições *a posteriori* condicionais (5.23), (5.24) e (5.25) utilizamos o algoritmo de Metropolis-Hastings, por não apresentarem formas fechadas.

Assim, considerando a transformação Box-Cox com erros normais não correlacionados para o modelo (5.13), assumimos as distribuições *a priori* para $\beta_{0j}, \beta_{1j}, \sigma_j^2, \sigma_w^2$ e λ_j com hiperparâmetros cujos valores foram $a_{0j} = 1; b_{0j} = b_{1j} = f_j = 1.000.000, a_{1j} = e_j = 0, c_j = d_j = a_w = b_w = 1.000$, geramos no *software* OPENBUGS 205.000 iterações, nas quais foram descartadas as 5.000 iterações iniciais (“burn-in”). Após o período de “burn-in”, consideramos saltos de tamanho 20 para obtermos amostras não correlacionadas, na qual resultou uma amostra final de tamanho 10.000 para cada parâmetro.

Na Tabela 5.3 apresentamos os sumários *a posteriori* dos parâmetros do modelo (5.13) e os valores correspondentes da estatística de teste z_G para o diagnóstico de convergência Geweke.

Tabela 5.3: Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke - Modelo 4.

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	-0,060	0,043	(-0,144; 0,026)	0,037
β_{02}	0,076	0,046	(-0,013; 0,166)	1,514
β_{03}	-0,069	0,032	(-0,132; -0,004)	1,636
β_{04}	0,064	0,013	(0,039; 0,092)	1,207
β_{05}	-0,156	0,038	(-0,234; -0,085)	-0,180
β_{11}	0,373	0,065	(0,249; 0,503)	0,860
β_{12}	-0,084	0,062	(-0,208; 0,036)	-1,067
β_{13}	0,166	0,044	(0,079; 0,252)	-1,868
β_{14}	0,274	0,040	(0,206; 0,365)	-1,346
β_{15}	0,092	0,049	(-0,004; 0,188)	0,308
σ_1	0,054	0,012	(0,035; 0,082)	-0,281
σ_2	0,055	0,011	(0,037; 0,081)	0,759
σ_3	0,028	0,006	(0,019; 0,041)	-0,441
σ_4	0,004	0,002	(0,002; 0,009)	-0,814
σ_5	0,035	0,009	(0,022; 0,057)	-0,399
σ_w	0,001	0,0007	(0,0003; 0,003)	-0,770
λ_1	0,075	0,381	(-0,681; 0,822)	-0,228
λ_2	-0,184	0,506	(-1,170; 0,829)	-0,676
λ_3	-0,723	0,610	(-1,940; 0,466)	0,635
λ_4	-1,523	0,483	(-2,446; -0,518)	-1,132
λ_5	-0,205	0,627	(-1,440; 1,012)	-0,119

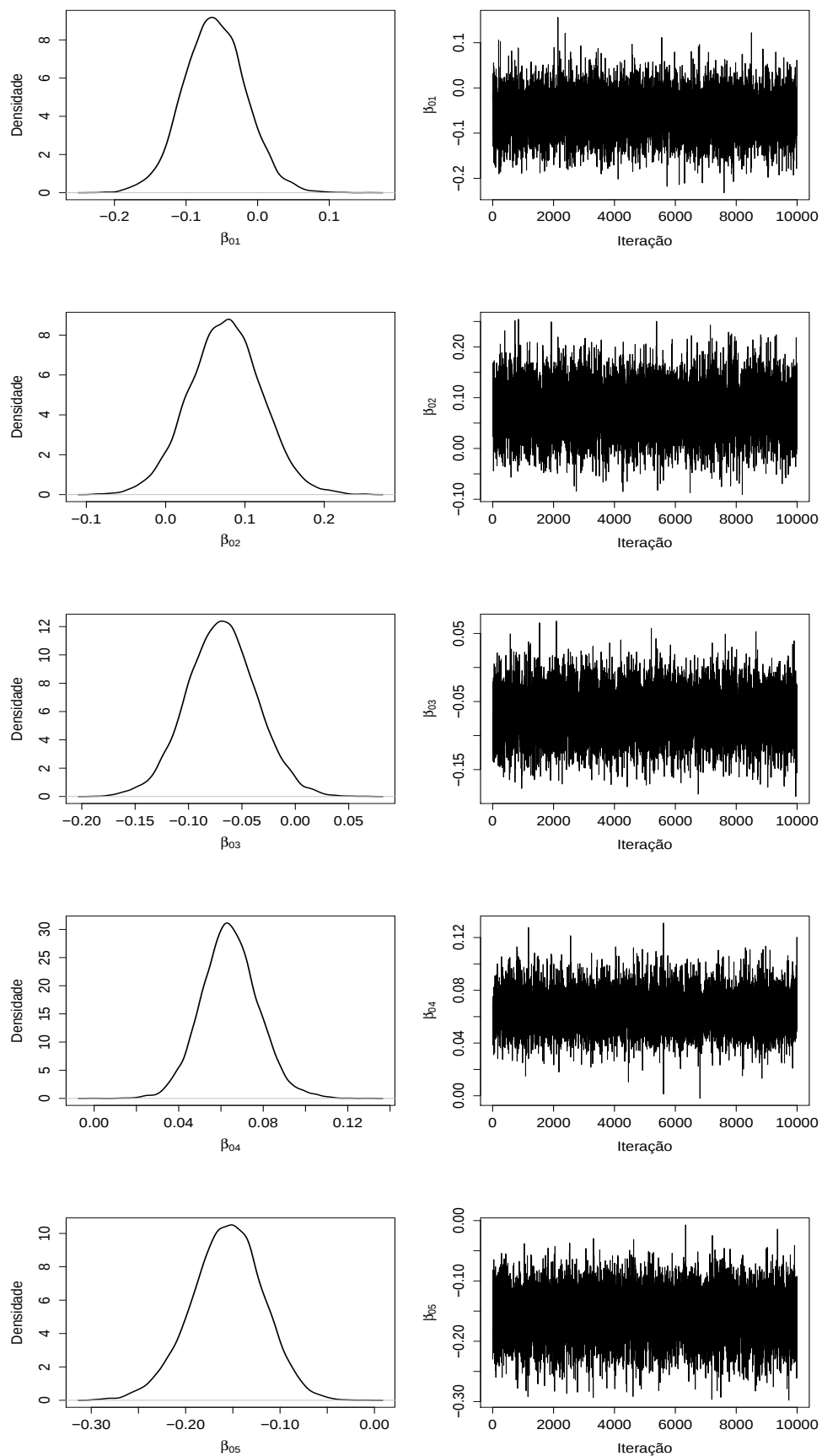


Figura 5.6: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros β_0 's no modelo 4.

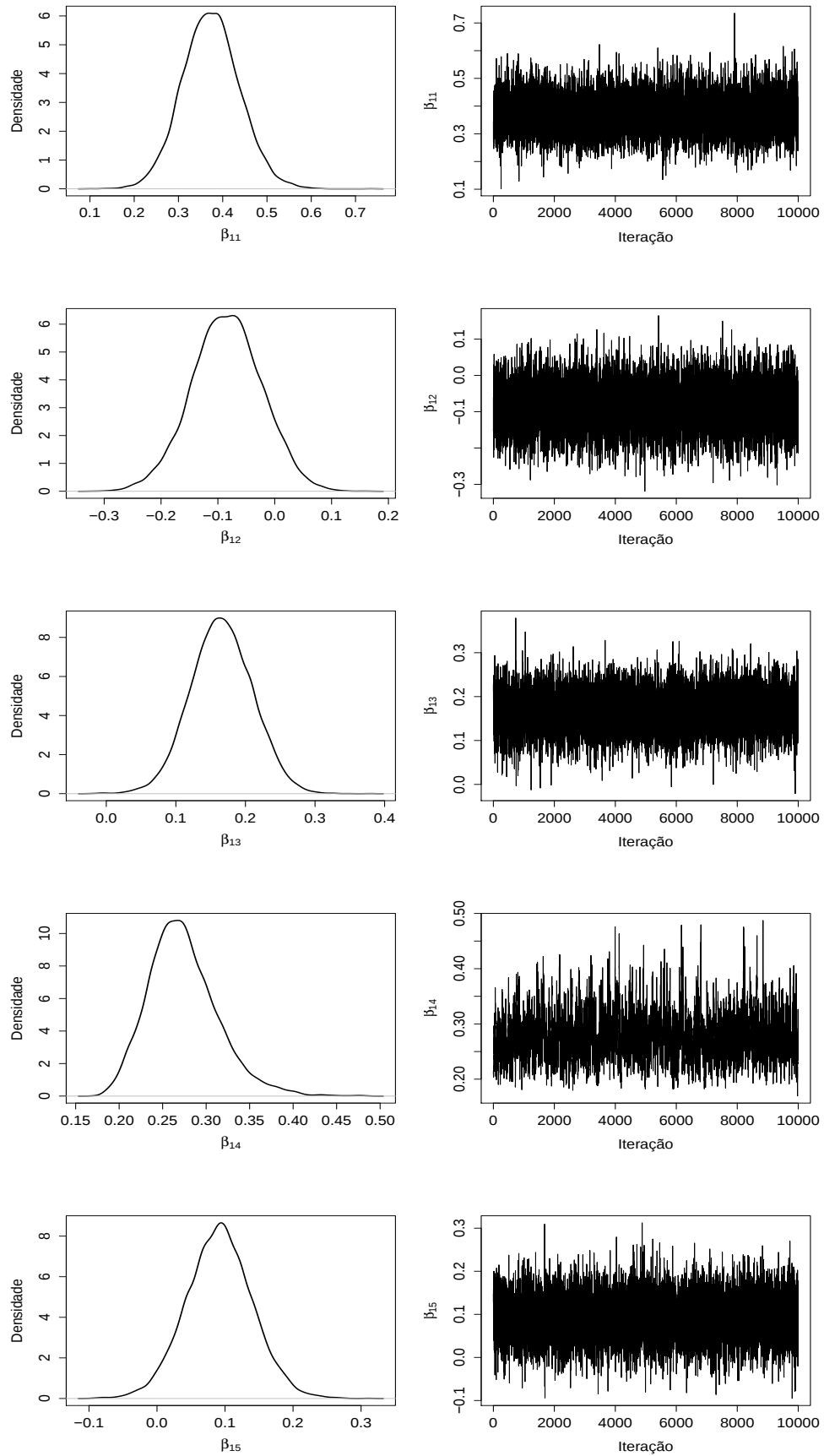


Figura 5.7: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros β_1 's no modelo 4.

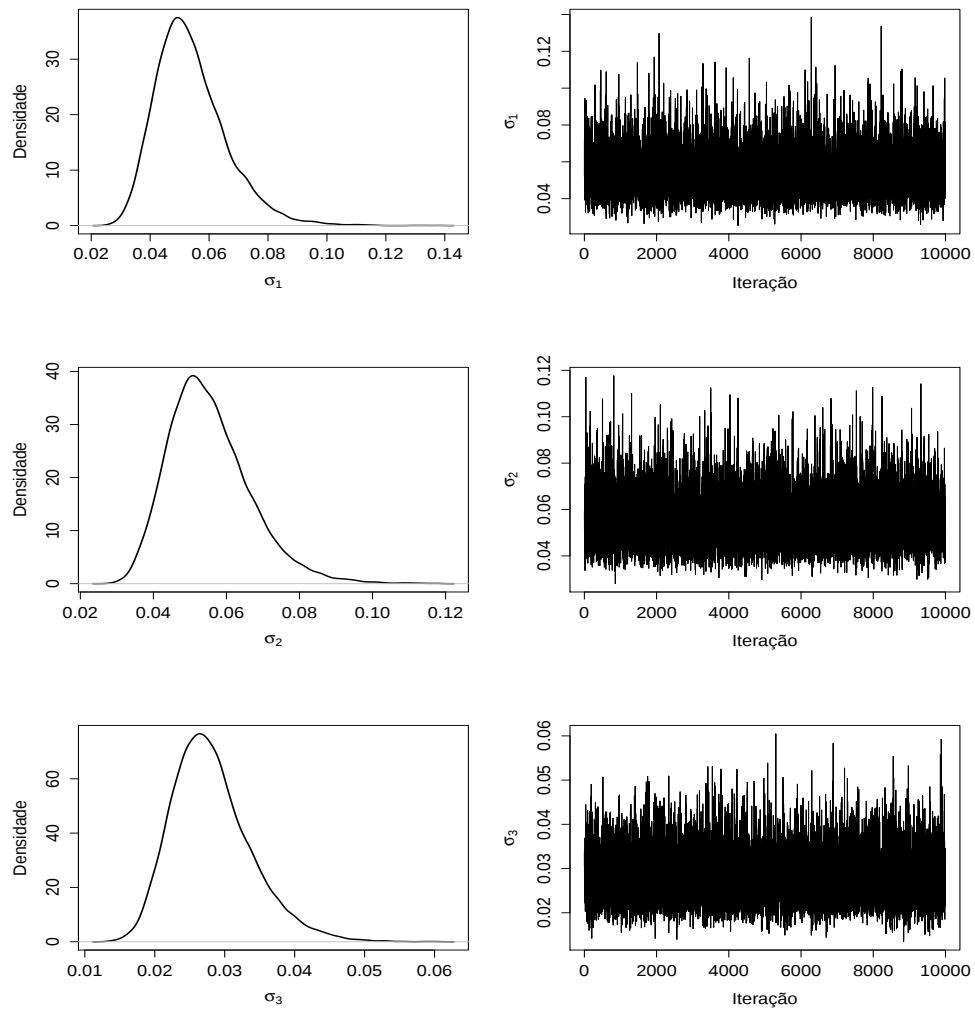


Figura 5.8: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros de variância no modelo 4.

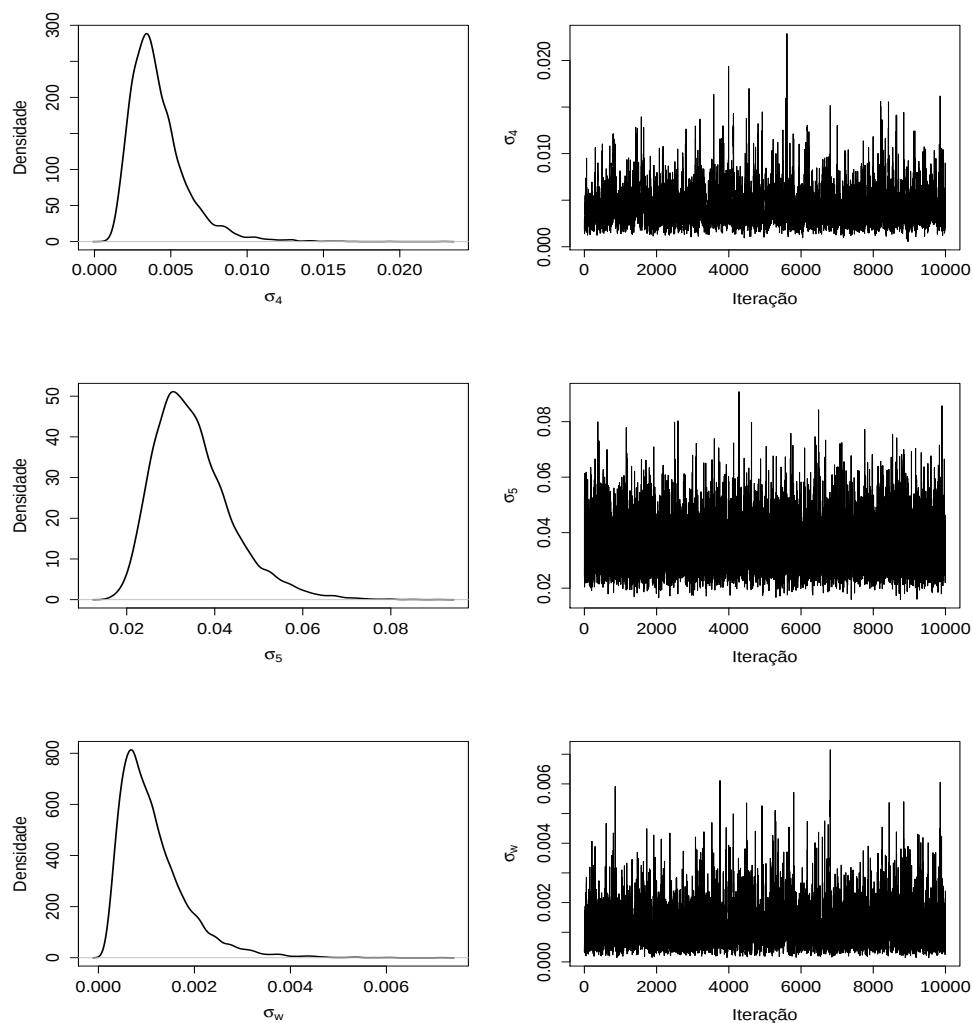


Figura 5.9: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros de variância no modelo 4.

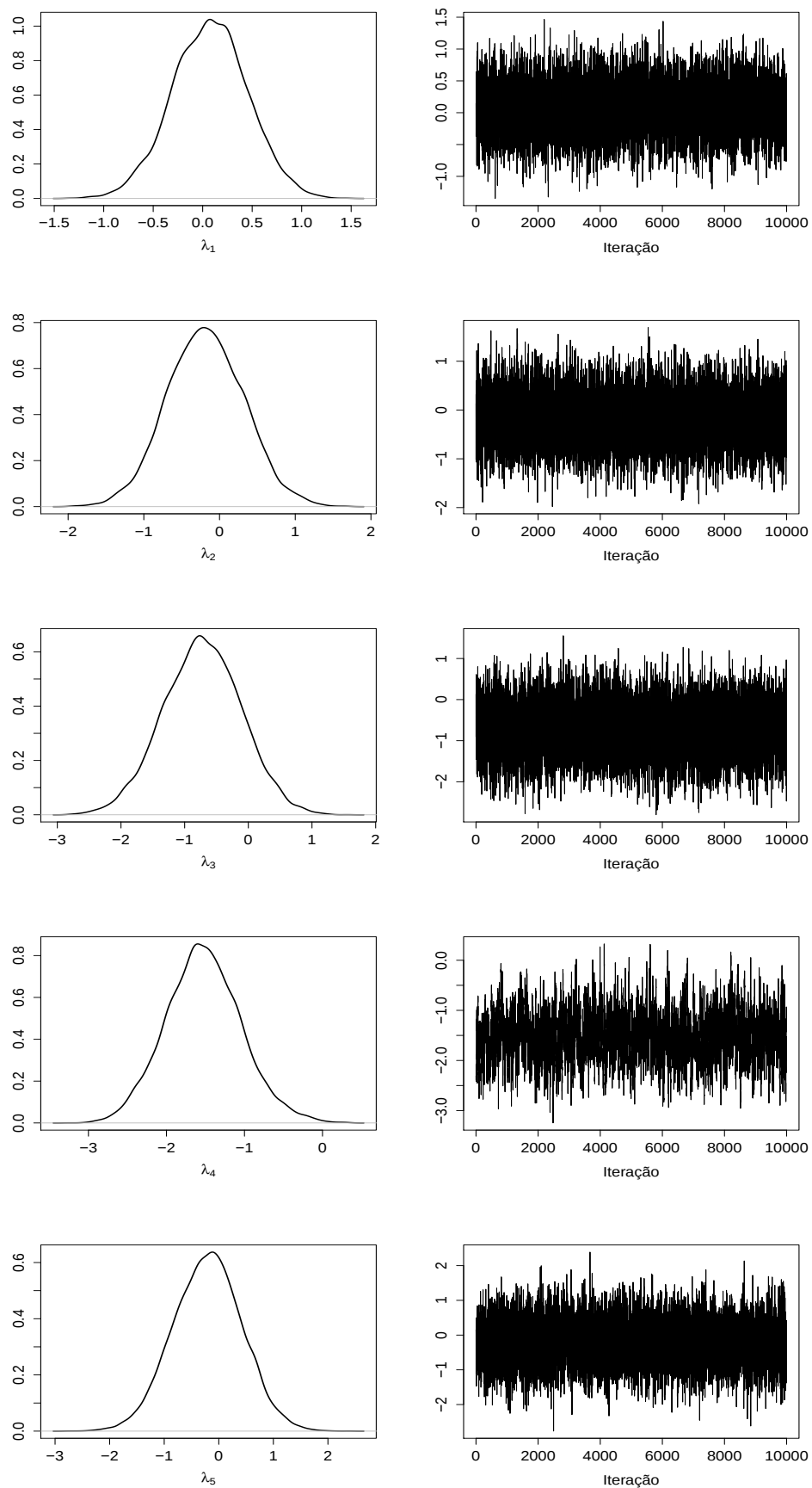


Figura 5.10: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros λ 's no modelo 4.

De acordo com os resultados da Tabela 5.3, observamos que as diferenças entre os dois grupos (antes e depois da dieta) dos componentes (pr/k), (ch/k) e (Ca/k) foram significativas, pois os intervalos de credibilidade para β_{11} , β_{13} e β_{14} não incluem o valor zero, levando ao mesmo resultado obtido através da transformação alr (Seção 5.1.1).

Podemos observar que apenas o λ_4 apresentou significância, indicando que para a maioria dos λ 's a transformação alr , que é um caso especial da transformação Box-Cox, seria mais adequada para os casos em que o intervalo de credibilidade incluem o valor zero ($\lambda = 0$).

Os valores das estatísticas de teste z_G do diagnóstico de convergência de Geweke são apresentados na Tabela 5.3, os gráficos das densidades *a posteriori* e o comportamento da cadeia gerada são apresentados nas Figuras 5.6, 5.7, 5.8, 5.9 e 5.10 para os parâmetros do modelo (5.13). Podemos observar que os traços *a posteriori* apresentaram certa uniformidade, indicando possível convergência, sendo confirmado através do diagnóstico de Geweke.

A Tabela 5.4 apresenta as estimativas das proporções verdadeiras para os componentes do leite: proteína, gordura do leite, carboidrato, cálcio, sódio e potássio para $z_i = 0$ (Grupo 1 - antes da dieta) e para $z_i = 1$ (Grupo 2 - depois da dieta), considerando a transformação Box-Cox com um efeito aleatório.

Tabela 5.4: Estimativas das proporções dos componentes - Modelo 4.

Componentes do leite	Proporções	$z = 0$ (Grupo 1)	Proporções	$z = 1$ (Grupo 2)
proteína	α_1	0,160	α_7	0,194
gordura do leite	α_2	0,183	α_8	0,142
carboidrato	α_3	0,159	α_9	0,158
cálcio	α_4	0,182	α_{10}	0,229
sódio	α_5	0,146	α_{11}	0,134
potássio	α_6	0,170	α_{12}	0,143

Percebe-se que os resultados da Tabela 5.4 são semelhantes aos resultados obtidos na Tabela 5.2, em que considerou a transformação alr com um efeito aleatório no modelo. Ou seja, as proporções dos componentes apresentam mesma tendência de aumentar ou diminuir após a dieta estabelecida no experimento.

5.3 Transformação *alr* Considerando Três Efeitos Aleatórios - Modelo 5

Nesta seção, consideramos a transformação *alr* para os dados composicionais (Apêndice B) e generalizando o modelo (5.1) com a introdução de dois efeitos aleatórios extras: v_{1i} que captura a dependência entre as medidas repetidas para o grupo 1 (antes de receber a dieta) para $i = 1, \dots, 30$ e v_{2i} que captura a dependência entre as medidas repetidas para o grupo 2 (depois de receber a dieta) para $i = 1, \dots, 30$. Desta forma, teremos três efeitos aleatórios no modelo, que é dado por

$$\mathbf{y}_i = \beta_0 + \beta_1 z_i + w_i + a_j v_{1i} + (1 - a_j) v_{2j} + \epsilon_i, \quad (5.26)$$

em que $a_j = 1$ (Grupo 1) e $a_j = 0$ (Grupo 2) para $j = 1, \dots, 5$. Assumimos a distribuição Normal $N(0, \sigma_{v1}^2)$ e $N(0, \sigma_{v2}^2)$, para v_{1i} e v_{2i} , respectivamente. Também assumimos as mesmas suposições consideradas para o modelo (5.1) e independência entre as quantidades aleatórias w_i , v_{1i} , v_{2i} e ϵ_{ji} , $i = 1, \dots, 60$; $j = 1, \dots, 5$. Denotamos o modelo definido em (5.26) como modelo 5.

Da mesma forma, que em (5.2), assumiremos $\epsilon_i \sim N(\mathbf{0}, \Sigma)$, em que $\mathbf{0}$ é o vetor de zeros e Σ é a matriz de variâncias e covariâncias.

Considerando o modelo apresentado em (5.13), assumiremos que os erros $\epsilon_{i1}, \dots, \epsilon_{i5}$ são mutuamente independentes. Assim, obtém-se,

$$\begin{aligned} \epsilon_{i1} &= y_{i1} - \beta_{01} - \beta_{11} z_i - w_i - a_j v_{1i} - (1 - a_j) v_{2j}, \\ &\vdots \\ \epsilon_{i5} &= y_{i5} - \beta_{05} - \beta_{15} z_i - w_i - a_j v_{1i} - (1 - a_j) v_{2j}. \end{aligned}$$

Assim,

$y_{ij} \sim N(\beta_{0j} + \beta_{1j} z_i, \sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)$, (Grupo 1 - antes da dieta);
 $y_{ij} \sim N(\beta_{0j} + \beta_{1j} z_i, \sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)$, (Grupo 2 - depois da dieta),
 para $j = 1, \dots, 5$ e $i = 1, \dots, 60$ ($n = n_1 + n_2$, em que $n_1 = 30$ para o Grupo 1 e $n_2 = 30$ para o Grupo 2).

De acordo com o resultado obtido em (5.5) dos estimadores das proporções verda-

deiras dos componentes, podemos estendê-lo para o modelo 5 (5.26) que considera a transformação *alr* com três efeitos aleatórios.

Desse modo, obtemos as estimativas das proporções verdadeiras dos componentes do leite, $\alpha_{i1}, \dots, \alpha_{i6}$.

Logo,

$$\begin{aligned} \alpha_{ij} &= \frac{e^{\beta_{0j} + \beta_{1j}z_i + w_i + a_jv_{1i} + (1-a_j)v_{2j}}}{\left(1 + e^{\beta_{01} + \beta_{11}z_i + w_i + a_jv_{1i} + (1-a_j)v_{2j}} + \dots + e^{\beta_{05} + \beta_{15}z_i + w_i + a_jv_{1i} + (1-a_j)v_{2j}}\right)}, \\ &\vdots \\ \alpha_{i6} &= \frac{1}{\left(1 + e^{\beta_{01} + \beta_{11}z_i + w_i + a_jv_{1i} + (1-a_j)v_{2j}} + \dots + e^{\beta_{05} + \beta_{15}z_i + w_i + a_jv_{1i} + (1-a_j)v_{2j}}\right)}, \end{aligned} \quad (5.27)$$

em que $i = 1, \dots, 60$ ($n = n_1 + n_2$, onde $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2) e $j = 1, \dots, 5$.

5.3.1 Análise Bayesiana - Modelo 5

A função densidade de probabilidade conjunta de $\mathbf{y}_1, \dots, \mathbf{y}_5$ dado o vetor de parâmetros $\mathbf{v}_5 = (\beta_{01}, \dots, \beta_{05}, \beta_{11}, \dots, \beta_{15}, \sigma_1^2, \dots, \sigma_5^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2)$ é dada por

$$\begin{aligned} f(\mathbf{y}_1, \dots, \mathbf{y}_5 | \mathbf{v}_5) &= \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)}} \exp \left\{ -\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} [y_{ij} - (\beta_{0j} + \beta_{1j}z_i)]^2 \right\} \\ &\times \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)}} \exp \left\{ -\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} [y_{ij} - (\beta_{0j} + \beta_{1j}z_i)]^2 \right\}, \end{aligned}$$

em que σ_{v1}^2 é a variância do efeito aleatório v_{1i} que captura a dependência entre as medidas repetidas do grupo 1 (antes da dieta) e σ_{v2}^2 é a variância do efeito aleatório v_{2i} que captura a dependência entre as medidas repetidas do grupo 2 (depois da dieta).

A função de verossimilhança para os parâmetros $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$, σ_w^2 , σ_{v1}^2 e σ_{v2}^2 do modelo (5.26) é dada por

$$\begin{aligned}
L(\beta_0, \beta_1, \boldsymbol{\sigma}^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2) &= \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} \sum_{i=1}^{30} \epsilon_{ij}^2 \right] \\
&\times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} \sum_{i=1}^{30} \epsilon_{ij}^2 \right],
\end{aligned} \tag{5.28}$$

em que

$$\sum_{i=1}^{30} \epsilon_{ij}^2 = \sum_{i=1}^{30} [y_{ij} - (\beta_{0j} + \beta_{1j} z_i)]^2.$$

Para a análise Bayesiana do modelo 5, consideramos as mesmas distribuições *a priori* dadas em (5.7) e (5.8). Para σ_{vl}^2 , assumimos a seguinte distribuição *a priori*,

$$\sigma_{vl}^2 \sim IG(a_{vl}, b_{vl}), \tag{5.29}$$

em que a_{vl} e b_{vl} são hiperparâmetros conhecidos, $l = 1, 2$. Assumimos independência *a priori* entre os parâmetros.

Desta forma, a partir do teorema de Bayes que associa a função de verossimilhança (5.28) com as distribuições *a priori* (5.7), (5.8) e (5.29), obtemos a distribuição *a posteriori* conjunta para os parâmetros $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \dots, \sigma_5^2)$, σ_w^2 , σ_{v1}^2 e σ_{v2}^2 dada por

$$\begin{aligned}
\pi(\beta_0, \beta_1, \boldsymbol{\sigma}^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2 | \mathbf{y}) &\propto \prod_{j=1}^5 \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \times \prod_{j=1}^5 \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \\
&\times \prod_{j=1}^5 (\sigma_j^2)^{-(c_j+1)} \exp \left(-\frac{d_j}{\sigma_j^2} \right) \times \prod_{j=1}^5 (\sigma_w^2)^{-(c_w+1)} \exp \left(-\frac{d_w}{\sigma_w^2} \right) \\
&\times \prod_{j=1}^5 (\sigma_{v1}^2)^{-(a_{v1}+1)} \exp \left(-\frac{b_{v1}}{\sigma_{v1}^2} \right) \times \prod_{j=1}^5 (\sigma_{v2}^2)^{-(a_{v2}+1)} \exp \left(-\frac{b_{v2}}{\sigma_{v2}^2} \right) \\
&\times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right] \\
&\times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right].
\end{aligned}$$

As distribuições *a posteriori* condicionais utilizadas para o amostrador de Gibbs (Gelfand e Smith, [20]) e para o Metropolis-Hastings são dadas por:

$$\begin{aligned}
i) \quad \pi(\beta_{0j} | \beta_1, \sigma^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2, \mathbf{y}) &\propto N(a_{0j}, b_{0j}^2) \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2)} \sum_{i=1}^n (\beta_{0j} - \mu_i^{(j)})^2 \right] \\
&\times \prod_{l=1}^2 \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n (\beta_{0j} - \mu_i^{(j)})^2 \right], \quad (5.30)
\end{aligned}$$

em que $\mu_i^{(j)} = y_{ij} - \beta_{1j} z_i$; $i = 1, \dots, n$, $j = 1, \dots, 5$ e $l = 1, 2$.

$$\begin{aligned}
ii) \quad \pi(\beta_{1j} | \beta_0, \sigma^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2, \mathbf{y}) &\propto N(a_{1j}, b_{1j}^2) \prod_{l=1}^2 \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n (\beta_{1j} z_i - \theta_i^{(j)})^2 \right], \\
&\quad (5.31)
\end{aligned}$$

em que $\theta_i^{(j)} = y_{ij} - \beta_{0j}$; $i = 1, \dots, n$, $j = 1, \dots, 5$ e $l = 1, 2$.

$$\begin{aligned}
iii) \quad \pi(\sigma_j^2 | \beta_0, \beta_1, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2, \mathbf{y}) &\propto (\sigma_j^2)^{-(c_j+1)} \exp \left(-\frac{d_j}{\sigma_j^2} \right) \\
&\times \prod_{l=1}^2 \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right], \quad (5.32)
\end{aligned}$$

para $i = 1, \dots, n$, $j = 1, \dots, 5$ e $l = 1, 2$.

$$\begin{aligned}
iv) \quad \pi(\sigma_w^2 | \beta_0, \beta_1, \sigma^2, \sigma_{v1}^2, \sigma_{v2}^2, \mathbf{y}) &\propto (\sigma_w^2)^{-(c_w+1)} \exp \left(-\frac{d_w}{\sigma_w^2} \right) \\
&\times \prod_{l=1}^2 (\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right], \\
&\quad (5.33)
\end{aligned}$$

para $i = 1, \dots, n$, $j = 1, \dots, 5$ e $l = 1, 2$.

$$\begin{aligned}
v) \quad \pi(\sigma_{v1}^2 | \beta_0, \beta_1, \boldsymbol{\sigma}^2, \sigma_w^2, \sigma_{v2}^2, \mathbf{y}) &\propto (\sigma_{v1}^2)^{-(a_{v1}+1)} \exp\left(-\frac{b_{v1}}{\sigma_{v1}^2}\right) \\
&\times (\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)^{-n/2} \exp\left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} \sum_{i=1}^n \epsilon_{ij}^2\right],
\end{aligned} \tag{5.34}$$

para $i = 1, \dots, n$ e $j = 1, \dots, 5$.

$$\begin{aligned}
vi) \quad \pi(\sigma_{v2}^2 | \beta_0, \beta_1, \boldsymbol{\sigma}^2, \sigma_w^2, \sigma_{v1}^2, \mathbf{y}) &\propto (\sigma_{v2}^2)^{-(a_{v2}+1)} \exp\left(-\frac{b_{v2}}{\sigma_{v2}^2}\right) \\
&\times (\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)^{-n/2} \exp\left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} \sum_{i=1}^n \epsilon_{ij}^2\right],
\end{aligned} \tag{5.35}$$

para $i = 1, \dots, n$ e $j = 1, \dots, 5$.

Podemos observar que as distribuições *a posteriori* condicionais (5.30), (5.31), (5.32), (5.33), (5.34) e (5.35) não apresentaram formas fechadas e, portanto, obtemos as estimativas dos parâmetros do modelo (5.26) através do algoritmo de Metropolis-Hastings.

Para a análise Bayesiana dos dados do Apêndice B, considera-se a transformação *alr* com erros normais não correlacionados para o modelo proposto (5.26).

Assumindo as distribuições *a priori* para $\beta_{0j}, \beta_{1j}, \sigma_j^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2$ com hiperparâmetros cujos valores foram $a_{0j} = 1$; $b_{0j} = b_{1j} = 1.000.000$, $a_{1j} = 0$, $c_j = d_j = a_w = b_w = a_{v1} = b_{v1} = a_{v2} = b_{v2} = 1.000$, geramos no *software* OPENBUGS 205.000 iterações, em que foram descartadas as 5000 iterações iniciais (“burn-in”). Após o período de “burn-in”, consideramos saltos de tamanho 20 para obtermos amostras não correlacionadas, na qual resultou uma amostra final de tamanho 10.000 para cada parâmetro. A convergência do algoritmo foi verificada pelos gráficos dos traços *a posteriori* das amostras de Gibbs simuladas e pelo diagnóstico de Geweke (Geweke, [23]).

Na Tabela 5.5 apresentamos os sumários *a posteriori* dos parâmetros do modelo (5.26) e os valores correspondentes da estatística de teste para o diagnóstico de convergência Geweke.

Dos resultados da Tabela 5.5 observamos que para os componentes do leite proteína, carboidrato e cálcio há diferenças significativas nos grupos antes e depois da dieta, sendo que os intervalos de credibilidade dos parâmetros β_{11} , β_{13} e β_{14} não incluem o valor zero.

As densidades *a posteriori* e o comportamento da cadeia gerada para os parâmetros do modelo (5.26) são apresentadas nas Figuras 5.11, 5.12, 5.13 e 5.14, em que observa-se que há uma certa uniformidade no comportamento das cadeias geradas.

A convergência dos algoritmos amostrador de Gibbs e Metropolis-Hastings foram verificadas utilizando o diagnóstico de Geweke para todos os parâmetros. Observamos que os valores da estatística do teste z_G estão no intervalo $(-1,96; 1,96)$ indicando convergência para todos os parâmetros estimados.

Tabela 5.5: Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke - Modelo 5.

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	-0,060	0,041	(-0,143; 0,021)	1,929
β_{02}	0,079	0,042	(-0,003; 0,159)	1,715
β_{03}	-0,058	0,030	(-0,177; 0,002)	1,582
β_{04}	0,070	0,020	(0,030; 0,111)	1,323
β_{05}	-0,150	0,036	(-0,221; -0,080)	-0,438
β_{11}	0,366	0,060	(0,249; 0,486)	-0,771
β_{12}	-0,082	0,060	(-0,199; 0,036)	-0,263
β_{13}	0,170	0,045	(0,081; 0,259)	-1,481
β_{14}	0,416	0,032	(0,354; 0,479)	-0,543
β_{15}	0,091	0,053	(-0,015; 0,196)	1,737
σ_1	0,049	0,010	(0,033; 0,072)	0,630
σ_2	0,050	0,010	(0,034; 0,073)	-0,255
σ_3	0,025	0,005	(0,016; 0,037)	-0,812
σ_4	0,010	0,003	(0,005; 0,017)	-0,499
σ_5	0,036	0,008	(0,024; 0,053)	0,761
σ_w	0,001	0,0009	(0,0003; 0,004)	0,635
σ_{v1}	0,001	0,0009	(0,0003; 0,004)	1,884
σ_{v2}	0,006	0,004	(0,001; 0,016)	-0,143

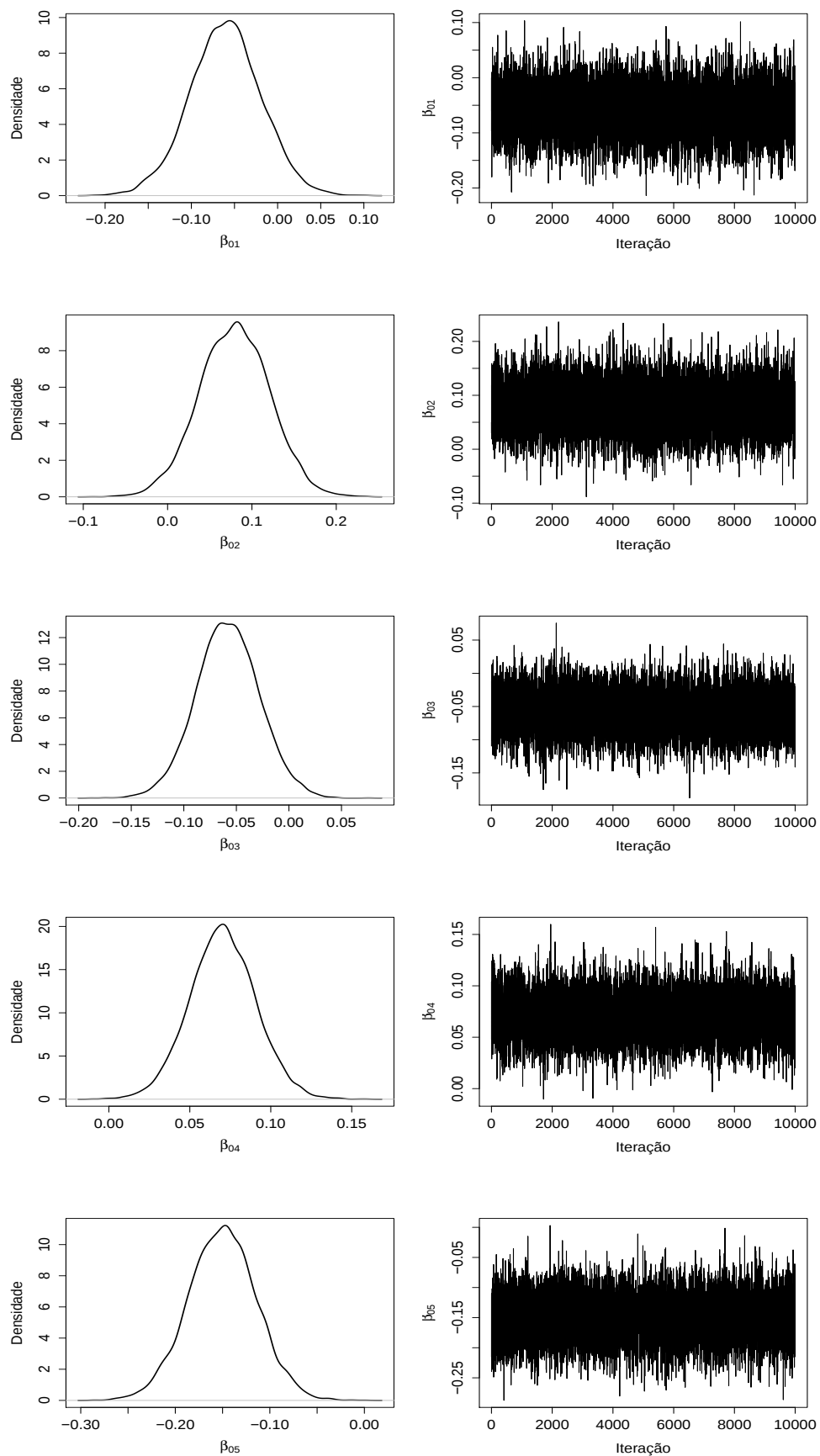


Figura 5.11: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros β_0 's no modelo 5.

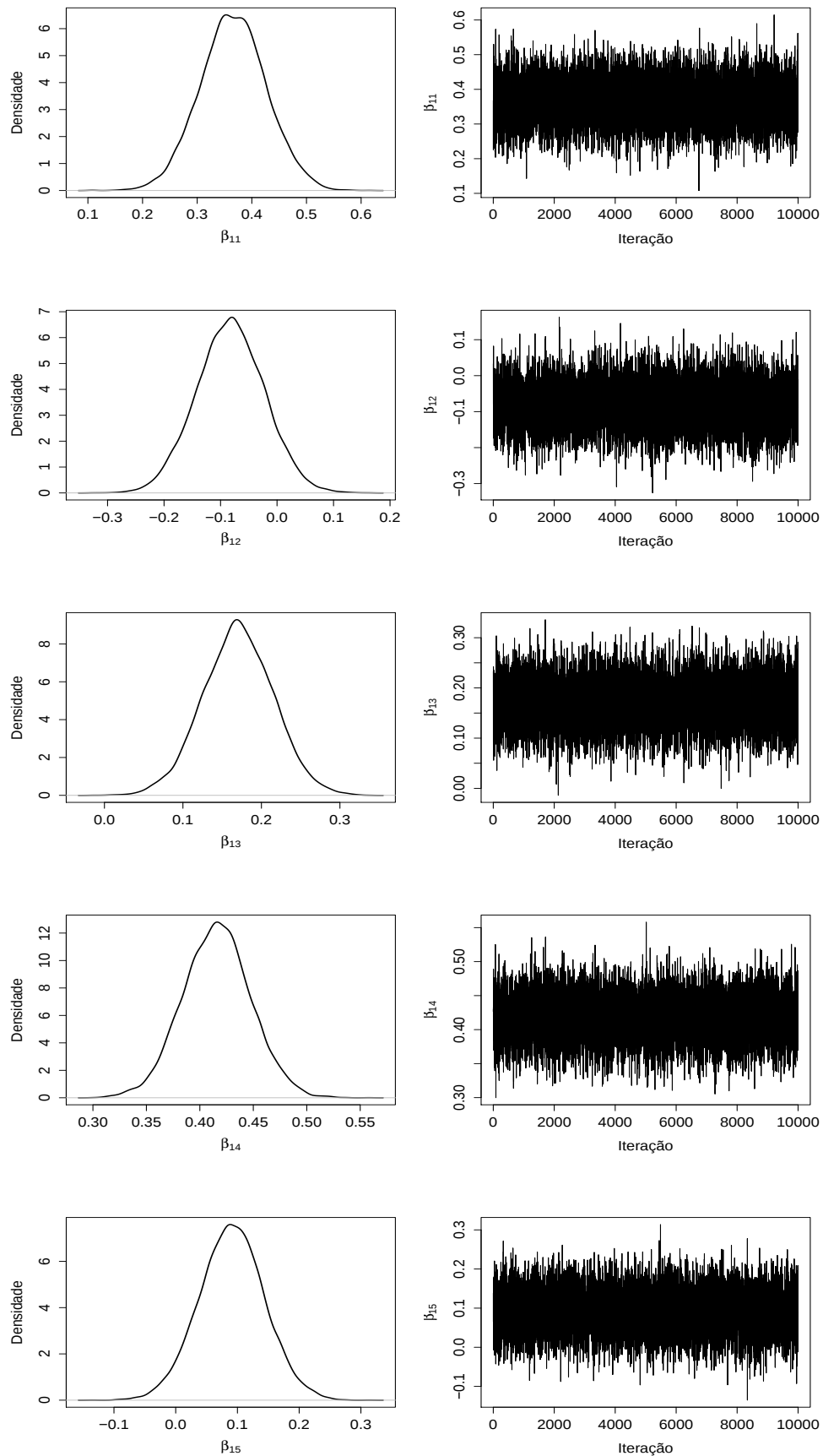


Figura 5.12: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros β_1 's no modelo 5.

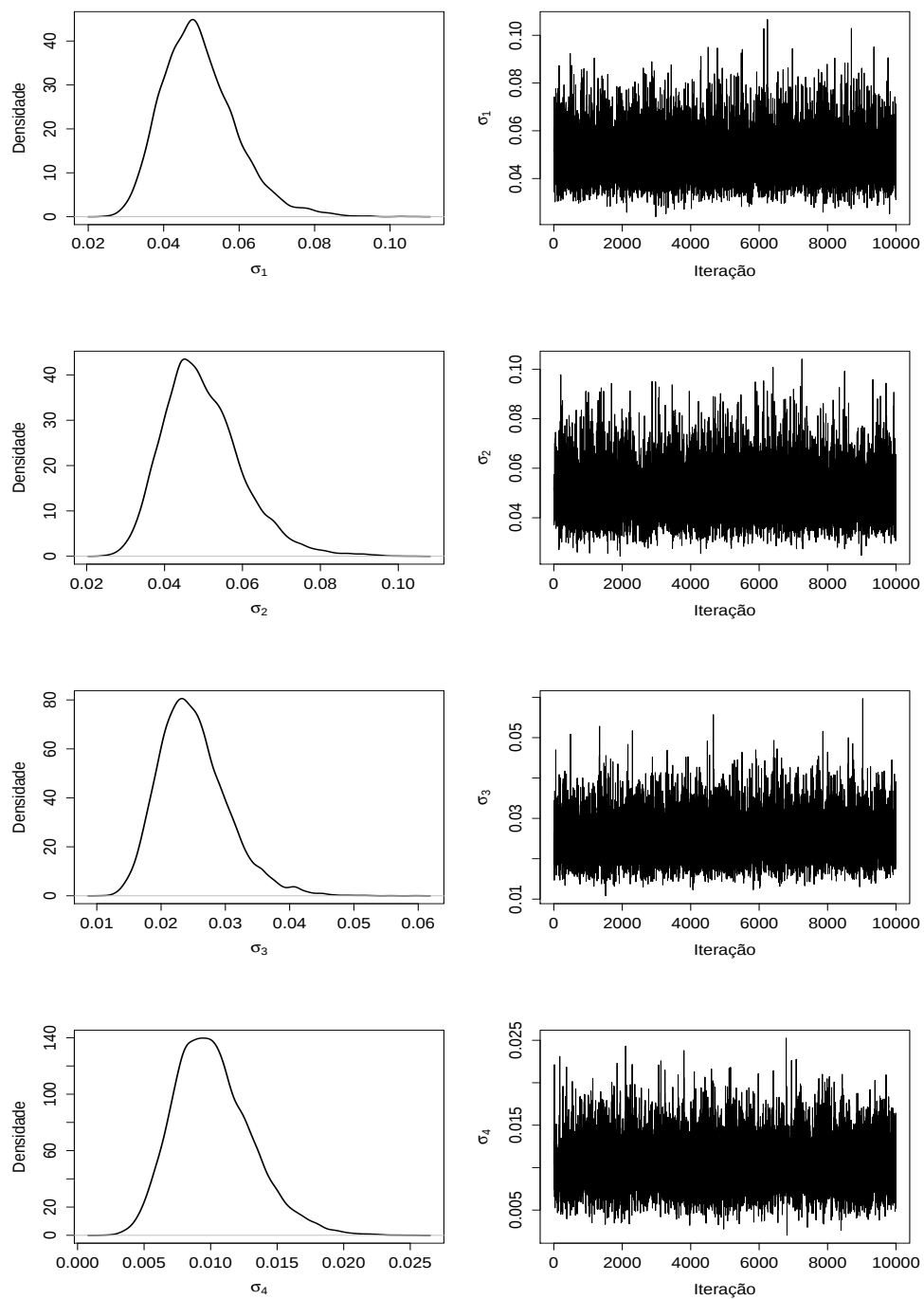


Figura 5.13: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros de variância no modelo 5.

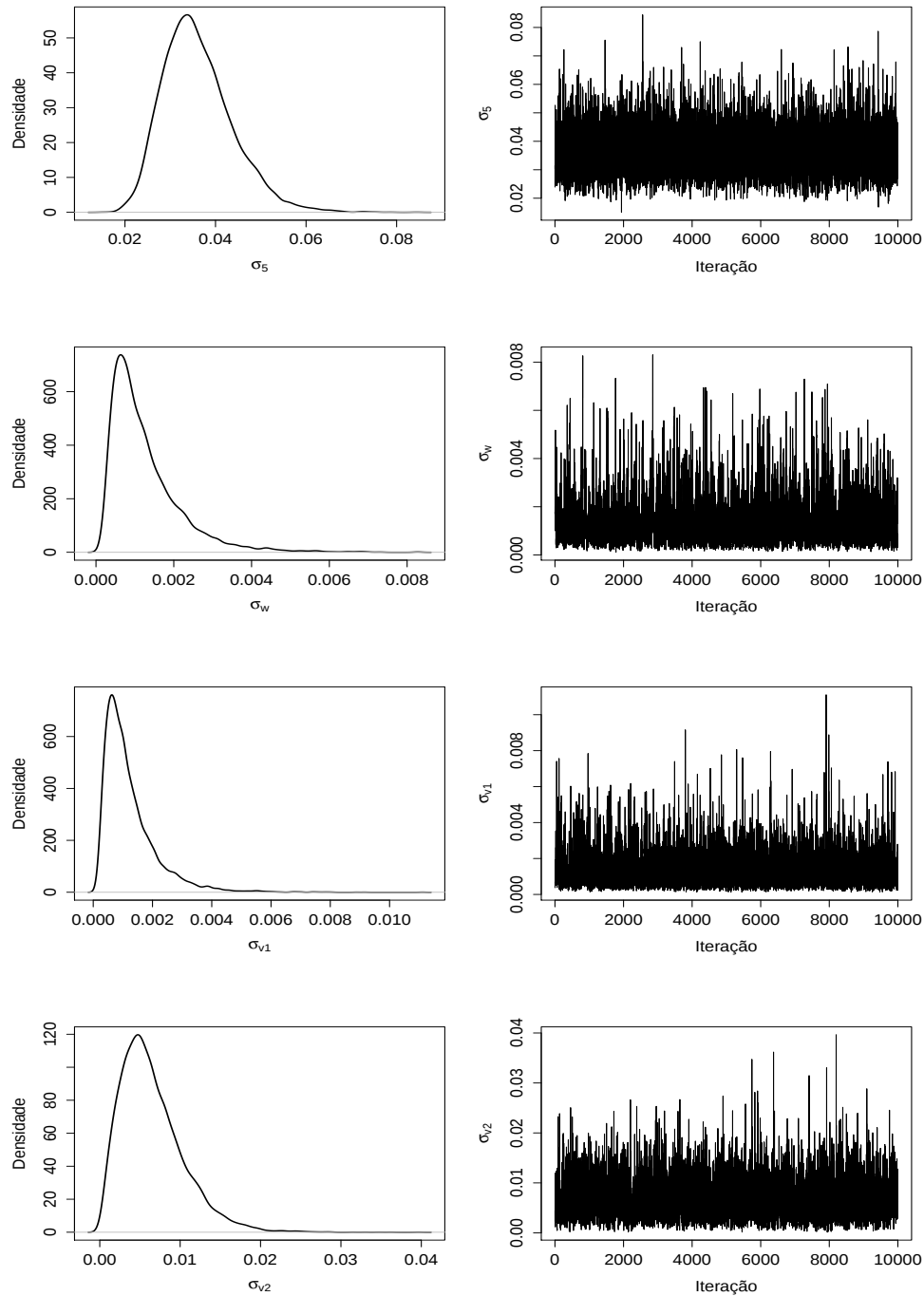


Figura 5.14: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros de variância no modelo 5.

A Tabela 5.6 apresenta as estimativas das proporções verdadeiras para os componentes de proteína, gordura do leite, carboidrato, cálcio, sódio e potássio para $z_i = 0$ (Grupo 1 - antes da dieta) e para $z_i = 1$ (Grupo 2 - depois da dieta), considerando a transformação *alr* com três efeitos aleatórios.

Tabela 5.6: Estimativas das proporções dos componentes - Modelo 5.

Componentes do leite	Proporções	$z = 0$ (Grupo 1)	Proporções	$z = 1$ (Grupo 2)
proteína	α_1	0,160	α_7	0,193
gordura do leite	α_2	0,183	α_8	0,142
carboidrato	α_3	0,160	α_9	0,159
cálcio	α_4	0,182	α_{10}	0,231
sódio	α_5	0,146	α_{11}	0,134
potássio	α_6	0,169	α_{12}	0,142

Mais uma vez, percebe-se que os resultados da Tabela 5.6 são semelhantes aos resultados obtidos nas Tabelas 5.2 e 5.4, considerando as transformações *alr* e Box-Cox com um efeito aleatório no modelo, respectivamente.

5.4 Transformação Box-Cox Considerando Três Efeitos Aleatórios - Modelo 6

Consideramos uma aplicação da transformação Box-Cox (3.3) nos dados composicionais do Apêndice B, x_{i1} , x_{i2} , x_{i3} , x_{i4} , x_{i5} e x_{i6} para $i = 1, \dots, 60$.

Conforme apresentado na Seção 5.2, obtemos $y_{i1}^{(\lambda_1)}, \dots, y_{i5}^{(\lambda_5)}$ através da definição da transformação de Box-Cox (3.3).

Para os dados transformados $y_{i1}^{(\lambda_1)}, \dots, y_{i5}^{(\lambda_5)}$, podemos definir o modelo 6 como

$$\mathbf{y}_i^{(\boldsymbol{\lambda})} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_i + w_i + a_j v_{1i} + (1 - a_j) v_{2j} + \boldsymbol{\epsilon}_i, \quad (5.36)$$

em que $a_j = 1$ (Grupo 1) e $a_j = 0$ (Grupo 2) para $j = 1, \dots, 5$. Assumimos a distribuição Normal $N(0, \sigma_{v1}^2)$ e $N(0, \sigma_{v2}^2)$, para v_{1i} e v_{2i} , respectivamente. Também assumimos as mesmas suposições consideradas para o modelo (5.1) e independência entre as quantidades aleatórias w_i , v_{1i} , v_{2i} e ϵ_{ji} , $i = 1, \dots, 60$; $j = 1, \dots, 5$.

Considerando o modelo apresentado em (5.36), assumiremos que os erros $\epsilon_{i1}, \dots, \epsilon_{i5}$ são mutuamente independentes. Assim, obtém-se,

$$\begin{aligned}
\epsilon_{i1} &= y_{i1}^{(\lambda_1)} - \beta_{01} - \beta_{11}z_i - w_i - a_j v_{1i} - (1 - a_j)v_{2j}, \\
&\vdots \\
\epsilon_{i5} &= y_{i5}^{(\lambda_5)} - \beta_{05} - \beta_{15}z_i - w_i - a_j v_{1i} - (1 - a_j)v_{2j},
\end{aligned}$$

$y_{ij}^{(\lambda_j)} \sim N(\beta_{0j} + \beta_{1j}z_i; \sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)$, (Grupo 1 - antes da dieta); e

$y_{ij}^{(\lambda_j)} \sim N(\beta_{0j} + \beta_{1j}z_i; \sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)$, (Grupo 2 - depois da dieta),

para $j = 1, \dots, 5$ e $i = 1, \dots, 60$ ($n = n_1 + n_2$, onde $n_1 = 30$ para o Grupo 1 e $n_2 = 30$ para o Grupo 2).

De acordo com os resultados obtidos em (5.17) e (5.27) dos estimadores das proporções verdadeiras dos componentes, podemos estendê-los para o modelo 6 (5.36) que considera a transformação Box-Cox com três efeitos aleatórios. Desse modo, obtemos os estimadores das proporções verdadeiras dos componentes do leite, $\alpha_{i1}, \dots, \alpha_{i6}$.

Logo,

$$\begin{aligned}
\alpha_{ij} &= \frac{[\lambda_j(\beta_{0j} + \beta_{1j}z_i + w_i + a_j v_{1i} + (1 - a_j)v_{2j}) + 1]^{\frac{1}{\lambda_j}}}{\left\{1 + [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i + a_j v_{1i} + (1 - a_j)v_{2j}) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i + a_j v_{1i} + (1 - a_j)v_{2j}) + 1]^{\frac{1}{\lambda_5}}\right\}}, \\
&\vdots \\
\alpha_{i6} &= \frac{1}{\left\{1 + [\lambda_1(\beta_{01} + \beta_{11}z_i + w_i + a_j v_{1i} + (1 - a_j)v_{2j}) + 1]^{\frac{1}{\lambda_1}} + \dots + [\lambda_5(\beta_{05} + \beta_{15}z_i + w_i + a_j v_{1i} + (1 - a_j)v_{2j}) + 1]^{\frac{1}{\lambda_5}}\right\}},
\end{aligned}$$

em que $a_j = 1$ (Grupo 1 - antes da dieta), $a_j = 0$ (Grupo 2 - depois da dieta) para $j = 1, \dots, 5$ e $i = 1, \dots, 60$ ($n = n_1 + n_2$, em que $n_1 = 30$ para o grupo 1 e $n_2 = 30$ para o grupo 2).

5.4.1 Análise Bayesiana - Modelo 6

A função densidade de probabilidade conjunta para os dados transformados $\mathbf{y}_1^{(\lambda_1)}, \dots, \mathbf{y}_5^{(\lambda_5)}$ dado o vetor de parâmetros $\mathbf{v}_6 = (\beta_{01}, \dots, \beta_{05}, \beta_{11}, \dots, \beta_{15}, \sigma_1^2, \dots, \sigma_5^2, \lambda_1, \dots, \lambda_5, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2)$ é dada por

$$f(\mathbf{y}_1^{(\lambda_1)}, \dots, \mathbf{y}_5^{(\lambda_5)} | \mathbf{v}_6) = \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)}} \exp \left\{ -\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} [y_{ij} - (\beta_{0j} + \beta_{1j}z_i)]^2 \right\} \\ \times \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)}} \exp \left\{ -\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} [(y_{ij} - (\beta_{0j} + \beta_{1j}z_i))]^2 \right\}$$

Por outro lado, a função densidade de probabilidade das observações não transformadas (variáveis originais) [Box & Tiao [9]], $y_{i1} = x_{i1}/x_{i6}$, $y_{i2} = x_{i2}/x_{i6}$, $y_{i3} = x_{i3}/x_{i6}$, $y_{i4} = x_{i4}/x_{i6}$ e $y_{i5} = x_{i5}/x_{i6}$ é dada por

$$f(\mathbf{y}_1, \dots, \mathbf{y}_5 | \mathbf{v}_6) = \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)}} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} (y_{ij}^{(\lambda_j)} - \beta_{0j} - \beta_{1j}z_i)^2 \right] \\ \times \prod_{j=1}^5 \frac{1}{\sqrt{2\pi(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)}} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} (y_{ij}^{(\lambda_j)} - \beta_{0j} - \beta_{1j}z_i)^2 \right] \\ \times \prod_{j=1}^5 \left(\frac{dy_{ij}^{(\lambda_j)}}{dy_{ij}} \right),$$

em que para $j = 1, \dots, 5$, $\frac{dy_{ij}^{(\lambda_j)}}{dy_{ij}} = \frac{\lambda_j y_{ij}^{\lambda_j-1}}{\lambda_j} = y_{ij}^{\lambda_j-1}$ são os elementos da diagonal da matriz Jacobiano.

A função de verossimilhança para os parâmetros $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2)$, $\lambda = (\lambda_1, \dots, \lambda_5)$, σ_w^2 , σ_{v1}^2 e σ_{v2}^2 do modelo (5.36) é dada por

$$L(\beta_0, \beta_1, \sigma^2, \lambda, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2) = \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} \sum_{i=1}^{30} \epsilon_{ij}^2 \right] \\ \times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} \sum_{i=1}^{30} \epsilon_{ij}^2 \right] \\ \times \left(\prod_{i=1}^n y_{ij}^{\lambda_j-1} \right), \quad (5.37)$$

em que

$$\sum_{i=1}^{30} \epsilon_{ij}^2 = \sum_{i=1}^{30} (y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j}z_i))^2.$$

Para a análise Bayesiana do modelo (5.36), consideramos as mesmas distribuições *a priori* dadas em (5.7), (5.8) e (5.29) e a função de verossimilhança (5.37). Dessa

forma, a distribuição *a posteriori* conjunta para os parâmetros $\beta_0 = (\beta_{01}, \dots, \beta_{05})$, $\beta_1 = (\beta_{11}, \dots, \beta_{15})$, $\sigma^2 = (\sigma_1^2, \dots, \sigma_5^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2)$ e $\lambda = (\lambda_1, \dots, \lambda_5)$ é dada por

$$\begin{aligned} \pi(\beta_0, \beta_1, \lambda, \sigma^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2 | \mathbf{y}^{(\lambda)}) &\propto \prod_{j=1}^5 \exp \left[-\frac{1}{2b_{0j}^2} (\beta_{0j} - a_{0j})^2 \right] \prod_{j=1}^5 \exp \left[-\frac{1}{2b_{1j}^2} (\beta_{1j} - a_{1j})^2 \right] \\ &\times \prod_{j=1}^5 (\sigma_j^2)^{-(c_j+1)} \exp \left(-\frac{d_j}{\sigma_j^2} \right) \prod_{j=1}^5 (\sigma_w^2)^{-(c_w+1)} \exp \left(-\frac{d_w}{\sigma_w^2} \right) \\ &\times \prod_{j=1}^5 (\sigma_{v1}^2)^{-(a_{v1}+1)} \exp \left(-\frac{b_{v1}}{\sigma_{v1}^2} \right) \prod_{j=1}^5 (\sigma_{v2}^2)^{-(a_{v2}+1)} \exp \left(-\frac{b_{v2}}{\sigma_{v2}^2} \right) \\ &\times \prod_{j=1}^5 \exp \left[-\frac{1}{2f_j^2} (\lambda_j - e_j)^2 \right] \times \left(\prod_{j=1}^5 \prod_{i=1}^n y_{ij}^{\lambda_j-1} \right) \\ &\times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right] \\ &\times \prod_{j=1}^5 (\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)^{-n/2} \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} \sum_{i=1}^n \epsilon_{ij}^2 \right]. \end{aligned}$$

As distribuições *a posteriori* condicionais são dadas por:

$$\begin{aligned} i) \quad \pi(\beta_{0j} | \beta_1, \sigma^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto N(a_{0j}, b_{0j}^2) \\ &\times \prod_{l=1}^2 \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n (\beta_{0j} - \mu_i^{(j)})^2 \right], \end{aligned} \quad (5.38)$$

em que $\mu_i^{(j)} = y_{ij}^{(\lambda_j)} - \beta_{1j} z_i$; $i = 1, \dots, n$, $j = 1, \dots, 5$ e $l = 1, 2$.

$$\begin{aligned} ii) \quad \pi(\beta_{1j} | \beta_0, \sigma^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto N(a_{1j}, b_{1j}^2) \\ &\times \prod_{l=1}^2 \exp \left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n (\beta_{1j} z_i - \theta_i^{(j)})^2 \right], \end{aligned} \quad (5.39)$$

em que $\theta_i^{(j)} = y_{ij}^{(\lambda_j)} - \beta_{0j}$; $i = 1, \dots, n$, $j = 1, \dots, 5$ e $l = 1, 2$.

$$\begin{aligned}
iii) \quad \pi(\sigma_j^2 | \beta_0, \beta_1, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto (\sigma_j^2)^{-(c_j+1)} \exp\left(-\frac{d_j}{\sigma_j^2}\right) \\
&\times \prod_{l=1}^2 \exp\left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n \epsilon_{ij}^2\right], \quad (5.40)
\end{aligned}$$

para $i = 1, \dots, n$, $j = 1, \dots, 5$ e $l = 1, 2$.

$$\begin{aligned}
iv) \quad \pi(\sigma_w^2 | \beta_0, \beta_1, \sigma^2, \sigma_{v1}^2, \sigma_{v2}^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto (\sigma_w^2)^{-(c_w+1)} \exp\left(-\frac{d_w}{\sigma_w^2}\right) \\
&\times \prod_{l=1}^2 (\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)^{-n/2} \exp\left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n \epsilon_{ij}^2\right], \quad (5.41)
\end{aligned}$$

para $i = 1, \dots, n$, $j = 1, \dots, 5$ e $l = 1, 2$.

$$\begin{aligned}
v) \quad \pi(\sigma_{v1}^2 | \beta_0, \beta_1, \sigma^2, \sigma_w^2, \sigma_{v2}^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto (\sigma_{v1}^2)^{-(a_{v1}+1)} \exp\left(-\frac{b_{v1}}{\sigma_{v1}^2}\right) \\
&\times (\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)^{-n/2} \exp\left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v1}^2)} \sum_{i=1}^n \epsilon_{ij}^2\right], \quad (5.42)
\end{aligned}$$

para $i = 1, \dots, n$ e $j = 1, \dots, 5$ (Grupo 1 - antes da dieta).

$$\begin{aligned}
vi) \quad \pi(\sigma_{v2}^2 | \beta_0, \beta_1, \sigma^2, \sigma_w^2, \sigma_{v1}^2, \lambda, \mathbf{y}^{(\lambda)}) &\propto (\sigma_{v2}^2)^{-(a_{v2}+1)} \exp\left(-\frac{b_{v2}}{\sigma_{v2}^2}\right) \\
&\times (\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)^{-n/2} \exp\left[-\frac{1}{2(\sigma_j^2 + \sigma_w^2 + \sigma_{v2}^2)} \sum_{i=1}^n \epsilon_{ij}^2\right], \quad (5.43)
\end{aligned}$$

para $i = 1, \dots, n$ e $j = 1, \dots, 5$ (Grupo 2 - depois da dieta).

$$vii) \quad \pi(\lambda_j | \beta_0, \beta_1, \sigma^2, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2, \mathbf{y}^{(\lambda)}) \propto N(e_j, f_j^2) \prod_{l=1}^2 \psi_j^{(l)}(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \sigma_{vl}^2, \lambda), \quad (5.44)$$

em que

$$\begin{aligned} \psi_j^{(l)}(\beta_0, \beta_1, \sigma^2, \sigma_w^2, \sigma_{vl}^2, \lambda) &= \exp \left[(\lambda_j - 1) \sum_{i=1}^n \ln y_{ij} + \frac{1}{(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \sum_{i=1}^n \left(y_{ij}^{(\lambda_j)} - (\beta_{0j} + \beta_{1j}) \right)^2 \right] \\ &= \exp \left[(\lambda_j - 1) \sum_{i=1}^n \ln y_{ij} + \frac{1}{(\sigma_j^2 + \sigma_w^2 + \sigma_{vl}^2)} \left(\sum_{i=1}^n y_{ij}^{(\lambda_j)^2} - 2\beta_{0j} \sum_{i=1}^n y_{ij}^{(\lambda_j)} - 2\beta_{1j} \sum_{i=1}^n y_{ij}^{(\lambda_j)} z_i \right) \right], \end{aligned}$$

para $j = 1, \dots, 5$ e $l = 1, 2$.

Podemos observar que as distribuições *a posteriori* condicionais (5.38), (5.39), (5.40), (5.41), (5.42), (5.43) e (5.44) não apresentaram formas fechadas e, portanto, obtemos as estimativas dos parâmetros do modelo (5.36) através do algoritmo de Metropolis-Hastings.

Para a análise Bayesiana dos dados do Apêndice B, considera-se a transformação Box-Cox com erros normais não correlacionados para o modelo proposto (5.36).

Assumindo a transformação Box-Cox para o modelo (5.36), e as distribuições *a priori* para $\beta_{0j}, \beta_{1j}, \sigma_j^2, \lambda_j, \sigma_w^2, \sigma_{v1}^2, \sigma_{v2}^2$ com hiperparâmetros cujos valores foram $a_{0j} = 1$; $b_{0j} = b_{1j} = f_j = 1.000.000$, $a_{1j} = e_j = 0$ e $c_j = d_j = a_w = b_w = a_{v1} = b_{v1} = a_{v2} = b_{v2} = 1.000$, geramos no *software* OPENBUGS 205.000 iterações, nas quais foram descartadas as 5.000 iterações iniciais (“burn-in”). Após o período de “burn-in”, consideramos saltos de tamanho 20 para obtermos amostras não correlacionadas, na qual resultou uma amostra final de tamanho 10.000 para cada parâmetro. A convergência do algoritmo foi verificada pelos gráficos dos traços *a posteriori* das amostras de Gibbs simuladas e pelo diagnóstico de Geweke (Geweke, [23]).

Na Tabela 5.7 apresentamos os sumários *a posteriori* dos parâmetros do modelo (5.26) e os valores correspondentes da estatística de teste para o diagnóstico de convergência Geweke.

Tabela 5.7: Resumos *a posteriori* e Estatística de Teste para o Diagnóstico de Geweke - Modelo 6.

Parâmetro	Média	Desvio Padrão	Intervalo de Credibilidade (95%)	z_G
β_{01}	-0,059	0,044	(-0,145; 0,029)	0,369
β_{02}	0,075	0,045	(-0,013; 0,165)	1,762
β_{03}	-0,067	0,032	(-0,129; -0,004)	-0,751
β_{04}	0,066	0,015	(0,037; 0,097)	0,166
β_{05}	-0,158	0,040	(-0,241; -0,083)	-0,986
β_{11}	0,371	0,066	(0,244; 0,502)	-1,649
β_{12}	-0,084	0,062	(-0,205; 0,038)	-0,776
β_{13}	0,166	0,045	(0,079; 0,254)	-0,638
β_{14}	0,304	0,047	(0,218; 0,406)	-0,097
β_{15}	0,091	0,053	(-0,012; 0,197)	-0,629
σ_1	0,053	0,012	(0,034; 0,081)	-0,240
σ_2	0,053	0,011	(0,036; 0,079)	1,346
σ_3	0,027	0,006	(0,018; 0,040)	0,812
σ_4	0,004	0,002	(0,001; 0,010)	0,764
σ_5	0,039	0,010	(0,023; 0,063)	-0,873
σ_w	0,0009	0,0006	(0,0002; 0,002)	-0,042
σ_{v1}	0,0009	0,0006	(0,0002; 0,002)	0,720
σ_{v2}	0,003	0,002	(0,0005; 0,009)	-0,125
λ_1	0,066	0,383	(-0,684; 0,826)	0,130
λ_2	-0,163	0,507	(-1,143; 0,839)	1,713
λ_3	-0,608	0,604	(-1,818; 0,549)	0,542
λ_4	-1,114	0,491	(-2,143; -0,170)	0,294
λ_5	-0,328	0,647	(-1,623; 0,920)	-0,209

As densidades *a posteriori* e o comportamento da cadeia gerada para os parâmetros do modelo (5.36) são apresentadas nas Figuras 5.15, 5.16, 5.17, 5.18 e 5.19. Além disso, a convergência do algoritmo foi verificada utilizando o diagnóstico de Geweke. Sendo assim, podemos observar que os valores da estatística do teste z_G estão no intervalo $(-1,96; 1,96)$ indicando convergência para todos os parâmetros estimados.

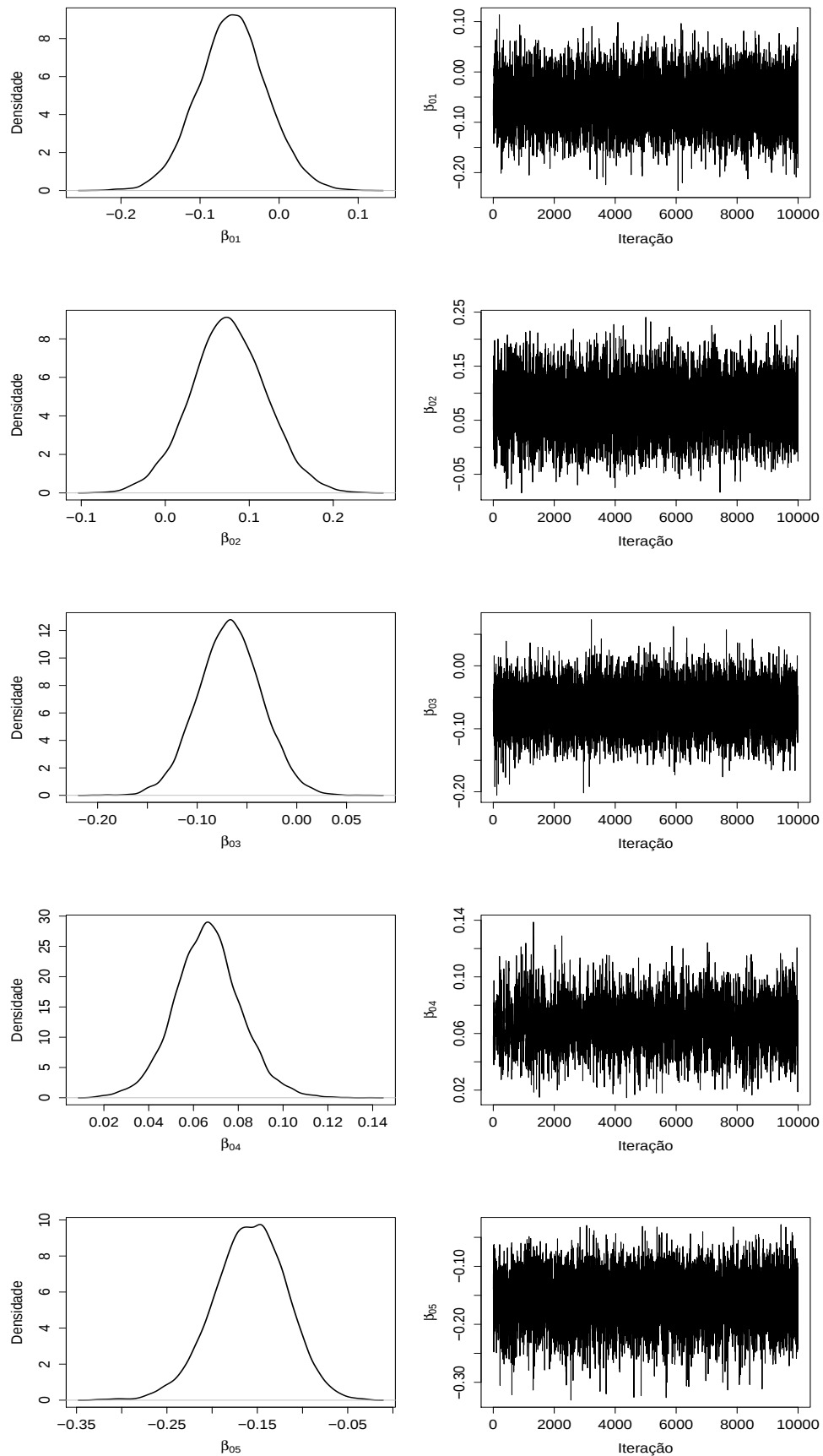


Figura 5.15: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros β_0 's no modelo 6.

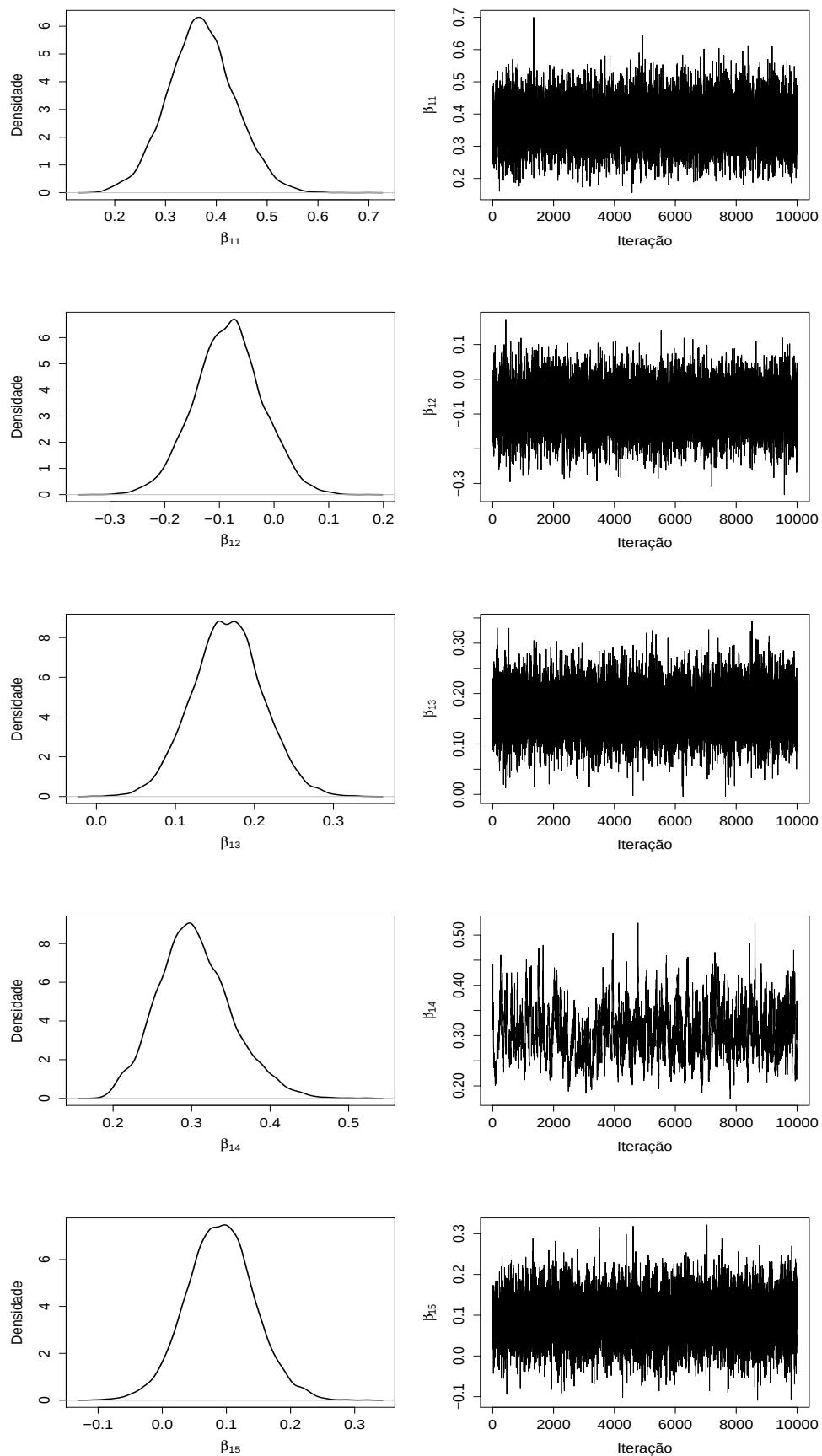


Figura 5.16: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros β_1 's no modelo 6.

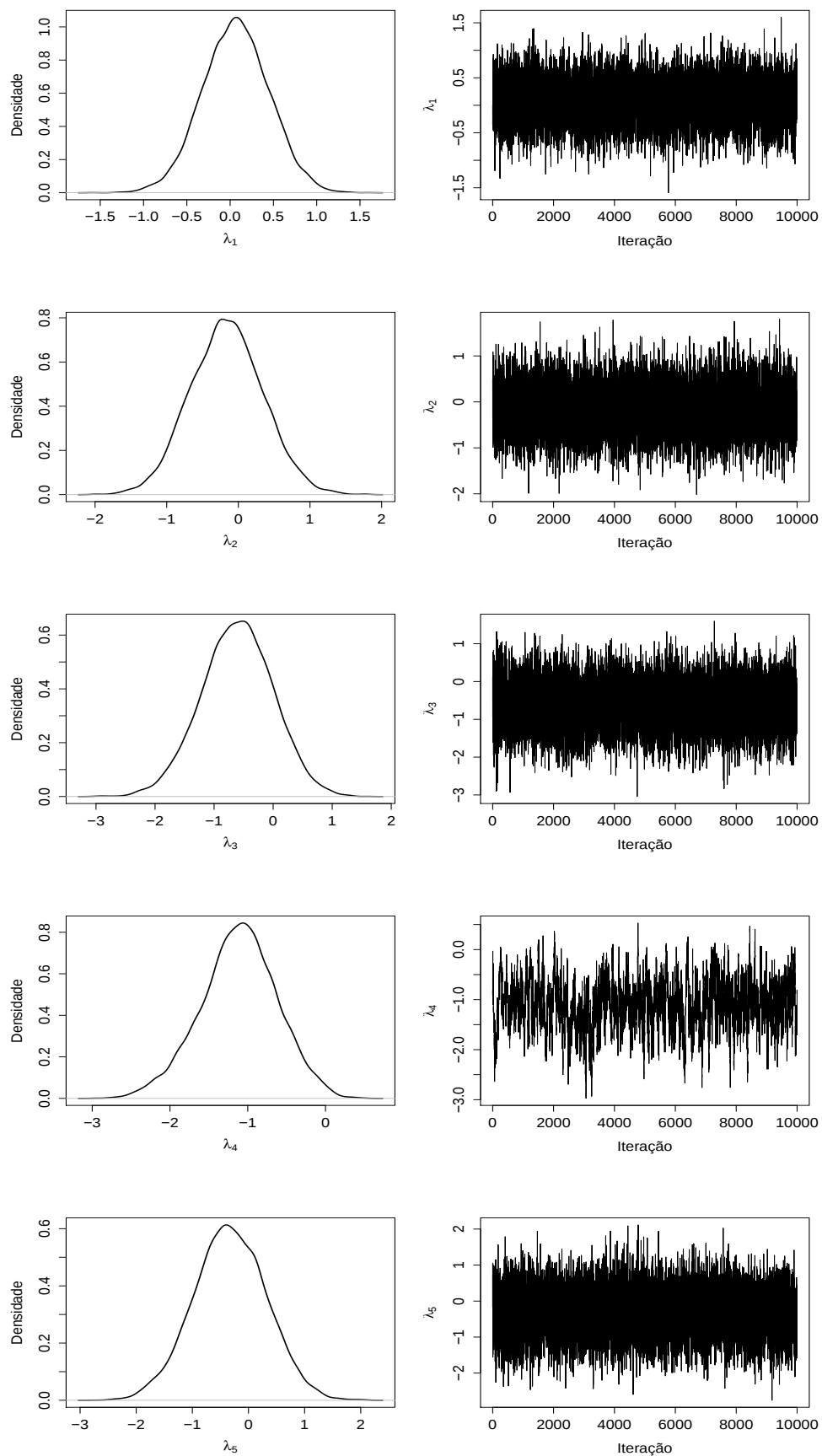


Figura 5.17: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros λ 's no modelo 6.

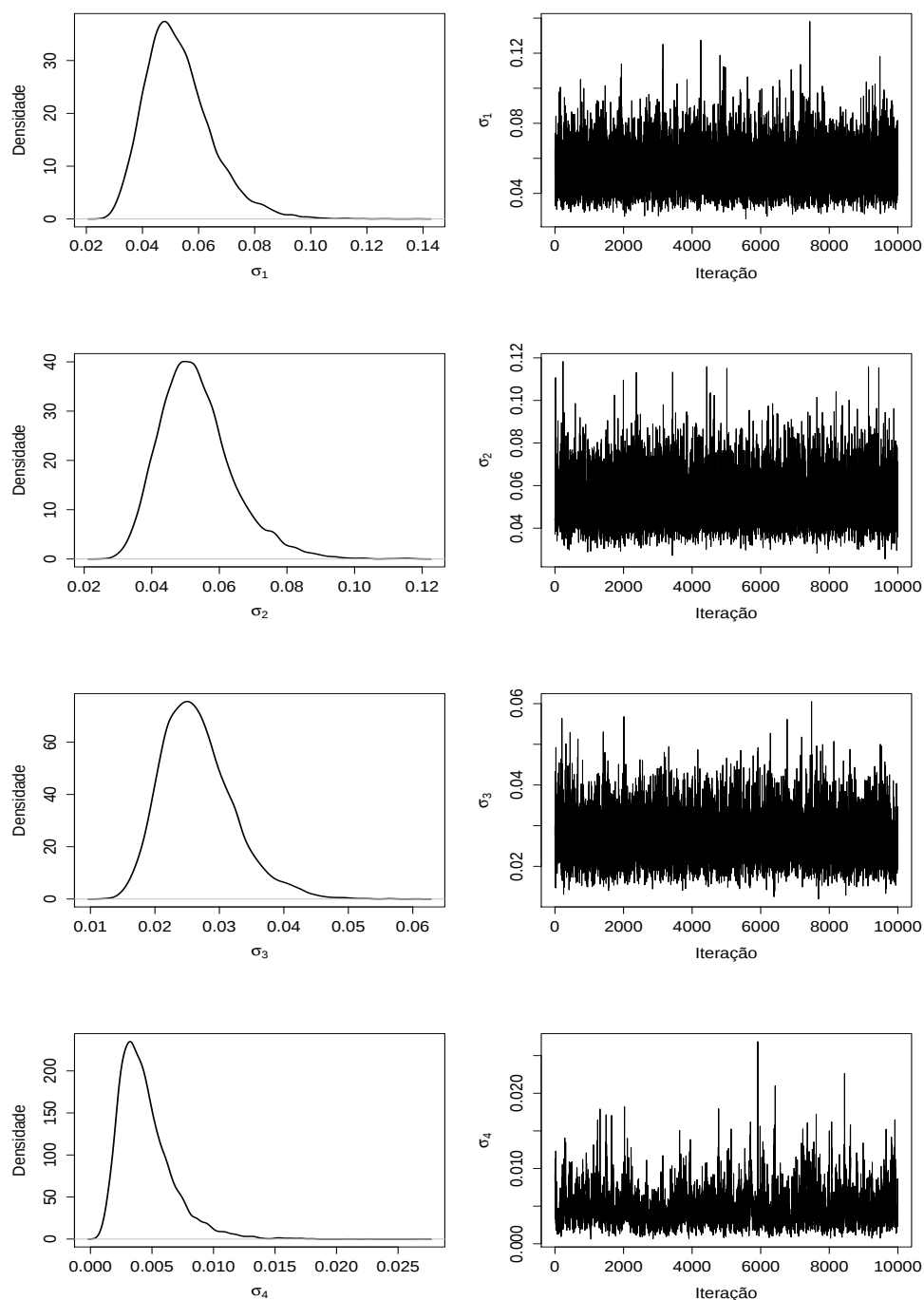


Figura 5.18: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros de variância no modelo 6.

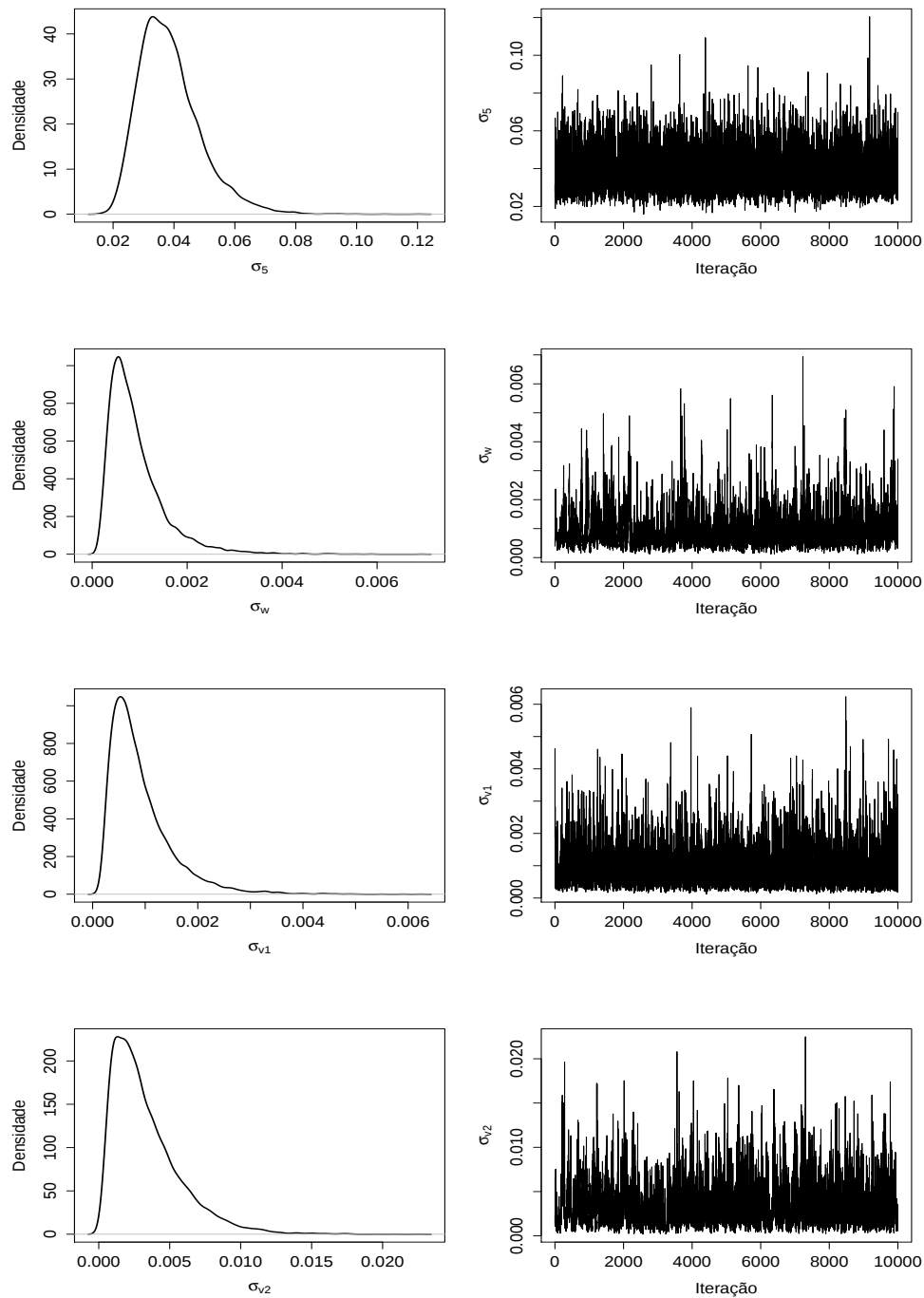


Figura 5.19: Densidades *a posteriori* via amostrador de Gibbs e traços *a posteriori* para os parâmetros de variância no modelo 6.

Dos resultados da Tabela 5.7 observamos que as diferenças entre os dois grupos (antes e depois da dieta) para os componentes proteína (pr), carboidrato (ch) e cálcio (Ca) foram significativas, pois os intervalos de credibilidade para β_{11} , β_{13} e β_{14} não incluem o valor zero.

Os parâmetros $\lambda_1, \lambda_2, \lambda_3$ e λ_5 não foram significativos, indicando que a transformação *alr*, que é um caso especial da transformação Box-Cox, para esses casos seria mais adequada ($\lambda = 0$).

A Tabela 5.8 apresenta as estimativas das proporções para os componentes de proteína, gordura do leite, carboidrato, cálcio, sódio e potássio para $z_i = 0$ (Grupo 1 - antes da dieta) e para $z_i = 1$ (Grupo 2 - depois da dieta), considerando a transformação Box-Cox com três efeitos aleatórios.

Tabela 5.8: Estimativas das proporções dos componentes - Modelo 6.

Componentes do leite	Proporções	$z = 0$ (Grupo 1)	Proporções	$z = 1$ (Grupo 2)
proteína	α_1	0,160	α_7	0,194
gordura do leite	α_2	0,183	α_8	0,141
carboidrato	α_3	0,159	α_9	0,158
cálcio	α_4	0,182	α_{10}	0,230
sódio	α_5	0,146	α_{11}	0,133
potássio	α_6	0,170	α_{12}	0,143

Podemos observar que os valores das estimativas não se alteraram em comparação aos resultados já apresentados nas Tabelas 5.2, 5.4 e 5.6, sendo que houve um aumento nas proporções para os componentes proteína e cálcio e uma diminuição das proporções de gordura do leite e potássio após a dieta balanceada.

5.5 Análise Clássica

Da mesma forma que no Capítulo 4, os dados composicionais do Apêndice B foram também analisados através de métodos clássicos aplicando a transformação *alr* em que utilizou-se a rotina *PROC NLMIXED* do *software* SAS para a estimação dos parâmetros do modelo (5.1) com o objetivo de comparação entre os métodos de estimação Bayesiano e clássico.

Para ilustrar a comparação, considerou-se o modelo (5.1), ou seja, o modelo com um efeito aleatório.

As estimativas dos parâmetros e seus respectivos intervalos de confiança são apresentados na Tabela 5.9.

Tabela 5.9: Estimativas dos parâmetros do modelo 3 - Método Clássico

Parâmetro	Estimativa	Desvio Padrão	Intervalo de Confiança (95%)
β_{01}	-0,026	0,017	(-0,062; 0,009)
β_{02}	0,034	0,018	(-0,002; 0,071)
β_{03}	-0,025	0,013	(-0,053; 0,002)
β_{04}	0,030	0,009	(0,012; 0,049)
β_{05}	-0,065	0,014	(-0,093; -0,037)
β_{11}	0,159	0,025	(0,109; 0,210)
β_{12}	-0,036	0,025	(-0,088; 0,016)
β_{13}	0,074	0,019	(0,035; 0,113)
β_{14}	0,180	0,013	(0,155; 0,206)
β_{15}	0,039	0,019	(0,0001; 0,078)
σ_1	0,095	0,009	(0,077; 0,114)
σ_2	0,098	0,009	(0,079; 0,118)
σ_3	0,073	0,007	(0,059; 0,087)
σ_4	0,049	0,005	(0,039; 0,058)
σ_5	0,074	0,007	(0,060; 0,088)
σ_w	0,012	0,008	(-0,004; 0,027)

Apesar de utilizar distribuições *a priori* não-informativas para a modelagem Bayesiana, os resultados da Tabela 5.9 não foram tão semelhantes com relação às estimativas Bayesianas. Além de apresentar diferenças antes e depois da dieta para os componentes proteína, carboidrato e cálcio, β_{15} foi significativo, ou seja, o componente sódio diferenciou após a dieta quando modelamos os dados através da inferência clássica. Percebe-se que o limite inferior do intervalo de confiança do σ_w é negativo, indicando problemas na estimação clássica para esse parâmetro.

Prosseguindo com a análise dos dados longitudinais referente a qualidade do leite após uma dieta estabelecida, ajustamos um modelo de regressão Dirichlet, sob o enfoque clássico em que as estimativas de máxima verossimilhança para os parâmetros do modelo proposto (3.4) foram obtidos no *software* R, através do pacote *DirichletReg*. Os resultados da estimação dos parâmetros do modelo estão dispostos na Tabela 5.10.

Na Tabela 5.10, podemos verificar que para os componentes gordura do leite, carboidrato, sódio e potássio, os coeficientes de regressão foram significativos, ou seja, houve diferença nesses componentes do leite após a aplicação da dieta. Podemos notar que esses coeficientes apresentaram valores negativos, indicando que a proporção para esse componentes no leite diminuiu após a dieta.

Tabela 5.10: Estimativas dos parâmetros do modelo de Regressão Dirichlet

Parâmetro	Estimativa	Desvio Padrão	Intervalo de Confiança (95%)
β_{01}	3,818	0,118	(3,587; 4,048)
β_{02}	3,957	0,117	(3,727; 4,187)
β_{03}	3,821	0,118	(3,590; 4,051)
β_{04}	3,948	0,117	(3,718; 4,178)
β_{05}	3,73	0,118	(3,499; 3,960)
β_{06}	3,878	0,117	(3,648; 4,108)
β_{11}	-0,225	0,167	(-0,552; 0,101)
β_{12}	-0,668	0,167	(-0,996; -0,341)
β_{13}	-0,419	0,167	(-0,746; -0,091)
β_{14}	-0,177	0,166	(-0,503; 0,149)
β_{15}	-0,497	0,167	(-0,825; -0,168)
β_{16}	-0,586	0,167	(-0,914; -0,259)

5.6 Discussão dos Resultados

Para a seleção de modelos, pode-se utilizar alguns critérios de discriminação de modelos, sendo que destacamos *Deviance Information Criterion* (DIC) (Spiegelhalter et al., [46]). Menores valores do DIC indicam o melhor modelo. Os valores de DIC foram obtidos através do *software* OPENBUGS.

Na Tabela 5.11, temos os valores de DIC para o modelo 3 (transformação *alr* considerando um efeito aleatório), modelo 4 (transformação Box-Cox considerando um efeito aleatório), modelo 5 (transformação *alr* considerando três efeitos aleatórios) e o modelo 6 (transformação Box-Cox considerando três efeitos aleatórios).

Tabela 5.11: Critério DIC - Modelos 3, 4, 5 e 6.

Modelo	DIC
Modelo 3	-162,5
Modelo 4	-123,1
Modelo 5	-169,6
Modelo 6	-124,7

De acordo com os resultados da Tabela 5.11, podemos observar que o melhor ajuste dos dados composicionais do Apêndice B foi para o modelo 5 (transformação *alr* considerando três efeitos aleatórios), que apresentou menor valor de DIC. E ainda, observa-se que os modelos que consideraram a transformação *alr* apresentaram menores valores do critério DIC que os modelos que abordaram a transformação Box-Cox.

Além disso, calculou-se a soma do quadrado das diferenças entre os valores observados e os valores ajustados dos modelos são apresentados na Tabela 5.12.

Tabela 5.12: Soma do quadrado das diferenças entre os valores observados e os valores ajustados.

Modelo	$\sum_{i=1}^n \sum_{j=1}^6 (y_{ij} - \hat{y}_{ij})^2$
Modelo 3 (transformação <i>alr</i> com 1 efeito aleatório - Bayesiano)	0,220812
Modelo 4 (transformação Box-Cox com 1 efeito aleatório - Bayesiano)	0,220743
Modelo 5 (transformação <i>alr</i> com 3 efeitos aleatórios - Bayesiano)	0,220705
Modelo 6 (transformação Box-Cox com 3 efeitos aleatórios - Bayesiano)	0,220877
Modelo (transformação <i>alr</i> com 1 efeito aleatório - Análise Clássica)	0,300242
Modelo de Regressão Dirichlet clássico	0,220714

Analisando os resultados obtidos, percebe-se que há pouca diferença entre os valores da Tabela 5.12.

Por outro lado, podemos afirmar que dentre os modelos ajustados, o modelo 5 (*alr* com 3 efeitos aleatórios) se ajusta melhor aos dados longitudinais (qualidade do leite) através do critério DIC e também pela soma do quadrado das diferenças entre os valores observados e os valores ajustados em que apresentou menor valor (diferença mínima em relação aos outros modelos). É importante salientar que os menores valores da Tabela 5.12 foram para os modelos em que os parâmetros foram estimados através da inferência Bayesiana.

A Figura 5.20 apresenta os valores ajustados e observados dos componentes do leite (proteína, gordura do leite, carboidrato, cálcio, sódio e potássio) para cada modelo proposto. Podemos observar que os valores ajustados através da inferência clássica são os mais distantes em relação à média observada.

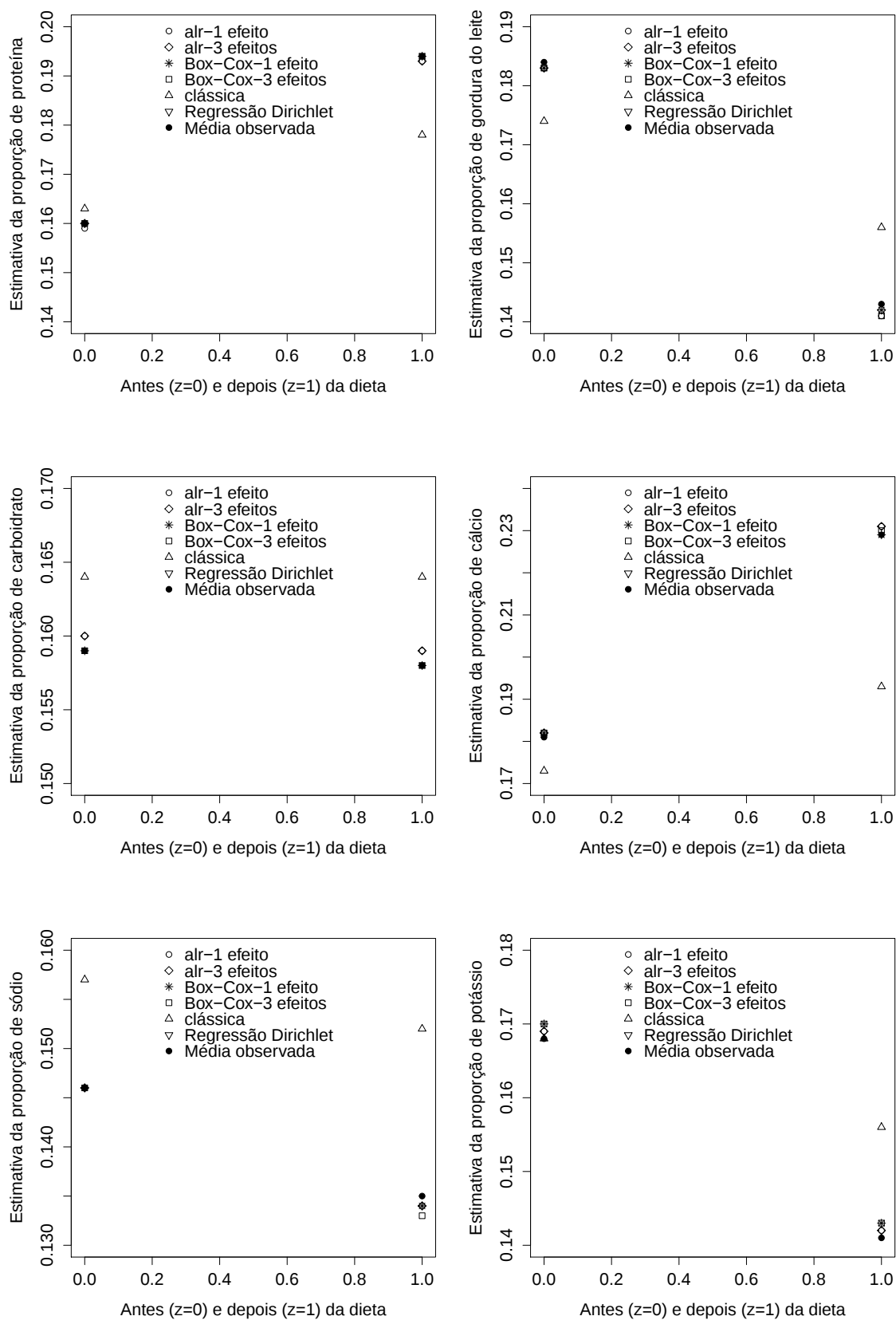


Figura 5.20: Gráficos dos valores observados e ajustados dos componentes do leite de acordo com o modelo

6 Considerações Finais

Neste trabalho, realizamos um estudo de dados composicionais aplicados em modelos de regressão sob o enfoque Bayesiano. Uma das restrições que os dados composicionais possui e que difere dos dados multivariados usuais, é que eles são definidos como um vetor em que seus elementos são positivos e correspondem à proporções ou frações, tal que a soma de todos eles é igual a 1. No entanto, essa restrição não deve ser ignorada para que os resultados da análise estatística não se tornem imprecisos.

Foram abordados transformações em dados composicionais, sendo elas, a transformação *alr* e Box-Cox, em que foram considerados erros não correlacionados com distribuição Normal para a modelagem estatística. A metodologia proposta foi aplicada em dois conjuntos de dados, sendo que um deles refere-se à dados longitudinais. Vale ressaltar que a abordagem da metodologia de dados composicionais para a análise dos dois conjuntos de dados do presente trabalho é inédita. Além disso, para efeito de comparação, abordamos o método clássico e o modelo de regressão Dirichlet (clássico).

Para o conjunto de dados referente aos jogos da Superliga de vôlei masculina, o modelo que obteve melhor ajuste foi o modelo de regressão Dirichlet, apesar de que o modelo 1 (transformação *alr*) apresentou melhor ajuste dentre os modelos que utilizou métodos Bayesianos. Para o conjunto de dados longitudinais (Apêndice B), o modelo 5 que considerou a transformação *alr* com três efeitos aleatórios foi mais adequado aos dados, seguido do modelo de regressão Dirichlet.

Há muitas dificuldades na obtenção de resultados na inferência clássica para tais modelos, principalmente quando há presença de um vetor de covariáveis. Com isso, a aplicação dos métodos Bayesianos para dados composicionais mostrou ser uma alternativa vantajosa para a sua análise, especialmente relacionados às restrições que a variável resposta possui, considerando ainda a aplicação dos métodos MCMC.

O uso do *software* OPENBUGS proporcionou facilidade na implementação dos algoritmos amostrador de Gibbs e Metropolis-Hastings para as aplicações apresentadas,

juntamente com o *software* R através do pacote *CODA* para a construção dos gráficos e dos resultados obtidos do diagnóstico de convergência de Geweke e o *software* SAS para a obtenção das estimativas através da inferência clássica. Os *software* utilizados mostraram ser ferramentas computacionais importantes para a obtenção dos resultados de interesse.

Como perspectiva futura de trabalho, uma delas envolve o estudo de modelos apropriados quando há presença de zeros em dados composicionais, assumir outros tipos de estrutura para os erros dos modelos, como por exemplo, distribuições assimétricas, *t*-Student multivariada.

Referências

- [1] ACHCAR, J. A.; OBAGE, S. C. Uma abordagem Bayesiana para dados composicionais considerando erros correlacionados. *Revista de Matemática e Estatística*, v. 23, n. 2, p. 95-107, 2005.
- [2] AITCHISON, J. The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 44, n. 2, p. 139-177, 1982.
- [3] AITCHISON, J. *The statistical analysis of compositional data*. Chapman & Hall, 1986.
- [4] AITCHISON, J. *A Concise Guide to Compositional Data Analysis*. CDA Workshop, Girona, 2005.
- [5] AITCHISON, J.; EGOZCUE, J. J. Compositional data analysis: Where are we and where should we be heading? *Mathematical Geology*, v. 37, n. 7, p. 829-850, 2005.
- [6] AITCHISON, J.; GREENACRE, M. Biplots of Compositional Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*. v. 51, n. 4, p. 375-392, 2002.
- [7] AITCHISON, J.; SHEN, S. M. Logistic-normal distributions: Some properties and uses. *Biometrika*. v. 67, n. 2, p. 261-272, 1980.
- [8] BOX, G. E. P.; COX, D. R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, v. 26, n. 2, p. 211-252, 1964.
- [9] BOX, G. E. P.; TIAO, G. C. *Bayesian inference in statistical analysis*. Reading: Addison-Wesley, 1973.
- [10] BREHM, J.; GATES, S.; GOMEZ, B. *A Monte Carlo Comparison of Methods for Compositional Data Analysis*. Political Methodology Society Annual Meetings, San Diego, 1998.
- [11] CAMARGO, A. P.; STERN, J. M.; LAURETTO, M. S. Estimation and Model Selection in Dirichlet Regression. In: *AIP Conference Proceedings*, v. 1443, p. 206-213, 2012.
- [12] CAMPBELL, G.; MOSIMANN, J. Multivariate methods for proportional shape. *ASA Proceedings of the Section on Statistical Graphics*, 10-17, 1987.
- [13] CASELLA, G.; BERGER, R. L. *Statistical Inference*. 2^a ed., Pacific Grove: Duxbury Advanced Series, 2002.
- [14] CASELLA, G.; GEORGE, E. I. Explaining the Gibbs Sampler. *The American Statistician*, v. 46, n. 3, p. 167-174, 1992.

- [15] CBV - Confederação Brasileira de Voleibol. **Dados Superliga de Vôlei Masculina**. Disponível em: <http://www.cbv.com.br/v1/superliga-1112/m-tabela.asp>. Acesso em: 23 de agosto de 2012.
- [16] CONGDON, P. *Bayesian Statistical Modelling*. Second edition, John Wiley & Sons, 2006.
- [17] EHLERS, R. S. *Introdução à Inferência Bayesiana*. 5ª ed, 2007.
- [18] GAMERMAN, D.; LOPES, H. F. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall/CRC, 2006.
- [19] GAMERMAN, D.; MIGON, H. S. *Inferência Estatística: Uma abordagem integrada*. Universidade Federal do Rio de Janeiro, 1997.
- [20] GELFAND, A. E.; SMITH, A. F. M. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association*, v. 85, n. 410, p. 398-409, 1990.
- [21] GELMAN, A.; RUBIN, D. B. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, v. 7, n. 4, p. 457-472, 1992.
- [22] GEMAN, S.; GEMAN, D. Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, n. 6, p. 721-741, 1984.
- [23] GEWEKE, J. *Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments*. In Bayesian Statistics 4, Oxford University Press, 1992.
- [24] GUEORGUIEVA, R.; ROSENHECK, R.; ZELTERMAN, D. Dirichlet component regression and its applications to psychiatric data. *Computational Statistics and Data Analysis*, v. 52, n. 12, p. 5344-5355, 2008.
- [25] HASTINGS, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, v. 57, n. 1, p. 97-109, 1970.
- [26] HIJAZI, R. An EM-Algorithm Based Method to Deal with Rounded Zeros in Compositional Data under Dirichlet Models. *Proceedings of the 4th International Workshop on Compositional Data Analysis*, 2011.
- [27] HIJAZI, R.; JERNIGAN, R. W. Modelling compositional data using Dirichlet regression models. *Journal of Applied Probability & Statistics*, v. 4, n. 1, p. 77-91, 2009.
- [28] IYENGAR, M.; DEY, D. K. *Bayesian analysis of compositional data*. Department of Statistics, University of Connecticut, Storrs, CT 06269-3120, 1996.
- [29] IYENGAR, M.; DEY, D. K. *Box-Cox transformations in Bayesian analysis of compositional data*. *Environmetrics*, v. 9, n. 6, p. 657-671, 1998.
- [30] JOHNSON, D. S.; HOETING, J. A.; POFF, N. L. Biological Monitoring: A Bayesian Model for multivariate compositional data. *Bayesian statistics and its applications*. Anamaya Publishers, New Delhi, India, p. 270-289, 2006.

- [31] JOHNSON, R.; WICHERN, D. *Applied multivariate statistical analysis*. New Jersey: Prentice Hall, 1998.
- [32] LUNN, D.; SPIEGELHALTER, D.; THOMAS, A.; BEST, N. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, v. 28, n. 25, p. 3049–3067, 2009.
- [33] MARTÍN-FERNÁNDEZ, J. A.; BARCELÓ-VIDAL, C.; PAWLOWSKY-GLAHN, V. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology*, v. 35, n. 3, p. 253–278, 2003.
- [34] MARTINS, A. B. T.; RIBEIRO Jr, P. J.; BONAT, W. H. Um modelo geoestatístico bivariado para dados composicionais. *Revista de Matemática e Estatística*, v. 27, n. 3, p. 456–477, 2009.
- [35] McALISTER, D. The law of the geometric mean. *Proceedings of the Royal Society of London*, v. 29, p. 367–376, 1879.
- [36] MELO, T. F. N.; VASCONCELLOS, K. L. P.; LEMONTE, A. J. Some restriction tests in a new class of regression models for proportions. *Computational Statistics & Data Analysis*, v. 53, n. 12, p. 3972–3979, 2009.
- [37] METROPOLIS, N.; ROSENBULTH, A. W.; ROSENBULTH, M. N.; TELLER, A. H.; TELLER, E. Equation of State Calculations by Fast Computing Machine. *Journal of Chemical Physics*, v. 21, n. 6, p. 1087–1092, 1953.
- [38] MIGON, H. S.; SOUZA, A. D. P.; SCHMIDT, A. M. *Modelos Hierárquicos e Aplicações*. 18º SINAPE, 279 p., 2008.
- [39] NEOCLEOUS, T.; AITKEN, C.; ZADORA, G. Transformations for compositional data with zeros with an application to forensic evidence evaluation. *Chemometrics and Intelligent Laboratory Systems*, v. 1, n. 1, p. 77–85, 2011.
- [40] NTZOUFRAS, I. *Bayesian Modeling using Winbugs*. New Jersey: John Wiley & Sons, 2009.
- [41] PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa, 2003.
- [42] PAWLOWSKY-GLAHN, V.; EGOZCUE, J. J.; TOLOSANA-DELGADO, R. *Lecture Notes on Compositional Data Analysis*. Universitat de Girona, Espanha, 2007.
- [43] PRESS, S. J. *Bayesian Statistics: Principles, Models, and Applications*. John Wiley & Sons, 1989.
- [44] RAYENS, W. S.; SRINIVASAN, C. Box-Cox transformations in the analysis of compositional data. *Journal of Chemometrics*, v. 5, n. 3, p. 227–239, 1991.
- [45] RAYENS, W. S.; SRINIVASAN, C. Estimation in compositional data analysis. *Journal of Chemometrics*, v. 5, n. 4, p. 361–374, 1991.
- [46] SPIEGELHALTER, D.; BEST, N. G.; CARLIN, B. P.; VAN DER LINDE, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 64, n. 4, p. 583–639, 2002.

-
- [47] SPIEGELHALTER, D.; THOMAS, A.; BEST, N.; LUNN, D. *Winbugs User Manual: Version 1.4*. <http://www.mrc-bsu.cam.ac.uk/bugs>. Relatório técnico, Cambridge: Medical Research Council Biostatistics, 2003.

APÊNDICE A – Conjunto de Dados da Superliga de Vôlei Masculina

Tabela A.1: Conjunto de dados referente aos jogos da Superliga de vôlei masculina 2011/2012.

Jogos	% ataque	% bloqueio	% saque	% erro	z	Jogos	% ataque	% bloqueio	% saque	% erro	z
1	48,00	12,00	2,67	37,33	1	65	53,27	9,35	5,61	31,78	1
2	53,06	14,29	7,14	25,51	1	66	51,09	6,52	4,35	38,04	1
3	44,00	13,33	8,00	34,67	0	67	55,26	6,58	6,58	31,58	1
4	52,63	14,74	7,37	25,26	0	68	55,10	12,24	3,06	29,59	0
5	56,00	8,00	5,33	30,67	1	69	64,94	6,49	5,19	23,38	1
6	65,63	10,16	2,34	21,88	0	70	57,33	10,67	10,67	21,33	1
7	54,67	14,67	2,67	28,00	1	71	56,70	12,37	7,22	23,71	0
8	50,00	12,50	9,62	27,88	1	72	44,74	9,21	7,89	38,16	0
9	52,58	15,46	3,09	28,87	0	73	62,50	13,54	4,17	19,79	1
10	57,33	13,33	4,00	25,33	1	74	56,52	5,22	6,09	32,17	1
11	56,60	11,32	6,60	25,47	0	75	53,19	14,89	4,26	27,66	0
12	49,33	9,33	14,67	26,67	1	76	57,33	9,33	6,67	26,67	1
13	56,67	8,33	6,67	28,33	0	77	50,67	18,67	2,67	28,00	1
14	66,67	4,00	1,33	28,00	1	78	53,95	10,53	6,58	28,95	0
15	56,70	10,31	3,09	29,90	0	79	44,00	16,00	2,67	37,33	1
16	62,67	6,67	4,00	26,67	1	80	48,00	5,33	6,67	40,00	1
17	49,35	5,19	3,90	41,56	1	81	56,00	8,00	2,67	33,33	0
18	48,00	13,33	6,67	32,00	1	82	57,84	11,76	2,94	27,45	0
19	56,12	14,29	4,08	25,51	0	83	47,25	12,09	4,40	36,26	1
20	44,00	9,33	5,33	41,33	0	84	59,81	11,21	3,74	25,23	0
21	54,95	10,81	5,41	28,83	1	85	53,06	11,22	5,10	30,61	1
22	59,34	15,38	3,30	21,98	0	86	54,55	7,27	4,55	33,64	0
23	51,52	12,12	1,01	35,35	0	87	59,18	6,12	3,06	31,63	1
24	57,69	14,10	5,13	23,08	0	88	46,88	13,54	3,13	36,46	1
25	53,33	14,67	2,67	29,33	0	89	57,50	11,67	1,67	29,17	0
26	55,45	13,86	5,94	24,75	1	90	48,00	20,00	8,00	24,00	0
27	48,04	10,78	3,92	37,25	0	91	54,00	13,00	2,00	31,00	1
28	48,00	17,33	9,33	25,33	0	92	54,67	6,67	8,00	30,67	1
29	57,33	8,00	6,67	28,00	1	93	56,12	7,14	3,06	33,67	1
30	53,78	8,40	5,04	32,77	1	94	56,00	8,00	8,00	28,00	1
31	44,74	10,53	7,02	37,72	1	95	50,67	13,33	6,67	29,33	0
32	53,57	13,10	3,57	29,76	1	96	52,00	12,00	1,33	34,67	1
33	59,41	5,94	6,93	27,72	1	97	58,06	10,75	2,15	29,03	0
34	59,21	9,21	6,58	25,00	0	98	45,33	6,67	2,67	45,33	0
35	51,04	13,54	6,25	29,17	0	99	51,90	15,19	2,53	30,38	0
36	50,51	13,13	3,03	33,33	0	100	52,78	15,74	1,85	29,63	1
37	55,34	16,50	4,85	23,30	0	101	54,29	9,52	7,62	28,57	0
38	61,33	4,00	2,67	32,00	0	102	50,67	14,67	5,33	29,33	0
39	48,98	17,35	2,04	31,63	1	103	48,00	20,00	2,67	29,33	1
40	55,32	9,57	5,32	29,79	1	104	62,03	6,33	2,53	29,11	1
41	56,25	7,29	5,21	31,25	1	105	40,00	12,00	9,33	38,67	0
42	56,19	7,62	2,86	33,33	0	106	59,09	8,18	4,55	28,18	1
43	51,72	8,62	6,03	33,62	1	107	49,33	9,33	8,00	33,33	1
44	49,46	13,98	3,23	33,33	1	108	58,25	6,80	3,88	31,07	0
45	47,27	11,82	5,45	35,45	0	109	60,78	9,80	6,86	22,55	0
46	57,33	13,33	1,33	28,00	0	110	55,77	13,46	1,92	28,85	0
47	56,84	14,74	3,16	25,26	0	111	56,00	8,00	4,00	32,00	0
48	60,61	10,10	3,03	26,26	0	112	55,88	10,78	5,88	27,45	0
49	60,18	10,62	3,54	25,66	0	113	64,13	5,43	7,61	22,83	0
50	52,43	9,71	3,88	33,98	0	114	46,67	8,00	9,33	36,00	1
51	50,67	13,33	4,00	32,00	0	115	56,14	8,77	4,39	30,70	1
52	63,30	8,26	4,59	23,85	1	116	49,00	9,00	6,00	36,00	1
53	54,46	4,95	2,97	37,62	1	117	54,67	10,67	5,33	29,33	1
54	56,25	8,33	4,17	31,25	0	118	48,65	18,02	4,50	28,83	1
55	69,89	5,38	5,38	19,35	0	119	64,22	7,34	1,83	26,61	1
56	65,82	15,19	3,80	15,19	0	120	56,58	7,89	11,84	23,68	1
57	57,89	5,26	11,84	25,00	1	121	56,38	14,89	5,32	23,40	0
58	36,84	12,63	7,37	43,16	0	122	48,45	15,46	5,15	30,93	0
59	50,00	14,42	1,92	33,65	0	123	52,83	6,60	6,60	33,96	0
60	52,08	6,25	5,21	36,46	0	124	60,00	5,33	9,33	25,33	0
61	53,33	13,33	5,33	28,00	0	125	59,63	9,17	1,83	29,36	1
62	44,00	16,00	10,67	29,33	1	126	54,67	10,67	1,33	33,33	1
63	58,67	10,67	8,00	22,67	1	127	54,67	6,67	6,67	32,00	0
64	54,00	8,00	5,00	33,00	0	128	46,91	6,17	4,94	41,98	1

APÊNDICE B – Conjunto de Dados Longitudinais

Tabela B.1: Composições alimentares do leite de 30 vacas (pr=proteína, mf=gordura do leite, ch=carboidrato, Ca=cálcio, Na=sódio, K=potássio) antes e depois de uma nova dieta.

Animal	Antes						Depois					
	pr	mf	ch	Ca	Na	K	pr	mf	ch	Ca	Na	K
1	0,1389	0,2278	0,1553	0,1699	0,1400	0,1680	0,1753	0,1459	0,1552	0,2122	0,1642	0,1473
2	0,1377	0,1661	0,1540	0,2066	0,1549	0,1807	0,2090	0,0937	0,1313	0,2341	0,1717	0,1603
3	0,1464	0,1525	0,1672	0,1976	0,1572	0,1792	0,2387	0,1207	0,1497	0,1832	0,1652	0,1426
4	0,1950	0,1564	0,1562	0,1993	0,1163	0,1768	0,2398	0,1345	0,1726	0,2310	0,0896	0,1326
5	0,1988	0,1423	0,1507	0,1869	0,1470	0,1742	0,1173	0,1647	0,1535	0,2482	0,1577	0,1586
6	0,1498	0,1979	0,1858	0,1782	0,1237	0,1645	0,1701	0,1063	0,1524	0,2508	0,1826	0,1379
7	0,1467	0,1552	0,1828	0,1778	0,1564	0,1812	0,2018	0,1109	0,1166	0,2800	0,1471	0,1436
8	0,1109	0,2690	0,1466	0,2046	0,0996	0,1693	0,2142	0,0944	0,1472	0,2488	0,1536	0,1418
9	0,1198	0,2005	0,1351	0,1984	0,1654	0,1807	0,1890	0,1622	0,2066	0,2182	0,0965	0,1274
10	0,2164	0,1624	0,1687	0,1818	0,1111	0,1597	0,2097	0,1431	0,1706	0,2082	0,1435	0,1249
11	0,1792	0,1585	0,1358	0,1907	0,1645	0,1713	0,1562	0,1611	0,1901	0,2452	0,1126	0,1349
12	0,1650	0,1836	0,1401	0,1896	0,1477	0,1740	0,1292	0,2046	0,1977	0,2104	0,1313	0,1268
13	0,1744	0,1999	0,1742	0,1608	0,1310	0,1597	0,2538	0,1314	0,1499	0,1665	0,1655	0,1328
14	0,1319	0,1689	0,1338	0,2055	0,1713	0,1886	0,1959	0,1289	0,1612	0,2324	0,1370	0,1448
15	0,1482	0,2426	0,1544	0,1524	0,1476	0,1549	0,2154	0,1707	0,1713	0,2456	0,0875	0,1095
16	0,1857	0,1891	0,1810	0,1829	0,1042	0,1571	0,1748	0,1715	0,1458	0,1959	0,1795	0,1326
17	0,1497	0,1552	0,1419	0,2027	0,1623	0,1883	0,1446	0,1634	0,1757	0,2265	0,1375	0,1523
18	0,1518	0,1703	0,1412	0,1656	0,2064	0,1646	0,1690	0,1918	0,1625	0,2510	0,0992	0,1264
19	0,1582	0,1437	0,1682	0,1852	0,1733	0,1713	0,1791	0,1607	0,1792	0,1990	0,1221	0,1599
20	0,1683	0,1832	0,1618	0,1659	0,1643	0,1565	0,2149	0,1210	0,1446	0,2589	0,1090	0,1516
21	0,1394	0,2128	0,1999	0,1619	0,1336	0,1523	0,1799	0,1545	0,1605	0,2272	0,1408	0,1371
22	0,1687	0,1570	0,1399	0,1883	0,1632	0,1829	0,1723	0,1566	0,1638	0,2460	0,1299	0,1314
23	0,1988	0,1436	0,1529	0,1809	0,1537	0,1700	0,1778	0,1285	0,1905	0,2468	0,1161	0,1403
24	0,1870	0,1770	0,1561	0,1754	0,1424	0,1622	0,2045	0,1670	0,1612	0,2124	0,1248	0,1301
25	0,1243	0,2008	0,1520	0,2043	0,1346	0,1840	0,2063	0,1206	0,1428	0,2287	0,1461	0,1555
26	0,1686	0,2286	0,1465	0,1641	0,1373	0,1549	0,2709	0,1018	0,1207	0,2491	0,1226	0,1349
27	0,1512	0,1692	0,1658	0,1865	0,1508	0,1766	0,2099	0,1188	0,1450	0,2617	0,1047	0,1598
28	0,2033	0,2042	0,1676	0,1460	0,1344	0,1445	0,2046	0,1370	0,1325	0,2779	0,1111	0,1369
29	0,1455	0,1817	0,1783	0,1519	0,1798	0,1628	0,2808	0,1252	0,1390	0,1813	0,1328	0,1408
30	0,1451	0,2350	0,1886	0,1696	0,1195	0,1422	0,1245	0,1871	0,1554	0,2084	0,1573	0,1672

APÊNDICE C – Programas

Este Apêndice apresenta alguns dos programas computacionais utilizados para a estimação dos parâmetros através do MCMC no OPENBUGS e a análise clássica realizada nos *software* SAS na *Procedure NLMIXED* e *R* para a estimação dos parâmetros do modelo de regressão Dirichlet.

C.1 OPENBUGS

C.1.1 Programa - Modelo 1

```
model
{
  for(i in 1:n) {
    y1[i] <- log(x1[i]/x4[i])
    y2[i] <- log(x2[i]/x4[i])
    y3[i] <- log(x3[i]/x4[i])
    y1[i] ~ dnorm(mu1[i], tau[1])
    y2[i] ~ dnorm(mu2[i], tau[2])
    y3[i] ~ dnorm(mu3[i], tau[3])
    mu1[i] <- beta0[1]+beta1[1]*z1[i]
    mu2[i] <- beta0[2]+beta1[2]*z1[i]
    mu3[i] <- beta0[3]+beta1[3]*z1[i]
  }
  beta0[1] ~ dnorm(0,0.001)
  beta0[2] ~ dnorm(0,0.001)
  beta0[3] ~ dnorm(0,0.001)
  beta1[1] ~ dnorm(0,0.001)
  beta1[2] ~ dnorm(0,0.001)
  beta1[3] ~ dnorm(0,0.001)
```

```

tau[1] ~ dgamma(1.0E-2,1.0E-2)
tau[2] ~ dgamma(1.0E-2,1.0E-2)
tau[3] ~ dgamma(1.0E-2,1.0E-2)
sigma[1] <- 1/tau[1]
sigma[2] <- 1/tau[2]
sigma[3] <- 1/tau[3]
}

```

C.1.2 Programa - Modelo 2

```

model
{
  for(i in 1:n) {
    L1[i] <- (sqrt(tau[1]/6.283185))*exp(-0.5*tau[1]*pow(dif1[i],2))*pow(y1[i],(lambda[1]-1))
    L2[i] <- (sqrt(tau[2]/6.283185))*exp(-0.5*tau[2]*pow(dif2[i],2))*pow(y2[i],(lambda[2]-1))
    L3[i] <- (sqrt(tau[3]/6.283185))*exp(-0.5*tau[3]*pow(dif3[i],2))*pow(y3[i],(lambda[3]-1))
    L[i] <- L1[i]*L2[i]*L3[i]
  }

  for(i in 1:n) {
    y1[i] <- (x1[i]/x4[i])
    y2[i] <- (x2[i]/x4[i])
    y3[i] <- (x3[i]/x4[i])

    transf1[i] <- (pow(y1[i],lambda[1])-1)/lambda[1]
    transf2[i] <- (pow(y2[i],lambda[2])-1)/lambda[2]
    transf3[i] <- (pow(y3[i],lambda[3])-1)/lambda[3]

    mu1[i] <- beta0[1]+beta1[1]*z1[i]
    mu2[i] <- beta0[2]+beta1[2]*z1[i]
    mu3[i] <- beta0[3]+beta1[3]*z1[i]

    dif1[i] <- transf1[i]-mu1[i]
    dif2[i] <- transf2[i]-mu2[i]
    dif3[i] <- transf3[i]-mu3[i]
  }

  beta0[1] ~ dnorm(0,1.0E-3)
  beta0[2] ~ dnorm(0,1.0E-3)
  beta0[3] ~ dnorm(0,1.0E-3)
  beta1[1] ~ dnorm(0,1.0E-3)

```

```

beta1[2] ~ dnorm(0,1.0E-3)
beta1[3] ~ dnorm(0,1.0E-3)
tau[1] ~ dgamma(1.0E-2,1.0E-2)
tau[2] ~ dgamma(1.0E-2,1.0E-2)
tau[3] ~ dgamma(1.0E-2,1.0E-2)
sigma[1] <- 1/tau[1]
sigma[2] <- 1/tau[2]
sigma[3] <- 1/tau[3]
lambda[1] ~ dnorm(0,1.0E-2)
lambda[2] ~ dnorm(0,1.0E-2)
lambda[3] ~ dnorm(0,1.0E-2)
for(i in 1:n) {
uns[i] ~ dbern(L[i])
}
}

```

C.1.3 Programa - Modelo 3

```

model
{
for(i in 1:n) {
y1[i] <- log(x1[i]/x6[i])
y2[i] <- log(x2[i]/x6[i])
y3[i] <- log(x3[i]/x6[i])
y4[i] <- log(x4[i]/x6[i])
y5[i] <- log(x5[i]/x6[i])
L1[i] <- (sqrt(tau[1]/6.283185))*exp(-0.5*tau[1]*pow(dif1[i],2))
L2[i] <- (sqrt(tau[2]/6.283185))*exp(-0.5*tau[2]*pow(dif2[i],2))
L3[i] <- (sqrt(tau[3]/6.283185))*exp(-0.5*tau[3]*pow(dif3[i],2))
L4[i] <- (sqrt(tau[4]/6.283185))*exp(-0.5*tau[4]*pow(dif4[i],2))
L5[i] <- (sqrt(tau[5]/6.283185))*exp(-0.5*tau[5]*pow(dif5[i],2))
L[i] <- L1[i]*L2[i]*L3[i]*L4[i]*L5[i]
}
for(i in 1:n)
{
dif1[i] <- y1[i]-mu1[i]

```

```
dif2[i] <- y2[i]-mu2[i]
dif3[i] <- y3[i]-mu3[i]
dif4[i] <- y4[i]-mu4[i]
dif5[i] <- y5[i]-mu5[i]
mu1[i] <- beta0[1]+beta1[1]*z[i]+w[i]
mu2[i] <- beta0[2]+beta1[2]*z[i]+w[i]
mu3[i] <- beta0[3]+beta1[3]*z[i]+w[i]
mu4[i] <- beta0[4]+beta1[4]*z[i]+w[i]
mu5[i] <- beta0[5]+beta1[5]*z[i]+w[i]
w[i] ~ dnorm(0, tau.w)
}
beta0[1] ~ dnorm(1,1.0E-6)
beta0[2] ~ dnorm(1,1.0E-6)
beta0[3] ~ dnorm(1,1.0E-6)
beta0[4] ~ dnorm(1,1.0E-6)
beta0[5] ~ dnorm(1,1.0E-6)
beta1[1] ~ dnorm(0,1.0E-6)
beta1[2] ~ dnorm(0,1.0E-6)
beta1[3] ~ dnorm(0,1.0E-6)
beta1[4] ~ dnorm(0,1.0E-6)
beta1[5] ~ dnorm(0,1.0E-6)
sigma[1] <- 1/tau[1]
sigma[2] <- 1/tau[2]
sigma[3] <- 1/tau[3]
sigma[4] <- 1/tau[4]
sigma[5] <- 1/tau[5]
sigma.w <- 1/tau.w
tau[1] ~ dgamma(0.001,0.001)
tau[2] ~ dgamma(0.001,0.001)
tau[3] ~ dgamma(0.001,0.001)
tau[4] ~ dgamma(0.001,0.001)
tau[5] ~ dgamma(0.001,0.001)
tau.w ~ dgamma(0.001,0.001)
for(i in 1:n) {
uns[i] ~ dbern(L[i])
```

```
}
}
```

C.1.4 Programa - Modelo 4

```
model
{
  for(i in 1:n) {
    y1[i] <- (x1[i]/x6[i])
    y2[i] <- (x2[i]/x6[i])
    y3[i] <- (x3[i]/x6[i])
    y4[i] <- (x4[i]/x6[i])
    y5[i] <- (x5[i]/x6[i])
    transf1[i] <- (pow(y1[i],lambda[1])-1)/lambda[1]
    transf2[i] <- (pow(y2[i],lambda[2])-1)/lambda[2]
    transf3[i] <- (pow(y3[i],lambda[3])-1)/lambda[3]
    transf4[i] <- (pow(y4[i],lambda[4])-1)/lambda[4]
    transf5[i] <- (pow(y5[i],lambda[5])-1)/lambda[5]
    L1[i] <- (sqrt(tau[1]/6.283185))*exp(-0.5*tau[1]*pow(dif1[i],2))*pow(y1[i],(lambda[1]-1))
    L2[i] <- (sqrt(tau[2]/6.283185))*exp(-0.5*tau[2]*pow(dif2[i],2))*pow(y2[i],(lambda[2]-1))
    L3[i] <- (sqrt(tau[3]/6.283185))*exp(-0.5*tau[3]*pow(dif3[i],2))*pow(y3[i],(lambda[3]-1))
    L4[i] <- (sqrt(tau[4]/6.283185))*exp(-0.5*tau[4]*pow(dif4[i],2))*pow(y4[i],(lambda[4]-1))
    L5[i] <- (sqrt(tau[5]/6.283185))*exp(-0.5*tau[5]*pow(dif5[i],2))*pow(y5[i],(lambda[5]-1))
    L[i] <- L1[i]*L2[i]*L3[i]*L4[i]*L5[i]
  }
  for(i in 1:n)
  {
    dif1[i] <- transf1[i]-mu1[i]
    dif2[i] <- transf2[i]-mu2[i]
    dif3[i] <- transf3[i]-mu3[i]
    dif4[i] <- transf4[i]-mu4[i]
    dif5[i] <- transf5[i]-mu5[i]
    mu1[i] <- beta0[1]+beta1[1]*z[i]+w[i]
    mu2[i] <- beta0[2]+beta1[2]*z[i]+w[i]
    mu3[i] <- beta0[3]+beta1[3]*z[i]+w[i]
    mu4[i] <- beta0[4]+beta1[4]*z[i]+w[i]
```

```
mu5[i] <- beta0[5]+beta1[5]*z[i]+w[i]
w[i] ~ dnorm(0, tau.w)
}
beta0[1] ~ dnorm(1,1.0E-6)
beta0[2] ~ dnorm(1,1.0E-6)
beta0[3] ~ dnorm(1,1.0E-6)
beta0[4] ~ dnorm(1,1.0E-6)
beta0[5] ~ dnorm(1,1.0E-6)
beta1[1] ~ dnorm(0,1.0E-6)
beta1[2] ~ dnorm(0,1.0E-6)
beta1[3] ~ dnorm(0,1.0E-6)
beta1[4] ~ dnorm(0,1.0E-6)
beta1[5] ~ dnorm(0,1.0E-6)
sigma[1] <- 1/tau[1]
sigma[2] <- 1/tau[2]
sigma[3] <- 1/tau[3]
sigma[4] <- 1/tau[4]
sigma[5] <- 1/tau[5]
sigma.w <- 1/tau.w
tau[1] ~ dgamma(0.001,0.001)
tau[2] ~ dgamma(0.001,0.001)
tau[3] ~ dgamma(0.001,0.001)
tau[4] ~ dgamma(0.001,0.001)
tau[5] ~ dgamma(0.001,0.001)
tau.w ~ dgamma(0.001,0.001)
lambda[1] ~ dnorm(0,1.0E-6)
lambda[2] ~ dnorm(0,1.0E-6)
lambda[3] ~ dnorm(0,1.0E-6)
lambda[4] ~ dnorm(0,1.0E-6)
lambda[5] ~ dnorm(0,1.0E-6)
for(i in 1:n) {
  uns[i] ~ dbern(L[i])
}
}
```

C.2 SAS - Proc NLMIXED

C.2.1 Programa - Dados Superliga de Vôlei

```

data volei;
input alr jogos y z;
datalines;
dados;
proc nlmixed data=volei;
parms beta01=1 beta02=1 beta03=1 beta11=0 beta12=0 beta13=0 s1=1 s2=1 s3=1;
mu1=beta01+beta11*z;
mu2=beta02+beta12*z;
mu3=beta03+beta13*z;
if alr=1 then like=1/sqrt(6.283185*(s1**2))*exp(-0.5*(1/(s1**2))*((y-mu1)**2));
if alr=2 then like=1/sqrt(6.283185*(s2**2))*exp(-0.5*(1/(s2**2))*((y-mu2)**2));
if alr=3 then like=1/sqrt(6.283185*(s3**2))*exp(-0.5*(1/(s3**2))*((y-mu3)**2));
ll=log(like);
model y ~ general(ll) subject=jogos;
run;

```

C.2.2 Programa - Dados Longitudinais (qualidade do leite)

```

data leite;
input alr y vaquinhas z;
datalines;
dados;
proc nlmixed data=leite;
bounds s1>0,s2>0,s3>0,s4>0,s5>0,sw>0;
parms beta01=1 beta02=1 beta03=1 beta04=1 beta05=1 beta11=0 beta12=0 beta13=0
beta14=0 beta15=0 s1=0.1 s2=0.1 s3=0.1 s4=0.1 s5=0.1 sw=0.11;
mu1=beta01+beta11*z+w;
mu2=beta02+beta12*z+w;
mu3=beta03+beta13*z+w;
mu4=beta04+beta14*z+w;
mu5=beta05+beta15*z+w;
if alr=1 then like=(1/sqrt(6.283185*(s1*s1)))*exp(-0.5*(1/(s1*s1))*((y-mu1)**2));

```

```

if alr=2 then like=(1/sqrt(6.283185*(s2*s2)))*exp(-0.5*(1/(s2*s2))*((y-mu2)**2));
if alr=3 then like=(1/sqrt(6.283185*(s3*s3)))*exp(-0.5*(1/(s3*s3))*((y-mu3)**2));
if alr=4 then like=(1/sqrt(6.283185*(s4*s4)))*exp(-0.5*(1/(s4*s4))*((y-mu4)**2));
if alr=5 then like=(1/sqrt(6.283185*(s5*s5)))*exp(-0.5*(1/(s5*s5))*((y-mu5)**2));
ll=log(like);
model y ~ general(ll);
random w ~ normal(0,sw*sw) subject=vaquinhas;
run;

```

C.3 R - Modelo de Regressão Dirichlet

```

leite <- read.table(file="Dados leite.csv",header=TRUE,sep=";")
a <- cbind(leite[1:30,],trat=0)
b <- cbind(leite[31:60,],trat=1)
dados.leite <- rbind(a[,,],b[,,])
dados.leite <- data.frame(dados.leite)
dados.leite[1:30,"trat"] <- "A"
dados.leite[31:60,"trat"] <- "D"

#####Regressão Dirichlet#####

library(DirichletReg)
leite_dir <- DR_data(dados.leite[,1:6])
summary(leite_dir)
regdirleite <- DirichReg(leite_dir ~ trat, dados.leite)
summary(regdirleite)
confint(regdirleite)
residuals(regdirleite)
fitted(regdirleite)

```