

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA

**ABORDAGEM BAYESIANA PARA MODELOS
MISTOS COM O USO DE “GIBBS SAMPLING”**

Monografia escrita para a
obtenção do grau de
Bacharel em Estatística

Autora:

Vanessa Bielefeldt Leotti

Orientadora:

Prof. Dra. Patrícia Klarmann Ziegelmann

Porto Alegre, Dezembro de 2005.

AGRADECIMENTOS

A Deus, meu Pai, ao Espírito Santo, meu Guia e a Jesus Cristo, meu Líder.

A minha família que tanto amo: meu pai, José, minha mãe, Leci, e meu irmão, Rafael. Pelo incentivo, apoio, conselhos e, principalmente, pelos sacrifícios feitos por mim para que eu chegasse aonde cheguei. Além disso, obrigada pelos valores ensinados e por constantemente cuidarem do meu caminho nesta vida.

Ao meu grande amor, Jean, por fazer-me sentir tão especial, por ter trazido a felicidade para a minha vida! Obrigada pela paciência, pelo incentivo, pela motivação e, sobretudo, pelo amor. Espero poder fazer o mesmo por ti nesta vida inteira que passaremos juntos.

Ao movimento católico de jovens CLJ, especialmente, a todos os padres, tios e jovens que estiveram comigo nestes seis anos de minha caminhada dentro do movimento. Devo a este grupo muito do que hoje tenho e sou. SHALOM!

Aos meus verdadeiros amigos, pela paciência e também por compreenderem minha ausência neste tempo todo. Em especial aos amigos da família Scheeren, pelo incentivo, pelas conversas, pela disponibilidade contínua em ajudar e também pelas caronas estes quatro anos. Deus abençoe a todos!

Aos colegas do curso, especialmente Marília e Kátia, pela ajuda e suporte nos momentos complicados. Que o futuro lhes reserve muita alegria!

Aos professores do curso, pelo valioso ensinamento, pela preocupação conosco demonstrada, pela disponibilidade contínua. Em especial pelo exemplo que quero seguir. Espero um dia ser colega de vocês! Desejo agradecer especialmente ao professor João Riboldi, pelo incentivo dado para que eu desenvolvesse meus potenciais e por ter me apresentado a pesquisa científica que hoje gosto tanto e quero seguir praticando; e também a professora Patrícia K. Ziegelmann, que tive o grande privilégio de ter como orientadora, pela paciência, incentivo, auxílio nas decisões que tive de tomar quanto ao meu futuro e em especial pelo grande exemplo. És um modelo para mim, tanto como profissional como quanto pessoa!

RESUMO

A principal diferença entre inferência Bayesiana e inferência clássica é que a primeira considera os parâmetros desconhecidos dos problemas estatísticos estudados como variáveis aleatórias, enquanto a segunda os considera constantes. Segundo a abordagem Bayesiana, antes de qualquer dado ser coletado deve-se assumir uma distribuição de probabilidade para os parâmetros desconhecidos do modelo, chamada de distribuição a priori, a qual deve ser especificada juntando-se toda informação existente sobre o parâmetro. Após os dados serem observados, a informação dada pela priori é combinada, através do teorema de Bayes, com a informação proveniente dos dados, que é representada pela função de verossimilhança dos parâmetros em função dos dados observados. Esta combinação resulta na distribuição a posteriori dos parâmetros desconhecidos que proverá toda a informação sobre estes após os dados terem sido observados. Para se chegar analiticamente à distribuição a posteriori, alguns cálculos são necessários, cálculos estes que nem sempre são exequíveis. Os métodos MCMC (abreviatura de “Markov Chain Monte Carlo”), que são métodos computacionais de simulação baseados em cadeias de Markov, são muito úteis para problemas Bayesianos em que a distribuição a posteriori não pode ser obtida analiticamente (e também para alguns problemas clássicos). Neste contexto, eles permitem que sejam simuladas observações provenientes da posteriori de interesse, e com estas observações, pode-se estimar momentos, quantis e até mesmo a densidade da distribuição. Um dos métodos MCMC é o algoritmo de “Gibbs Sampling”. Neste trabalho, o “Gibbs Sampling” será utilizado para analisar modelos mistos. O modelo misto é um modelo que compreende efeitos fixos e aleatórios em sua formulação e que assume distribuição multivariada para os erros do modelo. Desta forma, este modelo permite analisar dados que não são independentes tais como dados que são de alguma forma agrupados, como, por exemplo, quando se tomam múltiplas medidas de um mesmo sujeito. O modelo misto é um exemplo de modelo onde a distribuição a posteriori de interesse não é analiticamente derivável. Este trabalho trata da aplicação das técnicas Bayesianas ao modelo misto, em especial, da utilização do “Gibbs Sampling” para as análises deste modelo. Assim, o objetivo principal é introduzir os métodos MCMC, com enfoque no “Gibbs Sampling”, e estudar a sua aplicação em problemas estatísticos envolvendo modelos mistos. Além de abordar a parte teórica, serão apresentados exemplos envolvendo dados reais e dados simulados. No decorrer do trabalho, percebeu-se que o “Gibbs Sampling” é um algoritmo realmente eficiente e versátil para a análise de vários problemas Bayesianos, especialmente no caso dos modelos mistos.

SUMMARY

The main difference between Bayesian and classical inference is that the first considers the unknown parameters of the statistical model as random variables, while the second considers them as constants. In Bayesian framework, before any observation is collected, a probability distribution for the unknown parameters of the model must be assumed. This is called prior distribution, which must be specified by joining all the existent information about the parameter. After the data are collected, the information provided by the prior distribution is combined, through the Bayes theorem, with the information provided by the data, that is represented by the likelihood function of the parameters. The combination results in the posterior distribution of the unknown parameters of the model, which will provide all the information about them after the data is observed. Some calculus are necessary to obtain analytically the posterior distribution. However this is sometimes unfeasible. The MCMC methods (from "Markov Chain Monte Carlo") are computational methods of simulation based on Markov chains and are very useful in Bayesian problems when the posterior distribution can not be analytically obtained (and for some classical problems too). In this context, observations are simulated from the unknown posterior of interest. Therefore, moments, quantiles and even the full density are estimated. One of the MCMC methods is the Gibbs Sampling algorithm. In this work, "Gibbs Sampling" will be used to analyze mixed models. The mixed model is a model that involves fixed and random effects in its formulation and assumes a multivariate distribution for the errors of the model. As so, this model allows data that are either not independent or data that are in some way clustered to be analyzed, like, for example, when multiple measures are taken from the same subject. The mixed model is an example where the posterior distribution is not analytically derivable. This work discusses about the application of the Bayesian techniques to the mixed model, in special, about the use of Gibbs Sampling to analyze it. So, the main objective is to introduce the MCMC methods, with emphasis on Gibbs Sampling, and to study their application to statistical problems involving mixed models. Besides the discussion about the theory, examples involving real and simulated data will be presented. In the course of this work, it was noticed that Gibbs Sampling is a really efficient and versatile algorithm for the analyses of many Bayesian problems, in special in the case of mixed models.

ÍNDICE

1. INTRODUÇÃO.....	1
2. INTRODUÇÃO AOS MÉTODOS MCMC	5
2.1. CADEIAS DE MARKOV	6
2.2. ALGORITMO DE METROPOLIS-HASTINGS	7
2.3. “GIBBS SAMPLING”	9
2.3.1. <i>Definição e propriedades</i>	9
2.3.2. <i>Geração da amostra, valores iniciais e convergência</i>	10
2.3.3. <i>Análise dos resultados da simulação</i>	12
2.3.4. <i>Exemplo Normal Bivariada</i>	14
3. ABORDAGEM CLÁSSICA PARA MODELOS MISTOS NORMAIS.....	18
3.1. FORMULAÇÃO	18
3.2. ESTIMAÇÃO	20
3.2.1. <i>Estimação dos efeitos fixos e aleatórios</i>	21
3.2.2. <i>Estimação dos parâmetros de variâncias e covariâncias</i>	24
3.3. INTERVALOS DE CONFIANÇA E TESTE DE HIPÓTESES	26
3.4. EXEMPLOS.....	28
3.4.1. <i>Modelo de Componentes de Variância</i>	28
3.4.2. <i>Modelo para Dados Longitudinais – Banco de dados Weights</i>	30
4. ABORDAGEM BAYESIANA PARA MODELOS MISTOS NORMAIS.....	40
4.1. CONCEITOS BÁSICOS DE INFERÊNCIA BAYESIANA	41
4.2. ANÁLISE BAYESIANA PARA MODELOS MISTOS NORMAIS	44
4.3. DISTRIBUIÇÕES CONDICIONAIS COMPLETAS PARA MODELOS MISTOS NORMAIS.....	46
4.3.1. <i>Modelo de Componentes de Variância</i>	46
4.3.2. <i>Modelo para Dados Longitudinais</i>	47
4.4. EXEMPLOS.....	51
4.4.1. <i>Modelo de Componentes de Variância</i>	51
4.4.2. <i>Modelo para Dados Longitudinais – Banco de dados Weights</i>	54
4.5. ESTUDO DE SIMULAÇÃO DE DADOS LONGITUDINAIS	64
5. CONSIDERAÇÕES FINAIS	71
REFERÊNCIAS BIBLIOGRÁFICAS.....	73
APÊNDICE A – Programação em R para o Exemplo 2.3.4	75
APÊNDICE B – Programação em R para os Exemplos 3.4.1 e 4.4.1	77
APÊNDICE C – Programação no Winbugs para o Exemplo 4.4.2	81
APÊNDICE D – Programação em R para o Estudo de Simulação em 4.5.....	86

1. INTRODUÇÃO

A Inferência Bayesiana é uma abordagem estatística que considera os parâmetros desconhecidos dos diversos modelos como variáveis aleatórias. As inferências feitas sobre estes parâmetros baseiam-se na combinação de informações a priori sobre eles, provenientes de opiniões de especialistas, por exemplo, com os resultados amostrais. Esta combinação define a distribuição a posteriori dos parâmetros, que é o fundamento de toda análise estatística baseada na inferência Bayesiana.

O modelo misto é um modelo que compreende efeitos fixos e aleatórios em sua formulação. Efeitos fixos são variáveis independentes que se supõe que assumam valores constantes que podem ser escolhidos arbitrariamente pelo pesquisador. Efeitos aleatórios são variáveis independentes que assumem valores considerados aleatórios, e supõe-se uma distribuição para estas variáveis. Outra característica do modelo misto é o fato de que se assume uma distribuição multivariada para os erros do modelo, o que permite analisar dados que não sejam independentes. Por isto, o modelo misto é bastante útil quando se tem dados que são agrupados de alguma forma, como, por exemplo, alunos agrupados em turmas, pacientes agrupados em hospitais, etc. Nestes casos, a correlação inerente que existe entre observações de um mesmo grupo pode ser incorporada no modelo. Analogamente, é um modelo que tem encontrado grande aplicação para análise de medidas repetidas, que são estudos onde múltiplas medidas são feitas em mesmas unidades experimentais (indivíduos, objetos, etc). Para este tipo de dados, é coerente supor-se a existência de correlação entre as observações feitas na mesma unidade. Normalmente, supõe-se distribuição normal multivariada para estes erros, obtendo-se o modelo misto normal, que será o modelo focado neste trabalho.

Devido à complexidade da função de verossimilhança do modelo misto, não é possível se obter uma expressão fechada para a distribuição a posteriori dos parâmetros, mesmo utilizando-se prioris conjugadas. Assim, o algoritmo de “Gibbs Sampling”, um dos métodos MCMC – “Markov Chain Monte Carlo”, fornece um método computacional eficaz para se obter estimativas da distribuição a posteriori.

Os objetivos desta monografia são os seguintes:

- Introduzir os métodos MCMC;

- Abordar a técnica de “Gibbs Sampling”, descrevendo histórico, utilidade, descrição do algoritmo, embasamento teórico e um exemplo didático;

- Revisar a abordagem clássica para modelos mistos, apresentando a formulação do modelo, métodos de estimação e exemplos com solução computacional através do software SAS;

- Apresentar a abordagem Bayesiana para modelos mistos, exemplificando com dados reais e simulados analisados através dos softwares R e WinBUGS e discutindo o uso destes softwares;

- Comparar os resultados da estimação clássica com os obtidos através da análise Bayesiana supondo-se prioris não informativas. Neste trabalho não serão utilizadas prioris informativas.

Em relação aos métodos MCMC, as duas principais referências utilizadas nesta monografia foram os livros de Gamerman (1997) e Gilks et al. (1996). Gamerman (1997) apresenta uma abordagem mais teórica dos métodos MCMC, com exemplos mais voltados aos cálculos analíticos do que à aplicação dos métodos. É uma leitura muito interessante para aqueles que não tem familiaridade com a Inferência Bayesiana, especialmente o Capítulo 2 do livro, que é dedicado a revisar os principais conceitos da Inferência Bayesiana. Já Gilks et al. (1996) apresenta os métodos MCMC aplicados em várias áreas do conhecimento e exige um pouco mais de conhecimento Bayesiano prévio do leitor.

Especificamente sobre o “Gibbs Sampling”, o artigo de Casella e George (1992) é um bom texto introdutório do assunto, com uma linguagem muito acessível. Apresenta exemplos didáticos e aplicados, e a maneira de apresentar os conceitos envolvidos é de fácil compreensão.

Sobre o modelo misto, este trabalho tem como base os livros de Brown e Prescott (1999), Davidian e Giltinan (1995) e Searle et al. (1992). Brown e Prescott (1999) apresentam um texto bastante claro, não muito teórico e com bastantes exemplos voltados à área médica e analisados usando o software estatístico SAS. A abordagem Bayesiana dos modelos mistos também é tratada, mas de maneira superficial. Davidian e Giltinan (1995) e Searle et al. (1992) são livros bastante teóricos e de enfoque clássico. Discutem também sobre alguns tópicos da abordagem Bayesiana dos modelos mistos, através da idéia de formular um modelo

misto com estrutura hierárquica, idéia esta que será mais discutida no decorrer do trabalho. Davidian e Giltinan (1995) inclusive mostram como aplicar o “Gibbs Sampling” para alguns exemplos de modelo misto. Já Searle et al. (1992) e Brown e Prescott (1999) somente comentam sobre a possibilidade de usar o algoritmo.

O artigo de Gilks et al. (1993) e o Capítulo 17 do livro de Gilks et al. (1996) são duas referências que abordam o enfoque Bayesiano ao modelo misto para dados longitudinais, onde ambos usam o algoritmo de “Gibbs Sampling” para analisar seus exemplos.

Além desta introdução, este trabalho possui mais quatro capítulos. No Capítulo 2, os métodos MCMC serão introduzidos, explicando-se brevemente sua fundamentação teórica através de uma seção que tratará de Cadeias de Markov. Após, o capítulo enfoca o “Gibbs Sampling”, descrevendo o algoritmo e sua utilização, que não se restringe somente a problemas Bayesianos. Este capítulo, assim como os dois seguintes, encerra com a ilustração dos conceitos apresentados através de um exemplo.

No Capítulo 3, o modelo misto é apresentado em sua conceituação clássica. Sua definição teórica será descrita, e técnicas clássicas de análise, envolvendo métodos de estimação, também serão abordadas. O objetivo de apresentar a abordagem clássica do modelo misto é de meramente fornecer uma base de comparação entre o método clássico e o método Bayesiano de análise para este modelo. Ao final do capítulo, dois exemplos são analisados no software SAS, um com dados simulados e outro com dados reais.

No Capítulo 4, a abordagem Bayesiana para os modelos mistos é introduzida. O capítulo inicia com alguns conceitos básicos de inferência Bayesiana, contextualizando o leitor não muito habituado com esta abordagem estatística. Após, trata-se do caso dos modelos mistos, onde as distribuições a priori mais adequadas ao problema são mostradas e a impossibilidade de calcular-se a distribuição a posteriori analiticamente é discutida. O capítulo segue mostrando como aplicar o “Gibbs Sampling” para os modelos mistos, discutindo sobre os aspectos computacionais dos softwares utilizados, o R e o WinBUGS. Após, os dois exemplos analisados de forma clássica no Capítulo 3 são revisitados, aplicando-se os conceitos Bayesianos vistos no capítulo. Um estudo de simulação para analisar algumas propriedades dos estimadores Bayesianos é descrito na última seção do capítulo.

Finalmente, o Capítulo 5 apresenta as considerações finais do trabalho e sugestões para futuras pesquisas envolvendo o tema desta monografia.

Para auxiliar àqueles interessados nos métodos computacionais com que se procederam as análises, todos os códigos de programação utilizados no R e no WinBUGS para se analisar os exemplos estão apresentados nos apêndices.

2. INTRODUÇÃO AOS MÉTODOS MCMC

Atualmente, muitos problemas estatísticos são da categoria de modelos de grande dimensão, pois contêm muitos parâmetros. Gamerman (1997) cita como exemplo os modelos para dados espaciais. Até recentemente, trabalhar com a estrutura complexa destes modelos em muitas aplicações era difícil, pois requeria o desenvolvimento de metodologia e softwares específicos para cada caso (Gilks et al., 1996). Os métodos envolvendo simulação de Monte Carlo usando cadeias de Markov, conhecidos como métodos MCMC, da expressão em inglês “Markov Chain Monte Carlo”, fornecem uma metodologia única para analisar muitos destes problemas.

Os métodos Monte Carlo são úteis quando se tem uma variável aleatória X com distribuição conhecida e consegue-se, através de algum algoritmo, gerar dados dela, estando-se interessado na distribuição de uma função $g_X(\cdot)$, que pode ser impossível de ser trabalhada analiticamente. Assim, a simulação de Monte Carlo consiste basicamente em gerar valores independentes de X , calcular $g(x_i)$ para cada x_i simulado e se aproximar características da distribuição de $g_X(\cdot)$ através dos valores simulados. Utilizando-se MCMC, apesar de não se conseguir gerar observações diretamente da distribuição de interesse, geram-se amostras a partir de cadeias de Markov construídas de tal forma que, após um longo tempo, a distribuição das observações geradas será aquela de interesse. Com a amostra construída, aproximam-se as características de interesse da distribuição utilizando as respectivas características amostrais. As cadeias de Markov serão definidas e terão algumas de suas propriedades descritas na Seção 2.1.

Duas grandes vantagens dos métodos MCMC são: não é necessário conhecer a forma da distribuição de que se quer simular, o que ocorre em muitos problemas práticos na inferência Bayesiana; e há muitos algoritmos para se construir as cadeias de Markov necessárias para a simulação, sendo que todos atingem o objetivo de gerar dados da distribuição de interesse. Entretanto, todos os algoritmos são casos particulares do algoritmo Metropolis-Hastings, que será descrito na Seção 2.2. Um destes casos é o “Gibbs Sampling”, que será abordado e exemplificado na Seção 2.3.

Gilks et al. (1996) comenta que foram necessários cerca de 40 anos para o MCMC penetrar nos principais problemas práticos da Estatística. Sua origem está na literatura de física-estatística, e tem sido utilizado por mais de uma década em estatística espacial e análise

de imagens. Nos últimos anos, houve um grande crescimento no número de publicações envolvendo aplicações dos métodos MCMC na inferência Bayesiana e também na estatística clássica.

2.1. Cadeias de Markov

Nesta seção, alguns conceitos de cadeias de Markov serão brevemente apresentados. Para aqueles interessados em obter maior conhecimento sobre as cadeias de Markov sugere-se que vejam, por exemplo, Ross (1996).

Uma cadeia de Markov é um tipo especial de processo estocástico. Um processo estocástico pode ser definido como uma coleção de variáveis aleatórias $\{\theta^{(t)} : t \in T\}$ para algum conjunto T , que pode ser discreto ou contínuo. Aqui, somente se trabalhará com o caso discreto, mais especificamente onde $T \in \mathbf{N}$, e assim pode-se pensar o índice das variáveis aleatórias como o índice da iteração de uma simulação. O espaço dos possíveis valores de $\theta^{(t)}$ é denominado espaço de estados e denotado por S .

A principal característica de uma cadeia de Markov é que, a cada tempo $t \geq 0$, o próximo estado da cadeia, $\theta^{(t+1)}$, depende do passado somente através do atual estado, $\theta^{(t)}$. Isto pode ser formalmente definido como:

$$P(\theta^{(t+1)} = y | \theta^{(t)} = x_t, \theta^{(t-1)} = x_{t-1}, \dots, \theta^{(t-k)} = x_{t-k}, \dots, \theta^{(0)} = x_0) = P(\theta^{(t+1)} = y | \theta^{(t)} = x_t) \quad (2.1)$$

para todos $x_0, \dots, x_t, y \in S$, $k \in \mathbf{N}$.

Em geral, as probabilidades em (2.1) dependem de x_t, y e t . Quando não dependem de t , a cadeia é chamada de homogênea. Neste caso, a equação em (2.1) é denotada por P_{xy} e chamada de probabilidade de transição estacionária a um passo.

Estas probabilidades satisfazem:

a. $P_{xy} \geq 0, \forall x, y \in \mathbf{R}$;

b. $\sum_{y \in S} P_{xy} = 1, \forall x \in S$.

Define-se também $P(\theta^{(t+1)} | \theta^{(t)})$ como função, ou *kernel*, de transição da cadeia.

Um problema fundamental das cadeias de Markov no contexto da simulação é o estudo do comportamento assintótico da cadeia quando o número de iterações $t \rightarrow \infty$. Para entender este comportamento, é necessário apresentar-se o conceito de distribuição estacionária, denotada por π . A distribuição π é dita a distribuição estacionária de uma cadeia com probabilidades de transição P_{xy} se

$$\sum_{x \in S} \pi_x P_{xy} = \pi_y, \forall y \in S.$$

Esta distribuição também é conhecida como distribuição invariante ou de equilíbrio.

Sob certas condições de regularidade (que não serão abordadas aqui), tem-se a seguinte relação entre a probabilidade de transição a t passos, denotada por $P_{xy}^{(t)}$ e a distribuição estacionária:

$$\lim_{t \rightarrow \infty} P_{xy}^{(t)} = \pi_y. \quad (2.2)$$

Seja o estado inicial da cadeia $\theta^{(0)} = x$. Por (2.2), conclui-se que a cadeia irá gradualmente “esquecer” seu estado inicial e $P_x^{(t)}$ irá convergir para a distribuição estacionária, que se pode demonstrar ser única. Denota-se por $P_x^{(t)}$ a distribuição de probabilidade que fornece as probabilidades com que a cadeia sai de seu ponto inicial x e visita qualquer outro estado do espaço de estados, em t passos. Assim, à medida que t aumenta, os pontos $\theta^{(t)}$ gerados a partir da função de transição $P(\theta^{(t+1)} | \theta^{(t)})$ irão cada vez mais se parecer com amostras de π .

Logo, a questão essencial da simulação por MCMC é definir qual função de transição utilizar para que a distribuição estacionária π seja aquela em que estamos interessados. Como já dito anteriormente, há muitas maneiras de se escolher a função de transição para obter amostras de π , porém, todos eles são casos especiais do algoritmo de Metropolis–Hastings, que será visto a seguir.

2.2. Algoritmo de Metropolis-Hastings

Hastings (1970) generalizou o método inicialmente proposto por Metropolis et al. (1953), originando assim o algoritmo de Metropolis-Hastings. A idéia do algoritmo é que, a cada tempo t , gera-se um ponto candidato θ^c a partir de uma distribuição proposta $Q(\cdot | \theta^{(t)})$

e então aceita-se que o próximo estado da cadeia, $\theta^{(t+1)}$, seja igual a θ^c com probabilidade $\alpha(\theta^{(t)}, \theta^c)$, onde

$$\alpha(\theta^{(t)}, \theta^c) = \min\left(1, \frac{\pi(\theta^c)Q(\theta^{(t)} | \theta^c)}{\pi(\theta^{(t)})Q(\theta^c | \theta^{(t)})}\right). \quad (2.3)$$

Se o ponto candidato não for aceito, a cadeia não se move, isto é, $\theta^{(t+1)} = \theta^{(t)}$.

Portanto, o algoritmo é definido formalmente como:

1. Defina $t = 0$. Forneça um valor arbitrário para $\theta^{(0)}$;
2. Gere um valor θ^c a partir de $Q(\cdot | \theta^{(t)})$;
3. Gere um valor U de uma distribuição *Uniforme*(0,1);
4. Se $U \leq \alpha(\theta^{(t)}, \theta^c)$, faça $\theta^{(t+1)} = \theta^c$, caso contrário, faça $\theta^{(t+1)} = \theta^{(t)}$;
5. Incremente t ;
6. Volte ao passo 2 e continue as iterações até um tamanho m suficientemente grande.

Pode ser mostrado que, sob certas condições de regularidade, a distribuição estacionária da cadeia será $\pi(\cdot)$, independente da forma da distribuição proposta $Q(\cdot | \cdot)$. Entretanto, ao analisar-se a expressão em (2.3), percebe-se que é necessário conhecer $\pi(\cdot)$ para implementar o algoritmo, o que parece o tornar não muito útil.

Em problemas Bayesianos, $\pi(\cdot)$ será a distribuição a posteriori dos parâmetros, que é dada por:

$$\pi(\cdot) = \pi(\theta | x) = \frac{\pi(\theta)L(\theta | x)}{f(x)}, \quad (2.4)$$

onde $\pi(\theta)$ é a distribuição a priori, que sempre será conhecida, $L(\theta | x)$ é a função de verossimilhança de θ para os dados observados, que também sempre terá sua forma supostamente conhecida, e $f(x) = \int \pi(\theta)L(\theta | x)d\theta$ que pode não ter forma analítica fechada em alguns casos. Note, porém, que $f(x)$ é só uma constante de normalização, ou seja, ela nunca envolverá θ em sua expressão. Por isto, substituindo (2.4) em (2.3), teremos:

$$\frac{\pi(\theta^c)Q(\theta^{(t)}|\theta^c)}{\pi(\theta^{(t)})Q(\theta^c|\theta^{(t)})} = \frac{\pi(\theta^c|x)Q(\theta^{(t)}|\theta^c)}{\pi(\theta^{(t)}|x)Q(\theta^c|\theta^{(t)})} = \dots = \frac{\pi(\theta^c)L(\theta^c|x)Q(\theta^{(t)}|x)}{\pi(\theta^{(t)})L(\theta^{(t)}|x)Q(\theta^c|\theta^{(t)})}$$

que só envolve expressões conhecidas. Ou seja, $\pi(\theta|x)$ só será desconhecida se $f(x)$ não for derivável analiticamente.

A seguir será apresentada a técnica de “Gibbs Sampling”, também conhecida na Língua Portuguesa como Amostrador de Gibbs e que é adequada quando o modelo possui mais de um parâmetro desconhecido. Ela é basicamente o algoritmo de Metropolis-Hastings onde a distribuição proposta para cada parâmetro a ser simulado será a respectiva distribuição condicional completa e a probabilidade de aceitação de cada valor gerado será sempre igual a um.

2.3. “Gibbs Sampling”

A origem do “Gibbs Sampling” está no artigo de Geman e Geman (1984) que trata de processamento de imagens. Neste artigo, os autores utilizaram o algoritmo para simular dados de uma distribuição de Gibbs, por isto a denominação de “Gibbs Sampling”. Entretanto, somente com o trabalho de Gelfand e Smith (1990), a técnica encontrou aplicações em vários problemas estatísticos além da análise de imagens.

2.3.1. Definição e propriedades

Seja θ um vetor $p \times 1$, ou seja, $\theta' = (\theta_1, \dots, \theta_p)$, onde cada componente θ_i , $i = 1, \dots, p$, pode ser um escalar, um vetor ou uma matriz. O interesse está na distribuição $\pi(\theta)$ que, nos problemas práticos, pode não ter uma densidade de probabilidade analiticamente derivável ou possuir um algoritmo de geração direta muito trabalhoso, complicado ou simplesmente inexistente. Se as distribuições $\pi(\theta_i | \theta_{-i})$, $i = 1, \dots, p$ e $\theta'_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$, denominadas distribuições condicionais completas, são totalmente conhecidas e pode-se simular observações a partir delas, então se pode aplicar o “Gibbs Sampling” através do seguinte algoritmo:

1. Forneça valores iniciais para $\theta^{(0)} = (\theta_1^{(0)}, \dots, \theta_p^{(0)})$. Defina $t = 0$;
2. Obtenha o próximo estado da cadeia, $\theta^{(t+1)} = (\theta_1^{(t+1)}, \dots, \theta_p^{(t+1)})$, a partir das seguintes gerações sucessivas

$$\begin{aligned}\theta_1^{(t+1)} &\sim \pi(\theta_1 | \theta_2^{(t)}, \dots, \theta_p^{(t)}) \\ \theta_2^{(t+1)} &\sim \pi(\theta_2 | \theta_1^{(t+1)}, \theta_3^{(t)}, \dots, \theta_p^{(t)}) \\ &\vdots \\ \theta_p^{(t+1)} &\sim \pi(\theta_p | \theta_1^{(t+1)}, \dots, \theta_{p-1}^{(t+1)})\end{aligned}$$

3. Incremente t ;
4. Volte ao passo 2 e continue as iterações até um tamanho m que garanta a convergência do algoritmo (a próxima subseção tratará de convergência).

Como já mencionado, o “Gibbs Sampling” é um caso particular do algoritmo de Metropolis-Hastings, com distribuições propostas iguais às condicionais completas. Com alguns cálculos, demonstra-se que a probabilidade de aceitação de cada valor gerado através do “Gibbs Sampling” é igual a 1 e por isso a cadeia sempre se moverá.

Por sua estruturação, nota-se que o “Gibbs Sampling” reduz a dimensão do problema, pois ao invés de gerar amostras do vetor θ inteiro, gera valores para blocos de componentes de θ de menores dimensões.

Também se percebe que o algoritmo define uma cadeia de Markov, já que cada novo estado da cadeia, $\theta^{(t+1)}$, depende somente dos valores obtidos na iteração t . Demonstra-se que a distribuição estacionária da cadeia de Markov definida pelo “Gibbs Sampling” é π . Assim, quando $m \rightarrow \infty$, os valores simulados para θ são amostras de $\pi(\theta)$ e as observações geradas para cada θ_i , $i = 1, \dots, p$, são amostras de $\pi(\theta_i)$. Uma prova formal da convergência do algoritmo, bem como as condições para o seu acontecimento, pode ser encontrada em Tierney (1994).

2.3.2. Geração da amostra, valores iniciais e convergência

Como já mencionado, o “Gibbs Sampling” define uma cadeia de Markov, e por isto a amostra gerada através dele não é independente. Cada novo valor depende dos valores gerados na iteração anterior. Por este motivo, há duas maneiras apresentadas na literatura de se obter uma amostra aproximadamente independente de tamanho n da distribuição de interesse π . A primeira forma é replicar o algoritmo n vezes, em cada uma delas fazendo m iterações e então tomar o m -ésimo valor gerado de θ como um dos elementos da amostra de

tamanho n . Esta simulação precisará de mn gerações da cadeia. Se for estabelecido valores para os estados iniciais da cadeia $\theta_j^{(0)}$, $j = 1, \dots, n$, de forma independente, a amostra consiste em valores independentes de $\pi(\theta)$.

Outra forma é considerar uma única replicação do algoritmo, baseando-se no fato de que, após a convergência, os valores simulados da cadeia seguem a distribuição estacionária $\pi(\theta)$. Assim, os b primeiros valores gerados são considerados como um período de aquecimento do algoritmo (da expressão em inglês “burn-in”), necessário para a convergência do “Gibbs Sampling”, e são descartados, utilizando-se somente os n valores posteriores como uma amostra da distribuição de interesse. Para diminuir a dependência entre os valores simulados, pode-se determinar um valor k , conhecido como “thin” em inglês, pelo fato deste procedimento “emagrecer” a amostra. A amostra será formada pelos valores gerados da cadeia a cada k -ésima iteração após o “burn-in”, ou seja, $\theta_i^{(b+1)}, \theta_i^{(b+k+1)}, \theta_i^{(b+2k+1)}, \dots$. Entretanto, não se ganha em eficiência utilizando este procedimento e pode-se mostrar que a estimação é sempre menos precisa do que se mantendo na amostra todos os valores gerados após o “burn-in” (Gamerman, 1997).

Gilks et al. (1996) e Gamerman (1997) comentam que há conflitos na literatura sobre qual abordagem utilizar. Porém, ambos não recomendam a geração de n cadeias independentes, pois o custo computacional é maior, enquanto uma cadeia mais longa diminuiria a dependência do valor inicial.

Sobre o valor inicial, $\theta^{(0)}$, a ser estabelecido, não há muito material escrito, já que, teoricamente, ele não afeta a distribuição estacionária da cadeia. Devem ser escolhidos com mais cuidado para cadeias de convergência mais lenta, para se evitar um longo “burn-in” (Gilks et al., 1996).

O estabelecimento do tamanho do “burn-in” não pode ser realizado analiticamente na maioria dos casos. Entretanto, existem métodos informais e formais para se verificar a convergência da cadeia. O método informal mais utilizado baseia-se em gráficos da trajetória da cadeia, para verificar a partir de qual iteração os dados passam a apresentar um mesmo padrão de comportamento (veja exemplo na Subseção 2.3.4). Os métodos formais, conhecidos como diagnósticos de convergência, usam uma variedade de resultados teóricos e

aproximações, mas todos utilizam os valores simulados de alguma maneira. Cowles e Carlin (1994) e Brooks e Roberts (1998) apresentam uma revisão dos principais métodos.

Existem também algoritmos de diagnóstico de convergência, como o CODA, anagrama para a abreviatura da expressão “Convergence Diagnostics and Output Analysis”, introduzido por Best et al. (1995). Este programa está disponível como um pacote para o software R e inserido no WinBUGS. Este último é um software específico para a implementação do “Gibbs Sampling”, BUGS significa “Bayesian inference Using Gibbs Sampling”, cuja primeira versão foi apresentada por Spiegelhalter et al. (1995) e é distribuído gratuitamente na internet, assim como o R. Mais detalhes sobre a implementação do “Gibbs Sampling” nos pacotes R e WinBUGS serão dados na Seção 4.3.

Neste trabalho, adotou-se o procedimento de gerar uma única cadeia e desprezar os valores gerados até o “burn-in” escolhido, ou seja, definiu-se um “thin” igual a zero. A determinação do “burn-in” foi feita somente com base na inspeção visual de gráficos da trajetória da cadeia.

2.3.3. Análise dos resultados da simulação

Considerando-se uma simulação por “Gibbs Sampling” através do algoritmo descrito na Subseção 2.3.1, com uma única cadeia de tamanho m e um “burn-in” de tamanho b , pode-se estimar a média e a variância de θ_i como:

$$\bar{\theta}_i = \frac{1}{m-b} \sum_{t=b+1}^m \theta_i^{(t)} \quad (2.6)$$

$$e \ s_i^2 = \frac{1}{m-b-1} \sum_{t=b+1}^m (\theta_i^{(t)} - \bar{\theta}_i)^2. \quad (2.7)$$

Também qualquer quantil da distribuição, como, por exemplo, a mediana, é estimado obtendo-se o quantil correspondente na amostra constituída dos valores gerados após o “burn-in”.

A própria densidade marginal de θ_i pode ser estimada através da expressão a seguir:

$$\hat{\pi}(\theta_i) = \frac{1}{m-b} \sum_{t=b+1}^m \pi(\theta_i | \boldsymbol{\theta}_{-i}^{(t)}). \quad (2.8)$$

Este estimador para $\pi(\theta_i)$ é baseado na esperança da densidade condicional

$$E[\pi(\theta_i | \boldsymbol{\theta}_{-i})] = \int \pi(\theta_i | \boldsymbol{\theta}_{-i}) \pi(\boldsymbol{\theta}_{-i}) d\boldsymbol{\theta}_{-i} = \pi(\theta_i). \quad (2.9)$$

Logo, o estimador proposto por (2.8) imita (2.9), já que $\boldsymbol{\theta}_{-i}^{(t)}$, $t = b+1, \dots, m$ são considerados uma amostra de $\pi(\boldsymbol{\theta}_{-i})$ (Casella e George, 1992). Deve-se perceber que a expressão em (2.8) estima o valor da densidade de probabilidade $\pi(\theta_i)$ para um específico valor de θ_i . Para se obter uma estimativa da densidade de toda a distribuição, pode-se arbitrar um conjunto de valores representativos do espaço amostral de θ_i e avaliar a expressão em (2.8) para estes valores.

Quando θ_i é contínuo, a moda da sua distribuição não pode ser estimada obtendo-se o valor mais freqüente dentre os dados gerados a partir do “burn-in”. Isto acontece porque, ao se gerar números aleatórios com várias casas decimais, dificilmente serão obtidos valores idênticos. Porém, para solucionar este problema, pode-se calcular a densidade marginal estimada por (2.8) e observar o valor de θ_i que a maximiza. Outro método possível, computacionalmente mais rápido que a estimativa da densidade através de (2.8), seria estimar a densidade marginal de θ_i de maneira não-paramétrica, e também observar o valor que maximiza esta densidade.

Para uma função real de θ_i , denotada por $\psi = t(\theta_i)$, os estimadores para média e variância são análogos aos apresentados em (2.6) e (2.7). Ou seja,

$$\hat{E}(\psi) = \bar{\psi} = \frac{1}{m-b} \sum_{t=b+1}^m \psi_t = \frac{1}{m-b} \sum_{t=b+1}^m t(\theta_i^{(t)})$$

$$\text{e } s_{\psi}^2 = \frac{1}{m-b-1} \sum_{t=b+1}^m (\psi_t - \bar{\psi})^2.$$

Pode-se demonstrar que todos os estimadores apresentados nesta seção são consistentes (Gamerman, 1997).

A seguir um exemplo de utilização do “Gibbs Sampling” será apresentado. É um exemplo didático e de abordagem clássica, onde simularam-se dados de uma distribuição normal bivariada.

2.3.4. Exemplo Normal Bivariada

O objetivo principal deste exemplo é ilustrar a aplicação do “Gibbs Sampling” em uma situação fora do contexto Bayesiano, cujo intuito é meramente didático. Será simulada através de “Gibbs Sampling” a distribuição conjunta de X e Y , e as distribuições marginais de cada uma destas variáveis, sendo que as funções densidade de probabilidade de cada uma destas distribuições já são conhecidas.

Seja o par de variáveis aleatórias $(X, Y) \sim N\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}; \begin{bmatrix} \sigma_X^2 \\ \sigma_Y^2 \end{bmatrix}; \rho\right)$. Por isto, sabe-se que:

$$\blacksquare f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\};$$

$$\blacksquare f(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{-\frac{1}{2\sigma_X^2}(x-\mu_X)^2\right\}, \text{ portanto uma distribuição } N(\mu_X, \sigma_X^2);$$

$$\blacksquare f(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left\{-\frac{1}{2\sigma_Y^2}(y-\mu_Y)^2\right\}, \text{ portanto uma distribuição } N(\mu_Y, \sigma_Y^2);$$

$$\blacksquare f(x|y) = \frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}} \exp\left\{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left[x-\mu_X - \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y)\right]^2\right\},$$

portanto uma distribuição $N\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y-\mu_Y), \sigma_X^2(1-\rho^2)\right)$;

$$\blacksquare f(y|x) = \frac{1}{\sqrt{2\pi\sigma_Y^2(1-\rho^2)}} \exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left[y-\mu_Y - \frac{\rho\sigma_Y}{\sigma_X}(x-\mu_X)\right]^2\right\}, \text{ portanto}$$

uma distribuição $N\left(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2(1-\rho^2)\right)$.

Considere $\mu_X = 3; \sigma_X^2 = 1; \mu_Y = 0; \sigma_Y^2 = 25; \rho = 0,8$. Logo, as condicionais completas que são necessárias para aplicar o “Gibbs Sampling” serão:

$$(x|y) \sim N\left(3 + \frac{0,8}{5}(y); (1-0,8^2)\right) = N(3 + 0,16y; 0,36) \text{ e}$$

$$(y|x) \sim N(0,8(5)(x-3); 25(1-0,8^2)) = N(4(x-3); 9).$$

Propositadamente, foram escolhidos dois valores iniciais para as cadeias bastante extremos em relação à distribuição marginal verdadeira das variáveis aleatórias X e Y : $x^{(0)} = -5$; $y^{(0)} = 30$. Através do algoritmo “Gibbs Sampling”, implementado com o auxílio do software R (veja no Apêndice A a programação utilizada), primeiramente simulou-se uma cadeia de tamanho $m = 1000$. Na Figura 2.1 estão apresentados os gráficos das trajetórias das cadeias de Markov simuladas para X e para Y .

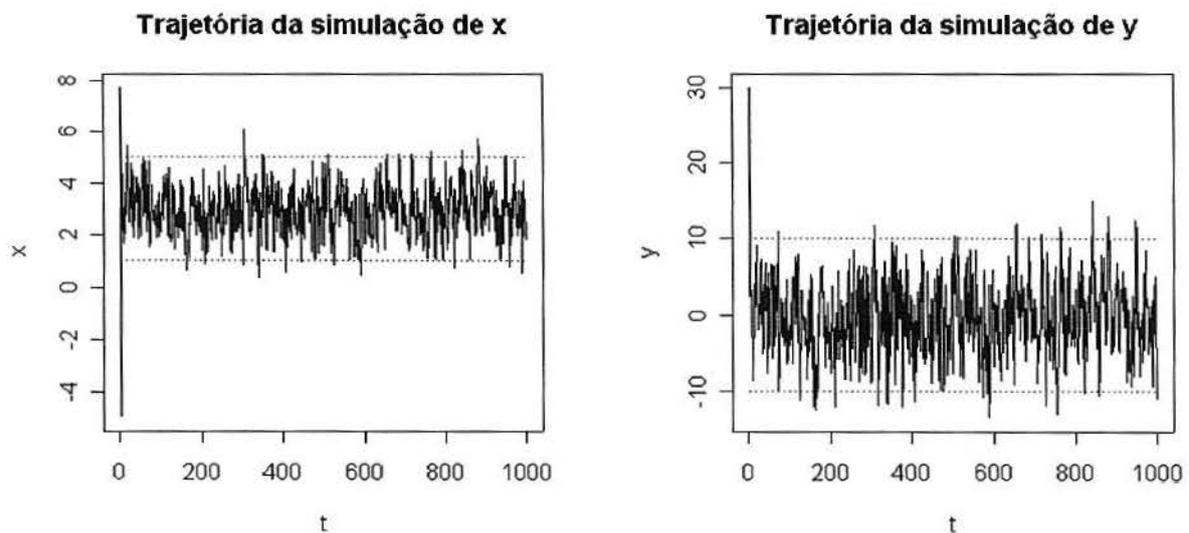


Figura 2.1: Trajetórias das cadeias de Markov de tamanho $m=1000$ geradas

As linhas horizontais pontilhadas presentes nos gráficos representam a média mais ou menos 2 vezes o desvio padrão respectivo de cada variável. Ou seja, espera-se que aproximadamente 95,45% dos pontos da cadeia esteja dentro destas duas linhas, com alguns pontos fora delas. Observa-se que, a partir da iteração $t=200$, a influência dos valores iniciais aparenta ter sido completamente eliminada e a convergência das cadeias aparenta ter sido atingida. Assim, consideraremos que o período de “burn-in” foi de $b=200$ iterações.

É notável que a análise da convergência considerando os desvios só pôde ser feita neste exemplo porque se conhece a distribuição marginal das variáveis, o que em problemas reais não ocorrerá. Em situações reais, os gráficos da trajetória da cadeia serão observados e se procurará a partir de qual iteração o comportamento dos dados aparenta ser similar, ou seja, a partir de qual iteração passa a existir um padrão na cadeia.

Para efeitos comparativos, aumentou-se o tamanho da cadeia para $m=5000$, e traçaram-se os gráficos das trajetórias das cadeias de Markov, mostrados na Figura 2.2.

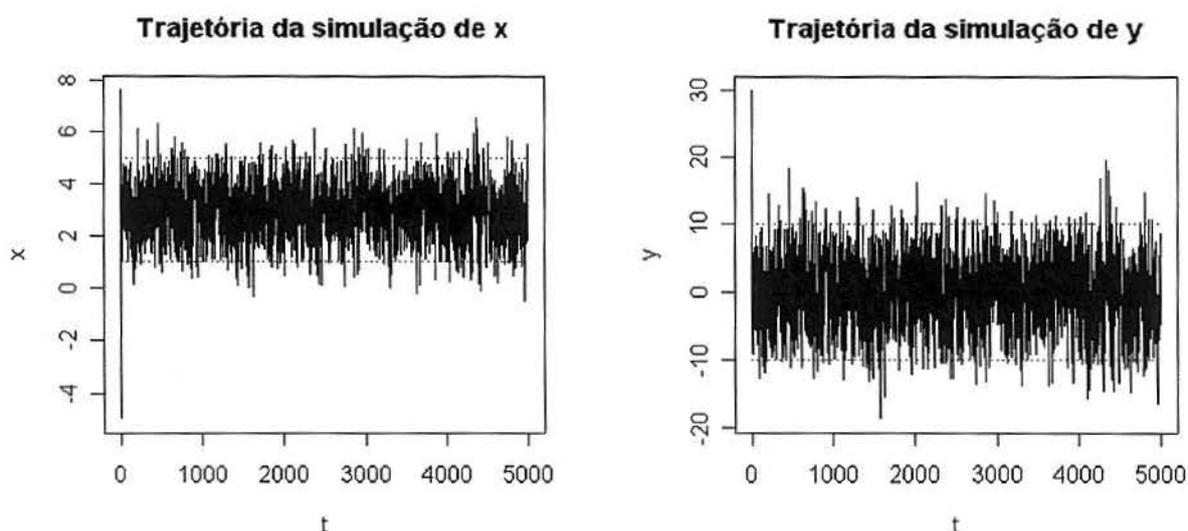


Figura 2.2: Trajetórias das cadeias de Markov de tamanho $m=5000$ geradas

Analisando-se as duas novas cadeias, decidiu-se considerar que o período de “burn-in” foi de $b=500$ iterações. Nas duas simulações, com as cadeias de tamanho 1000 e 5000, desconsideraram-se os valores gerados no “burn-in” e, com os valores restantes, calcularam-se a média de X , a variância de X , a média de Y , a variância de Y e o coeficiente de correlação amostral entre as duas variáveis, através dos métodos descritos na Subseção 2.3.3. Os resultados obtidos estão mostrados na Tabela 2.1.

Estatística	Parâmetro	$m=1000$	$m=5000$
		$b=200$	$b=500$
Média de X	3	2.916066	2.963873
Variância de X	1	0.9083664	0.9953891
Média de Y	0	-0.3334430	-0.204162
Variância de Y	25	23.72509	24.33253
Correlação	0,8	0.7804972	0.8040394

Tabela 2.1: Resultados obtidos nas duas simulações

Claramente, percebe-se que se utilizando um tamanho maior para a cadeia de Markov, obtiveram-se resultados mais próximos aos parâmetros das distribuições de X e Y .

Como mencionado na Subseção 2.3.3, pode-se estimar as densidades marginais das variáveis simuladas através da equação (2.8). Aplicando-se esta equação para os dados simulados nas duas etapas, produzimos os gráficos da Figura 2.3 e da Figura 2.4.

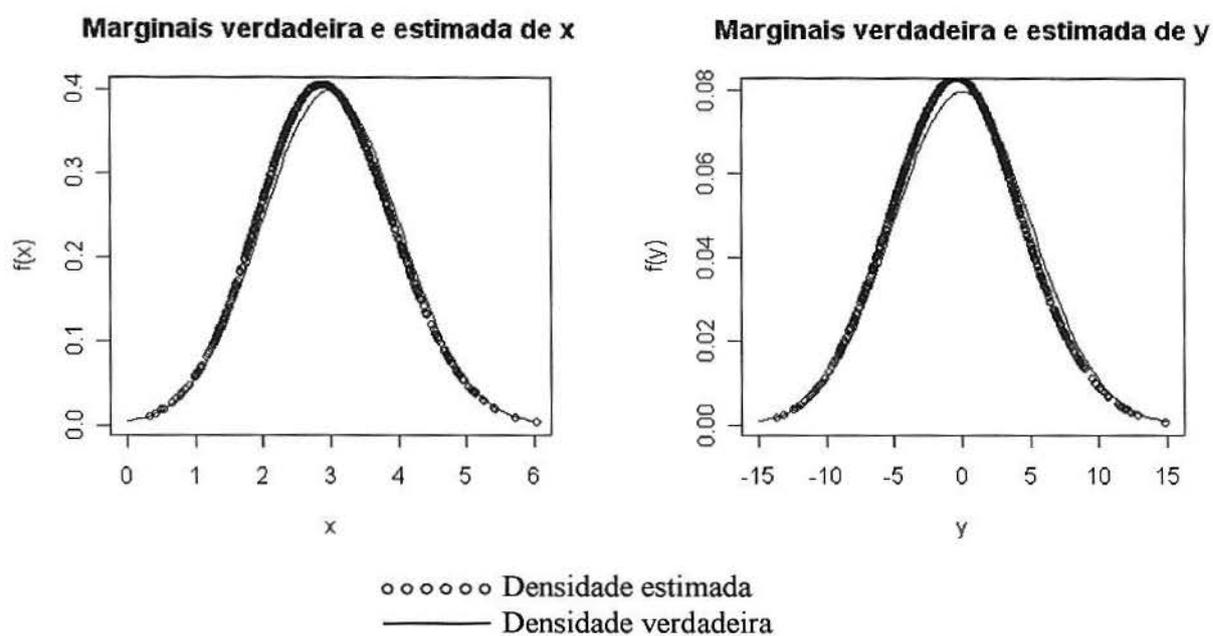


Figura 2.3: Densidades marginais estimadas para a cadeia com $m=1000$ e $b=200$

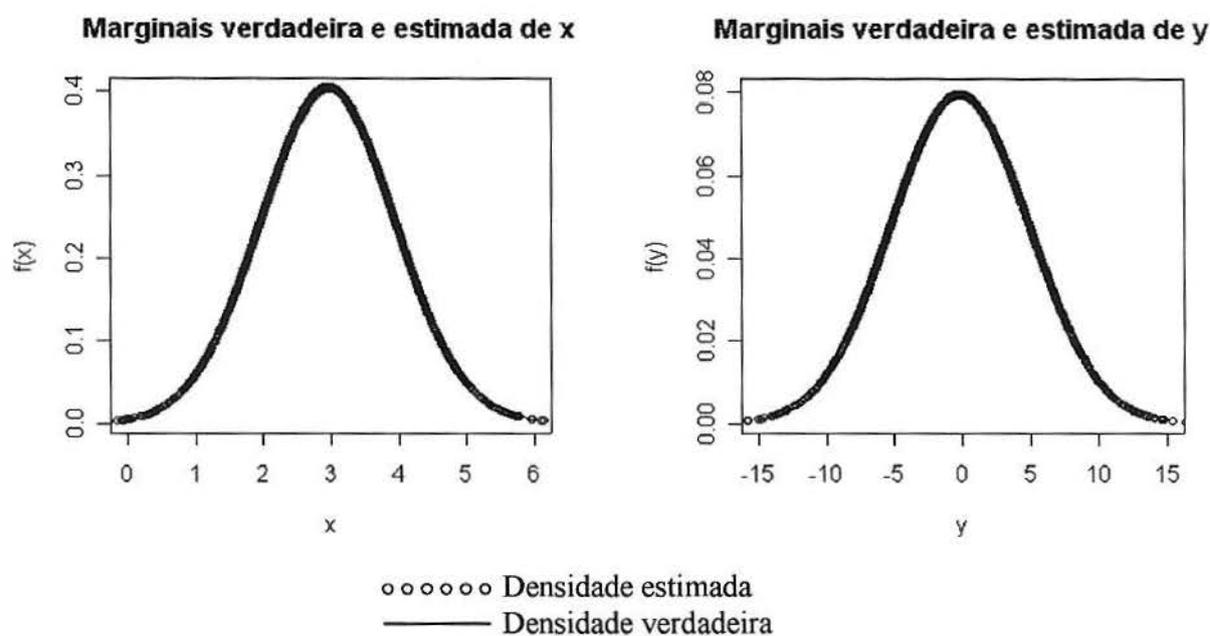


Figura 2.4: Densidades marginais estimadas para a cadeia com $m=5000$ e $b=500$

Novamente, parece evidente que a densidade marginal foi estimada com menor erro quando o tamanho da cadeia era maior.

3. ABORDAGEM CLÁSSICA PARA MODELOS MISTOS NORMAIS

Para muitas aplicações estatísticas, o uso do modelo linear geral (GLM) é uma metodologia muito útil. Este modelo supõe que os termos de erro sejam normais, independentes e com variância constante. Dois modelos conhecidos que podem ser escritos desta forma são a análise de variância e a análise de regressão. Entretanto, há casos em que a suposição de erros independentes é muito restritiva e não pode ser considerada como válida. Um exemplo é a análise de dados longitudinais, onde tomam-se observações em uma mesma unidade de observação ao longo do tempo e, por este fato, a suposição de independência das observações é comprometida. Assim, os modelos mistos permitem analisar estes casos, já que se podem assumir erros correlacionados e com variâncias heterogêneas. Além disto, é possível modelar-se efeitos fixos e aleatórios, enquanto o modelo linear geral somente modela efeitos fixos. Assim como o GLM, o modelo misto também pode ser utilizado em casos de dados não balanceados. Entretanto, somente com o maior desenvolvimento de softwares como SAS e R, a metodologia de modelos mistos se tornou mais acessível. Os modelos mistos normais, que serão abordados neste trabalho, são aqueles em que se assume distribuição normal para os erros.

Henderson (1984) aponta como maior vantagem do modelo misto a unificação da análise de várias técnicas estatísticas, permitindo que se utilize os mesmos métodos de estimação, testes de hipótese e predição para muitos problemas práticos.

Neste capítulo se apresentará na Seção 3.1 a formulação do modelo misto, seguido dos métodos de estimação dos efeitos fixos, aleatórios e parâmetros de variâncias e covariâncias, na Seção 3.2 e das técnicas de inferência na Seção 3.3. Por fim, exemplos serão apresentados na Seção 3.4, solucionados através do software SAS.

3.1. Formulação

Em notação matricial, pode-se expressar o modelo misto como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \quad (3.1)$$

onde

\mathbf{y} é um vetor $n \times 1$ de dados observados;

\mathbf{X} é uma matriz $n \times p$ conhecida;

β é um vetor $p \times 1$ de efeitos fixos geralmente desconhecidos;

Z é uma matriz $n \times q$ conhecida;

u é um vetor $q \times 1$ de efeitos aleatórios geralmente desconhecidos e

ε é um vetor $n \times 1$ de erros aleatórios não-observáveis.

As suposições do modelo são: $u \sim N(0, D)$, $\varepsilon \sim N(0, R)$ e $Cov(u, \varepsilon') = 0$, onde D e R são matrizes de variâncias e covariâncias de dimensões $q \times q$ e $n \times n$, respectivamente. Nota-se, por (3.1), que o modelo linear geral é um caso especial de modelo misto onde Z é nula e $R = I_n \sigma^2$, sendo I_n a matriz identidade $n \times n$.

Pela formulação do modelo em (3.1) e as suposições citadas acima, temos que

$$E(\mathbf{y}) = \mathbf{X}\beta,$$

$$Var(\mathbf{y}) = \mathbf{ZDZ}' + \mathbf{R} = \mathbf{V},$$

$$\mathbf{y} \sim N(\mathbf{X}\beta, \mathbf{V}).$$

Na análise de modelos mistos, pode-se supor um certo padrão para as matrizes de variâncias e covariâncias D e R . Isto é feito modelando-se uma estrutura de covariância para elas. Existem vários tipos de estruturas de covariância que podem ser modeladas, tais como:

1. *Componentes de Variância* (VC – “Variance Components”): permite modelar variâncias diferentes para grupos de efeitos ou observações e supõe covariâncias nulas;

2. *Autoregressiva de 1ª ordem* (AR(1) – “Autoregressive (1)”): supõe variâncias iguais, mas covariâncias que diminuem à medida que os elementos se afastam da diagonal principal da matriz;

3. *Não-estruturada* (UN – “Unstructured”): é o caso mais geral possível de estrutura de covariância, permitindo variâncias e covariâncias diferentes.

A seguir, as três estruturas estão apresentadas matricialmente.

$$VC = \begin{bmatrix} \sigma_A^2 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & \ddots & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & \sigma_A^2 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \sigma_B^2 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & \ddots & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & \sigma_B^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}; AR(1) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots \\ \rho & 1 & \rho & \rho^2 & \dots \\ \rho^2 & \rho & 1 & \rho & \dots \\ \rho^3 & \rho^2 & \rho & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix};$$

$$UN = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} & \dots \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} & \dots \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} & \dots \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

Pode-se especificar o modelo misto em (3.1) da forma de um modelo hierárquico. Modelos hierárquicos, em linhas gerais, são modelos especificados em estágios, onde cada estágio possui um submodelo. São muito úteis, assim como os modelos mistos, para dados que se apresentam agrupados em níveis hierárquicos onde cada estágio modelará as relações entre as variáveis naquele nível. Entretanto, apesar da formulação dos dois ser diferente, eles são equivalentes. No caso dos modelos mistos, tem-se que:

$$\text{Estágio 1: } \mathbf{y} | \mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}; \mathbf{R}) \quad (3.2)$$

$$\text{Estágio 2: } \mathbf{u} \sim N(\mathbf{0}; \mathbf{D}). \quad (3.3)$$

Esta hierarquia será útil para se compreender algumas propriedades do modelo misto bem como a aplicação das técnicas Bayesianas a este modelo, assunto que será abordado no Capítulo 4.

3.2. Estimação

Na inferência clássica, os métodos de estimação usualmente baseiam-se na maximização de uma função de verossimilhança, segundo a idéia de que a estimativa mais adequada de um parâmetro seria aquele valor mais provável de ter gerado o particular conjunto de dados observados.

Em modelos em que se supõem observações independentes, a verossimilhança é simplesmente o produto da função densidade de probabilidade para cada observação. Como no modelo misto as observações nem sempre são independentes, a verossimilhança precisa ser

baseada em uma distribuição multivariada. Assim, para o modelo misto normal, a função de verossimilhança é baseada em uma distribuição normal multivariada, como segue:

$$L(\boldsymbol{\beta}, \mathbf{V} | \mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\mathbf{V}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\}, \quad (3.4)$$

onde $|\mathbf{V}|$ é o determinante da matriz \mathbf{V} .

Para a maximização da função de verossimilhança, é mais útil se tomar o seu logaritmo natural, como abaixo:

$$\ln L(\boldsymbol{\beta}, \mathbf{V} | \mathbf{y}) = -\frac{1}{2}n \ln(2\pi) - \frac{1}{2}[\ln|\mathbf{V}| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]. \quad (3.5)$$

Há interesse em se fazer inferências sobre três questões quando se trabalha com modelos mistos: os parâmetros dos efeitos fixos, os parâmetros dos efeitos aleatórios e os parâmetros de variâncias e covariâncias do modelo.

3.2.1. Estimação dos efeitos fixos e aleatórios

Há três maneiras de se obter uma solução para os efeitos fixos: a primeira delas consiste em obter o valor de $\boldsymbol{\beta}$ que maximiza (3.5), a segunda é obter uma solução através de mínimos quadrados generalizados e a terceira é solucionar as equações do modelo misto. Será mostrado a seguir que todas as três maneiras obtêm o mesmo estimador para $\boldsymbol{\beta}$.

Para os efeitos aleatórios, pode-se estimá-los através das equações do modelo misto.

- Método da Máxima Verossimilhança

Para maximizar a função de verossimilhança, deve-se diferenciar (3.5) em relação a $\boldsymbol{\beta}$ considerando-se \mathbf{V} conhecida e igualar o resultado a zero, obtendo uma solução que depende dos parâmetros de variâncias e covariâncias do modelo:

$$\mathbf{X}' \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \mathbf{0}.$$

Fazendo algumas manipulações na expressão acima, obtém-se

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y}, \quad (3.6)$$

cuja esperança e variância são:

$$E(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}E(\mathbf{y}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta} \quad (3.7)$$

$$Var(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Var(\mathbf{y})\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \dots = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \quad (3.8)$$

Percebe-se que o estimador em (3.6) é não-viciado, considerando-se que \mathbf{V} é conhecida. Posteriormente, será discutido o caso em que \mathbf{V} não é conhecida.

- Método dos Mínimos Quadrados Generalizados

O método de mínimos quadrados generalizados, no caso dos modelos mistos, consiste em minimizar o produto dos resíduos completos, $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, ponderados pelo inverso da matriz de variâncias, \mathbf{V}^{-1} . O termo resíduos completos é devido a diferenciação entre esses e os resíduos ordinários, $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}$. Logo, um estimador para $\boldsymbol{\beta}$ é obtido minimizando-se a seguinte expressão em relação a $\boldsymbol{\beta}$:

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (3.9)$$

Sem fazer-se nenhum cálculo, pode-se notar que o estimador obtido por mínimos quadrados generalizados é o mesmo obtido por máxima verossimilhança, já que maximizar (3.5) em relação a $\boldsymbol{\beta}$ nada mais é do que minimizar (3.9) também em relação a $\boldsymbol{\beta}$.

- Equações do Modelo Misto

As equações do modelo misto são devidas a Henderson (1984) e são úteis para obterem-se tanto soluções para os parâmetros dos efeitos fixos quanto para os parâmetros dos efeitos aleatórios. São obtidas através da maximização da função de verossimilhança em termos de $\boldsymbol{\beta}$, \mathbf{u} , \mathbf{D} e \mathbf{R} , construída através da idéia de ver o modelo misto como um modelo hierárquico.

Assim, usando (3.2) e (3.3), temos:

$$L(\boldsymbol{\beta}, \mathbf{u}, \mathbf{D}, \mathbf{R} | \mathbf{y}) = f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y} | \mathbf{u})f(\mathbf{u}) = L(\boldsymbol{\beta}, \mathbf{R} | \mathbf{u}, \mathbf{y}) \times L(\mathbf{D} | \mathbf{u}) \quad (3.10)$$

$$= \frac{1}{(2\pi)^n |\mathbf{R}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} [\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})]' \mathbf{R}^{-1} [\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})]\right\} \times \frac{1}{(2\pi)^q |\mathbf{D}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \mathbf{u}' \mathbf{D}^{-1} \mathbf{u}\right\}.$$

Como no caso da maximização de (3.5), é mais útil se tomar o logaritmo de (3.10), como segue:

$$l = \ln L(\boldsymbol{\beta}, \mathbf{u}, \mathbf{D}, \mathbf{R} | \mathbf{y}) = \frac{n+q}{2} \ln(2\pi) - \frac{1}{2} \left\{ \ln|\mathbf{R}| + [\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})]' \mathbf{R}^{-1} [\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})] + \ln|\mathbf{D}| + \mathbf{u}' \mathbf{D}^{-1} \mathbf{u} \right\}. \quad (3.11)$$

Derivando-se (3.11) em relação a $\boldsymbol{\beta}$ e a \mathbf{u} , considerando-se \mathbf{D} e \mathbf{R} conhecidas, e substituindo $\boldsymbol{\beta}$ por $\hat{\boldsymbol{\beta}}$ e \mathbf{u} por $\hat{\mathbf{u}}$, tem-se que:

$$\begin{aligned} \begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \mathbf{u}} \end{bmatrix} &= \begin{bmatrix} -\mathbf{X}' \mathbf{R}^{-1} \mathbf{y} + \mathbf{X}' \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{X}' \mathbf{R}^{-1} \mathbf{Z} \hat{\mathbf{u}} \\ -\mathbf{Z}' \mathbf{R}^{-1} \mathbf{y} + \mathbf{Z}' \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} \hat{\mathbf{u}} + \mathbf{D}^{-1} \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ & \begin{bmatrix} \mathbf{X}' \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{X}' \mathbf{R}^{-1} \mathbf{Z} \hat{\mathbf{u}} \\ \mathbf{Z}' \mathbf{R}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} \hat{\mathbf{u}} + \mathbf{D}^{-1} \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}' \mathbf{R}^{-1} \mathbf{y} \end{bmatrix} \\ & \begin{bmatrix} \mathbf{X}' \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}' \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}' \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}' \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}. \end{aligned}$$

Com algumas manipulações, e lembrando que $\mathbf{V} = \mathbf{Z}\mathbf{D}\mathbf{Z}' + \mathbf{R}$, obtém-se os estimadores:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{y} \quad \text{e} \quad (3.12)$$

$$\hat{\mathbf{u}} = \mathbf{D}\mathbf{Z}' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}). \quad (3.13)$$

Percebe-se que o estimador para os parâmetros dos efeitos fixos em (3.12) é idêntico àquele em (3.6), e sua variância, por conseqüência, idêntica àquela em (3.8).

A variância de $\hat{\mathbf{u}}$ é

$$\text{Var}(\hat{\mathbf{u}}) = \mathbf{D}\mathbf{Z}' \mathbf{V}^{-1} \mathbf{Z}\mathbf{D} - \mathbf{D}\mathbf{Z}' \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} \mathbf{Z}\mathbf{D}. \quad (3.14)$$

Note, por (3.12) e (3.13), que os estimadores de $\boldsymbol{\beta}$ e \mathbf{u} supõem que a matriz \mathbf{V} seja conhecida. Em problemas práticos, raramente a matriz \mathbf{V} será conhecida. Nestas situações, $\boldsymbol{\beta}$ e \mathbf{u} são estimados substituindo \mathbf{V} por $\hat{\mathbf{V}}$ (uma estimativa para \mathbf{V}) em, respectivamente,

(3.12) e (3.13). Para as estimativas dos parâmetros dos efeitos fixos e aleatórios, este procedimento não acarreta problemas, porém, ocasiona subestimativas para as variâncias dos estimadores (Brown e Prescott, 1999).

Se \mathbf{D} e \mathbf{R} são conhecidas, $\hat{\boldsymbol{\beta}}$ é o melhor estimador linear não-viciado (BLUE – “Best Linear Unbiased Estimator”) de $\boldsymbol{\beta}$ e $\hat{\mathbf{u}}$ é o melhor preditor linear não-viciado (BLUP – “Best Linear Unbiased Predictor”) de \mathbf{u} , onde “melhor” significa mínimo erro quadrático médio. De qualquer modo, \mathbf{D} e \mathbf{R} são usualmente desconhecidas e estimadas utilizando um dos métodos de estimação de parâmetros de variâncias e covariâncias a serem vistos a seguir. Neste caso, as abreviaturas BLUE e BLUP não são adequadas, mas a palavra empírico é freqüentemente acrescentada para indicar tal aproximação. A abreviação apropriada se torna então EBLUE e EBLUP (SAS Institute Inc., 2002).

3.2.2. *Estimação dos parâmetros de variâncias e covariâncias*

Existem disponíveis vários métodos para estimar os parâmetros de variâncias e covariâncias do modelo misto. Os três métodos mais conhecidos são: o método da ANOVA, ou dos momentos; o método da máxima verossimilhança (ML) e o método da máxima verossimilhança restrita, ou residual (REML).

- Método da ANOVA

Consiste basicamente em igualar os quadrados médios observados aos seus valores esperados, em uma tabela de análise de variância. Note que os valores esperados são combinações lineares dos componentes de variância. Assim, obtêm-se equações lineares que envolvem os componentes de variância, e as soluções são tomadas como seus estimadores. Os estimadores produzidos pelo método da ANOVA são sempre não-viciados e de variância mínima. Porém, podem produzir estimativas negativas de componentes de variância, o que não é coerente.

- Método da Máxima Verossimilhança (ML)

Maximiza a função de verossimilhança (3.4) em relação aos parâmetros de variâncias e covariâncias considerando os parâmetros dos efeitos fixos como constantes. Após obter as estimativas para os parâmetros de variâncias e covariâncias, os efeitos fixos são então estimados maximizando-se a mesma verossimilhança, porém considerando os parâmetros de

variâncias e covariâncias como constantes. Este método produz estimativas viciadas para os parâmetros de variâncias e covariâncias, mais especificamente, subestima os parâmetros. Este vício é maior quanto menores os graus de liberdade utilizados para se estimar os parâmetros de variâncias e covariâncias (Brown e Prescott, 1999).

- Método da Máxima Verossimilhança Residual ou Restrita (REML)

Utiliza uma função de verossimilhança que não envolve β , ou seja, é definida somente em termos dos parâmetros de variâncias e covariâncias. O primeiro passo é obter uma função de verossimilhança a partir dos resíduos completos estimados, $\mathbf{y} - \mathbf{X}\hat{\beta}$, onde $\hat{\beta}$ está definido em (3.6). Pode ser mostrado que $\mathbf{y} - \mathbf{X}\hat{\beta}$ e $\hat{\beta}$ são independentes e que $\hat{\beta}$ segue uma distribuição normal multivariada com média e variância dadas por (3.7) e (3.8). Assim, a verossimilhança REML é escrita como:

$$L(\mathbf{V} | \mathbf{y} - \mathbf{X}\hat{\beta}) = \frac{L(\beta, \mathbf{V} | \mathbf{y})}{L(\beta | \hat{\beta}, \mathbf{V})},$$

onde $L(\beta, \mathbf{V} | \mathbf{y})$ foi definida em (3.4) e

$$L(\beta | \hat{\beta}, \mathbf{V}) = \frac{1}{(2\pi)^{\frac{p}{2}} \left| (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \right|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\hat{\beta} - \beta)' \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} (\hat{\beta} - \beta) \right\}.$$

Portanto,

$$L(\mathbf{V} | \mathbf{y} - \mathbf{X}\hat{\beta}) = \frac{1}{(2\pi)^{\frac{n-p}{2}} \left| (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \right|^{\frac{1}{2}} |\mathbf{V}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}) \right\}$$

e tomando-se seu logaritmo natural:

$$\ln L(\mathbf{V} | \mathbf{y} - \mathbf{X}\hat{\beta}) = -\frac{1}{2} \left\{ (n-p) \ln(2\pi) - \ln \left| (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \right| + \ln |\mathbf{V}| + (\mathbf{y} - \mathbf{X}\hat{\beta})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\hat{\beta}) \right\}. \quad (3.15)$$

Apesar de que $\hat{\beta}$ aparece em (3.15), a função de verossimilhança só depende dos parâmetros de variâncias e covariâncias, já que $\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Deve-se notar que a diferença entre o logaritmo da verossimilhança do método ML e o logaritmo da verossimilhança do método REML está no termo adicional que esta última apresenta,

$\ln|(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}|$, que é o logaritmo do determinante da $\text{Var}(\hat{\boldsymbol{\beta}})$. Por incorporar este termo, as estimativas produzidas pelo REML são não-viciadas.

As estimativas para os parâmetros dos efeitos fixos não costumam diferir muito se as estimativas de variância forem produzidas pelo REML ou pelo ML. Entretanto, seus erros padrões serão sempre subestimados se os parâmetros de variâncias e covariâncias forem os estimados pelo ML. Assim, o método REML é preferível em relação ao ML (Brown e Prescott, 1999). Em casos balanceados, a solução obtida através do método REML é idêntica à obtida pelo método da ANOVA (Searle et al., 1992).

Tanto pelo ML, quanto pelo REML, não há uma equação que especifique as soluções para os parâmetros de variâncias e covariâncias, como existia para os efeitos fixos e aleatórios. Assim, necessita-se de um método iterativo para maximizar as duas funções de verossimilhança, como o algoritmo de Newton-Raphson, disponível no SAS.

3.3. Intervalos de Confiança e Teste de Hipóteses

Para testarem-se hipóteses sobre os efeitos fixos, deve-se construir um vetor de contrastes \mathbf{L} . Por exemplo, se há três efeitos fixos, μ , β_1 e β_2 , para testar-se a diferença $\beta_1 - \beta_2$, toma-se $\mathbf{L} = [0 \ 1 \ -1]$. E a hipótese nula do teste será $H_0 : \mathbf{L}\boldsymbol{\beta} = 0$.

Utiliza-se a estatística de teste

$$t = \frac{\mathbf{L}\hat{\boldsymbol{\beta}}}{\sqrt{\mathbf{L}\text{var}(\hat{\boldsymbol{\beta}})\mathbf{L}'}} \quad (3.16)$$

onde $\text{var}(\hat{\boldsymbol{\beta}})$ é equivalente à dada pela expressão (3.8) trocando-se \mathbf{V} por $\hat{\mathbf{V}}$. Em geral, a estatística de teste acima segue uma distribuição t aproximada, e seus graus de liberdade devem ser estimados. O método mais conhecido para estimar estes graus de liberdade é devido a Satterthwaite (1946) e está disponível no SAS. Demonstra-se que, no caso de dados balanceados e alguns casos especiais de dados desbalanceados (ver SAS Institute Inc., 2002), a estatística de teste dada em (3.16) segue uma distribuição t exata.

Um intervalo com $100(1 - \alpha)\%$ de confiança pode ser construído facilmente a partir de (3.16):

$$\mathbf{L}\hat{\boldsymbol{\beta}} \pm t_{\hat{\nu}, \alpha/2} \sqrt{\mathbf{L} \text{var}(\hat{\boldsymbol{\beta}}) \mathbf{L}'},$$

onde $\hat{\nu}$ são os graus de liberdade estimados.

No caso dos efeitos aleatórios, as expressões para o teste de hipóteses e o intervalo de confiança são análogas. Elas são dadas por

$$H_0 : \mathbf{L}\mathbf{u} = 0,$$

$$t = \frac{\mathbf{L}\mathbf{u}}{\sqrt{\mathbf{L} \text{var}(\hat{\mathbf{u}}) \mathbf{L}'}}$$

e

$$\mathbf{L}\mathbf{u} \pm t_{\hat{\nu}, \alpha/2} \sqrt{\mathbf{L} \text{var}(\hat{\mathbf{u}}) \mathbf{L}'}$$

Para fazer inferências sobre os parâmetros de variâncias e covariâncias, uma estatística comumente usada é a Z de Wald, que é calculada dividindo-se a estimativa do parâmetro por seu erro padrão assintótico. O erro padrão assintótico é calculado a partir da inversa da matriz de derivadas segundas da verossimilhança em relação a cada um dos parâmetros. Este teste é válido para amostras grandes, mas pode não ser confiável para conjuntos de dados pequenos e para parâmetros de variância, pois se sabe que a distribuição do estimador destes parâmetros é assimétrica ou truncada (SAS Institute Inc., 2002).

Uma melhor alternativa é fazer o teste para os parâmetros de variância e covariância através da comparação das verossimilhanças dos modelos incluindo o componente de variância (L_1) e excluindo-o (L_2). Um resultado muito utilizado em outras técnicas estatísticas, como a regressão logística, por exemplo, é que, sob a hipótese nula de que o termo retirado do modelo não é significativo, a diferença dos logaritmos das verossimilhanças tem distribuição $1/2\chi_{1gl}^2$. Ou seja,

$$2(\ln(L_1) - \ln(L_2)) \sim \chi_{1gl}^2.$$

Este teste é conhecido como Teste da Razão de Verossimilhanças.

3.4. Exemplos

3.4.1. Modelo de Componentes de Variância

O objetivo principal deste exemplo é mostrar a análise de um modelo com poucos parâmetros através da formulação em modelo misto. No Capítulo 4, este exemplo será revisitado no contexto Bayesiano.

O modelo de componentes de variância é definido como:

$$y_{ij} = \theta_i + \varepsilon_{ij}, \quad i = 1, \dots, K, \quad j = 1, \dots, J$$

onde $\theta_i \sim N(\mu_\theta, \sigma_\theta^2)$, $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$ e, portanto, $y_{ij} | \theta_i \sim N(\theta_i, \sigma_\varepsilon^2)$.

Este modelo pode ser visto em sua estrutura hierárquica, onde:

$$\text{Estágio 1: } y_{ij} | \theta_i \sim N(\theta_i, \sigma_\varepsilon^2) \quad (3.17)$$

$$\text{Estágio 2: } \theta_i \sim N(\mu_\theta, \sigma_\theta^2). \quad (3.18)$$

Um conjunto de dados será simulado segundo este modelo para exemplificar a sua análise. Definiu-se os seguintes valores para os parâmetros e número de observações: $K = 5$, $J = 10$, $\mu_\theta = 10$, $\sigma_\theta^2 = 15$ e $\sigma_\varepsilon^2 = 3$. Esta etapa foi feita através do software R e a programação utilizada está apresentada no Apêndice B.

O primeiro passo foi simular valores para $\theta_i \sim N(10, 15)$, $i = 1, \dots, 5$. Gerou-se então, os valores: $\theta_1 = 12,961318$, $\theta_2 = 6,371229$, $\theta_3 = 9,998053$, $\theta_4 = 6,660848$ e $\theta_5 = 13,973707$. A partir destes valores, geraram-se os dados para $y_{ij} \sim N(\theta_i, 3)$, $i = 1, \dots, 5$ e $j = 1, \dots, 10$, obtendo-se os dados apresentados na Tabela 3.1:

$i \setminus j$	1	2	3	4	5
1	12,023331	9,124187	8,927173	7,444252	14,56684
2	9,723175	10,592689	9,677705	7,037576	12,943671
3	12,416579	5,729597	12,697276	5,023533	14,632859
4	14,262978	9,817653	7,761687	6,907652	13,891107
5	12,79859	5,766354	6,883161	5,302723	14,307992
6	15,054712	5,547308	8,460689	6,737171	16,198396
7	12,34766	7,238691	9,103424	9,233624	16,215717
8	16,345757	7,582859	7,813017	6,349764	19,96081
9	11,439246	4,087789	9,584939	7,097709	10,881694
10	12,39707	7,940003	10,410913	6,555402	14,01892

Tabela 3.1: Dados simulados do modelo de Componentes de Variância

O modelo de componentes de variância pode também ser expresso da forma de modelo misto dada em (3.1), onde $\mathbf{R} = \mathbf{I}_n \sigma_\varepsilon^2$, sendo \mathbf{I}_n a matriz identidade $n \times n$ e $n = KJ$. Os vetores \mathbf{y} , $\boldsymbol{\beta}$, \mathbf{u} e $\boldsymbol{\varepsilon}$, e as matrizes \mathbf{X} e \mathbf{Z} para este exemplo estão a seguir especificados.

$$\mathbf{y} = \begin{bmatrix} y_{1,1} \\ y_{1,2} \\ \vdots \\ y_{1,10} \\ \vdots \\ y_{5,1} \\ y_{5,2} \\ \vdots \\ y_{5,10} \end{bmatrix} = \begin{bmatrix} 12,023331 \\ 9,723175 \\ \vdots \\ 12,39707 \\ \vdots \\ 14,56684 \\ 12,943671 \\ \vdots \\ 14,01892 \end{bmatrix}; \mathbf{X} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}; \boldsymbol{\beta} = [\mu_\theta]; \mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ \dots & \vdots & \vdots \\ 0 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 1 \end{bmatrix}; \mathbf{u} = \begin{bmatrix} \theta_1 - \mu_\theta \\ \theta_2 - \mu_\theta \\ \vdots \\ \theta_3 - \mu_\theta \\ \theta_4 - \mu_\theta \\ \theta_5 - \mu_\theta \end{bmatrix} \text{ e } \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{1,1} \\ \varepsilon_{1,2} \\ \vdots \\ \varepsilon_{1,10} \\ \vdots \\ \varepsilon_{5,1} \\ \varepsilon_{5,2} \\ \vdots \\ \varepsilon_{5,10} \end{bmatrix}.$$

Para este exemplo, tem-se que $n = KJ = 50$, $p = 1$ e $q = 5$. Deve-se notar que os efeitos aleatórios contidos no vetor \mathbf{u} são $\theta_i - \mu_\theta$ e não simplesmente θ_i para que se possa assumir $\mathbf{u} \sim N(\mathbf{0}, \mathbf{D})$, onde, neste caso, $\mathbf{D} = \sigma_\theta^2 \mathbf{I}_q$.

Analisando-se este exemplo através do Proc MIXED do SAS e utilizando-se os três métodos de estimativas de variâncias descritos anteriormente, obtiveram-se os resultados apresentados na Tabela 3.2.

Estimativa	Valor Verdadeiro	REML	ML	ANOVA
$\hat{\mu}_\theta$	10	10,1773	10,1773	10,1773
$\hat{\theta}_1 = \hat{u}_1 + \mu_\theta$	12,961318	12,8032	12,7838	12,8032
$\hat{\theta}_2 = \hat{u}_2 + \mu_\theta$	6,371229	7,4242	7,4446	7,4242
$\hat{\theta}_3 = \hat{u}_3 + \mu_\theta$	9,998053	9,1621	9,1696	9,1621
$\hat{\theta}_4 = \hat{u}_4 + \mu_\theta$	6,660848	6,8669	6,8914	6,8669
$\hat{\theta}_5 = \hat{u}_5 + \mu_\theta$	13,973707	14,6301	14,5971	14,6301
$\hat{\sigma}_\theta^2$	15	11,9153	9,4617	11,9153
$\hat{\sigma}_\varepsilon^2$	3	3,5259	3,5259	3,5259

Tabela 3.2: Resultados do Proc MIXED do SAS para o exemplo do Modelo de Componentes de Variância

Analisando os resultados da Tabela 3.2, nota-se que os obtidos através do método REML e método da ANOVA foram iguais, como era esperado já que os dados deste problema são balanceados. As estimativas foram próximas aos valores verdadeiros. A exceção foi o parâmetro σ_θ^2 , que teve uma diferença mais discrepante entre o valor definido e a estimativa produzida pelos três métodos. Além disso, nota-se que o método ML produziu um valor mais afastado ao valor especificado para σ_θ^2 que o REML e o método da ANOVA.

3.4.2. Modelo para Dados Longitudinais – Banco de dados *Weights*

Dados longitudinais são observações feitas sequencialmente ao longo do tempo nos mesmos sujeitos, indivíduos ou objetos, ou seja, nas mesmas unidades experimentais. O modelo misto tem características bastante úteis para a análise de dados longitudinais, pois não tem restrições caso os dados sejam não balanceados (medidas obtidas em tempos diferentes nas unidades experimentais, observações faltantes, número de unidades experimentais diferentes por tratamento são casos de desbalanceamento) e permite que se modelem correlações entre as observações feitas em uma mesma unidade experimental, o que é bastante coerente.

A aplicação dos modelos mistos para a análise de dados longitudinais foi primeiramente apresentada por Laird e Ware (1982), que especificaram o seguinte modelo:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\varepsilon}_i, \quad (3.19)$$

onde

$\mathbf{y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,j}, \dots, y_{i,n_i})$ é o vetor das n_i observações feitas no sujeito i , $i = 1, \dots, n$;

\mathbf{X}_i é uma matriz $n_i \times p$ conhecida;

$\boldsymbol{\beta}$ é o vetor dos efeitos fixos, de dimensão $p \times 1$;

\mathbf{Z}_i é uma matriz $n_i \times q$ conhecida;

\mathbf{u}_i é o vetor dos efeitos aleatórios específicos do sujeito i , de dimensão $q \times 1$;

$\boldsymbol{\varepsilon}_i$ é o vetor de erros aleatórios não-observáveis, de dimensão $n_i \times 1$.

Para completar a especificação do modelo, supõe-se que $\mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D})$ e $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i)$. A matriz \mathbf{R}_i depende do índice i somente em sua dimensão, e não de qualquer outra maneira (Davidian e Giltinan, 1995). É através da escolha de estruturas de covariância para a matriz \mathbf{R}_i que as correlações entre as observações feitas na mesma unidade experimental são modeladas. Na abordagem clássica, usualmente adota-se a estrutura VC para a matriz \mathbf{D} .

O modelo em (3.19) pode ser especificado da forma de um modelo hierárquico e a hierarquia ajuda a entender as duas fontes de variabilidade existentes no problema envolvendo dados longitudinais. Isto pode ser visto através das expressões a seguir:

$$\text{Estágio 1: } \mathbf{y}_i | \mathbf{u}_i \sim N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i, \mathbf{R}_i)$$

$$\text{Estágio 2: } \mathbf{u}_i \sim N(\mathbf{0}, \mathbf{D}).$$

O estágio 1 modela a variabilidade dentro de sujeitos, ou seja, explica-se o porquê da variável resposta não ser constante para as diversas observações no mesmo sujeito. O estágio 2 modela a variabilidade entre sujeitos, ou seja, explica-se o porquê da variável resposta não assumir os mesmos valores para as medidas obtidas no mesmo tempo em sujeitos diferentes.

Para exemplificar a análise de dados longitudinais através do modelo misto, utilizou-se um banco de dados da biblioteca do software SAS chamado *Weights*. Estes dados foram obtidos a partir de um estudo onde 57 indivíduos receberam um de três programas de levantamento de peso (tratamentos) disponíveis. A força de cada indivíduo foi medida uma vez a cada dois dias durante duas semanas a contar do início do programa. Os sujeitos no

primeiro programa, denotado por CONT, não participaram do levantamento de peso. O segundo programa, denotado por RI, aumenta o número de repetições conforme o sujeito torna-se mais forte. Finalmente, o terceiro programa, denotado por WI, aumenta a quantidade de peso à medida que o sujeito torna-se mais forte. O objetivo desta análise é investigar o efeito que cada programa de levantamento de peso tem no aumento da força ao longo do tempo.

Neste estudo, as observações foram tomadas nos mesmos momentos em todas as unidades experimentais, e não há observações faltantes. Porém, há um número diferente de sujeitos para cada tratamento (20 no CONT, 16 no RI e 21 no WI). Por isso, é um conjunto de dados desbalanceado.

Uma das maneiras de se analisar este conjunto de dados é ajustar uma reta de regressão para cada sujeito e supor que os interceptos e coeficientes angulares de cada sujeito são variáveis aleatórias cuja esperança depende do programa de levantamento de peso que o indivíduo recebeu.

Para facilitar a compreensão do exemplo, antes de apresentar a formulação do modelo misto para este conjunto de dados, a especificação de um modelo hierárquico para o mesmo será apresentada.

Assim, pode-se especificar o modelo hierárquico proposto como

$$\text{Estágio 1: } y_{ijk} | \alpha_{suj.i}, \beta_{suj.i} \stackrel{iid}{\sim} N(\alpha_{suj.i} + \beta_{suj.i} \text{tempo}_j, \sigma_\varepsilon^2) \quad (3.20)$$

$$\text{Estágio 2: } \begin{cases} \alpha_{suj.i} \stackrel{iid}{\sim} N(\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \dots + \alpha_{(t-1)} t_{(t-1)}, \sigma_\alpha^2) \\ \beta_{suj.i} \stackrel{iid}{\sim} N(\beta_0 + \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_{(t-1)} t_{(t-1)}, \sigma_\beta^2) \end{cases} \quad (3.21)$$

onde $i = 1, \dots, n$, onde n é o número de unidades experimentais (sujeitos), no caso, $n = 57$;

$j = 1, \dots, n_i$, onde n_i é o número observações feitas no sujeito i , no caso, $n_i = g = 7, \forall i$;

$k = 1, \dots, t$, onde t é o número de tratamentos, no caso, $t = 3$, sendo que o programa CONT é identificado por $k = 1$, o programa RI, por $k = 2$ e o programa WI, por $k = 3$;

$t_l, l=1, \dots, (t-1)$ são variáveis indicadoras dos tratamentos, tal que

$$t_l = \begin{cases} 1, \text{ se } k = l \\ 0, \text{ caso contrário} \end{cases}$$

Percebe-se, por (3.20) e (3.21), que a covariável tratamento foi incluída no estágio 2 deste modelo. A inclusão é feita de maneira análoga às idéias da regressão múltipla de usar variáveis indicadoras para covariáveis qualitativas e as próprias covariáveis quando estas forem contínuas. Neste exemplo somente foi incluída a covariável tratamento. Outras covariáveis tais como idade, peso, etc., também podem ser incluídas no estágio 2 para explicar a variabilidade entre sujeitos.

Este mesmo modelo hierárquico pode ser escrito da forma de modelo misto apresentada em (3.19), onde $p = 6, q = 2$ e:

$$\mathbf{y}_i = \begin{bmatrix} y_{i1k} \\ y_{i2k} \\ \vdots \\ y_{in_k} \end{bmatrix}; \mathbf{u}_i = \begin{bmatrix} \alpha_{suj.i} - (\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \dots + \alpha_{(t-1)} t_{(t-1)}) \\ \beta_{suj.i} - (\beta_0 + \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_{(t-1)} t_{(t-1)}) \end{bmatrix}; \mathbf{Z}_i = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & n_i \end{bmatrix} \forall i;$$

$$\boldsymbol{\beta} = \begin{bmatrix} \alpha_0 \\ \beta_0 \\ \vdots \\ \alpha_{(t-1)} \\ \beta_{(t-1)} \end{bmatrix}; \mathbf{X}_i = [\mathbf{Z}_i | t_1 \mathbf{Z}_i | \dots | t_{(t-1)} \mathbf{Z}_i]; \boldsymbol{\varepsilon}_i = \begin{bmatrix} y_{i1k} \\ y_{i2k} \\ \vdots \\ y_{in_k} \end{bmatrix}; \mathbf{D} = \begin{bmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}; \mathbf{R}_i = \sigma_\varepsilon^2 \mathbf{I}_{n_i}.$$

Note que se está adotando estruturas VC para as matrizes \mathbf{R}_i e \mathbf{D} . Como, para este conjunto de dados, $n_i = g = 7, \forall i$, as matrizes \mathbf{R}_i serão iguais a $\mathbf{R} = \sigma_\varepsilon^2 \mathbf{I}_7, \forall i$.

Os interceptos e coeficientes angulares para cada tratamento são:

$$\alpha_{trat1} = \alpha_0 + \alpha_1; \alpha_{trat2} = \alpha_0 + \alpha_2; \dots; \alpha_{trat t} = \alpha_0 \text{ e}$$

$$\beta_{trat1} = \beta_0 + \beta_1; \beta_{trat2} = \beta_0 + \beta_2; \dots; \beta_{trat t} = \beta_0.$$

- Análise Preliminar dos Dados

Para verificar o comportamento da variável força ao longo do tempo em cada tratamento, fez-se, como análise exploratória, o gráfico das médias das observações de cada tempo para cada tratamento. Este gráfico está mostrado na Figura 3.1.

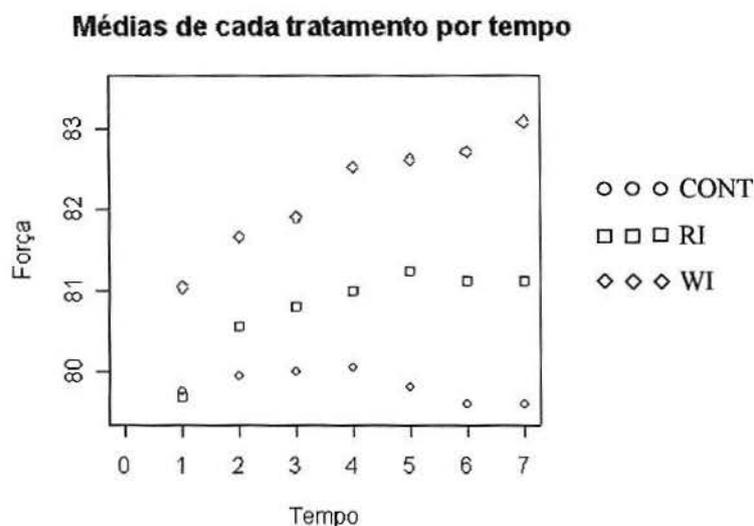


Figura 3.1: Gráfico das médias de força por tempo e por tratamento

Pela Figura 3.1, aparentemente a força dos sujeitos aumenta ao longo do tempo somente para os programas RI e WI. Para o programa CONT, a força parece ser constante. Aparentemente, o programa WI proporciona um aumento de força mais rapidamente que o programa RI. O ajuste linear parece ser adequado para cada um dos três tratamentos. Apesar de que há uma leve curvatura para o programa RI, não há sentido em assumir uma curva quadrática, em que a partir de determinado tempo o sujeito irá diminuir sua força.

O banco de dados *Weights* possui as colunas: Program (tratamento), Subj (sujeito), Time (Tempo) e Strength (Força). A coluna Subj não está numerada de forma contínua, ou seja, de 1 a 57. A numeração depende do tratamento, ou seja, para o tratamento CONT, vai de 1 a 20, para o RI, de 1 a 16 e, para o WI, de 1 a 21. Por isto, sempre que se quiser o efeito de cada sujeito, deve-se aninhar (com o comando Nest) o efeito de Sujeito com o efeito de Tratamento. Para analisar este modelo misto através do Proc MIXED do SAS, deve-se especificar como efeitos fixos Program e a interação Program*Time, e como efeitos aleatórios Subj(Program) e a interação Subj(Program)*Time. Para ambos os efeitos fixos e aleatórios deve-se determinar que o software não estime a média geral (Intercept). Note que, o efeito de Program irá ajustar um intercepto para cada tratamento, o efeito de Program*Time um

coeficiente angular para cada tratamento, o efeito de Subj(Program) um intercepto para cada sujeito e o efeito de Subj(Program)*Time um coeficiente angular para cada sujeito.

A seguir, procede-se a análise do banco de dados *Weights* através do Proc MIXED do SAS, utilizando o método de estimação de parâmetros de variâncias e covariâncias REML. Em um primeiro momento, supôs-se estruturas VC para as matrizes **D** e **R** (Análise 1). Após, supôs-se VC para **D** e AR(1) para **R** (Análise 2).

- Análise 1

Supõe estruturas VC para as matrizes **D** e **R**.

Abaixo, apresenta-se o teste de hipóteses para os efeitos fixos, cuja hipótese nula postula que os efeitos são iguais a zero.

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
program	3	54.6	13031.4	<.0001
time*program	3	54.6	6.74	0.0006

Conclui-se, a 95% de confiança, que pelo menos um dos interceptos e um dos coeficientes angulares dos tratamentos é significativamente diferente de zero.

Outros resultados obtidos estão mostrados na Tabela 3.3.

Est.	Estimativa	IC 95%	Est.	Estimativa	IC 95%
$\hat{\alpha}_{trat1}$	80,0143	(78,6394; 81,3892)	$\hat{u}_{1,1}$	4,5882	(2,8104; 6,3661)
$\hat{\alpha}_{trat2}$	79,9554	(78,4181; 81,4926)	$\hat{u}_{21,1}$	-0,9988	(-2,8937; 0,8960)
$\hat{\alpha}_{trat3}$	80,9456	(79,6038; 82,2874)	$\hat{u}_{37,1}$	3,0328	(1,2781; 4,7876)
$\hat{\beta}_{trat1}$	-0,04821	(-0,2179; 0,1215)	$\hat{u}_{1,2}$	0,3502	(0,04750; 0,6530)
$\hat{\beta}_{trat2}$	0,2098	(0,02010; 0,3995)	$\hat{u}_{21,2}$	-0,1250	(-0,4361; 0,1861)
$\hat{\beta}_{trat3}$	0,3197	(0,1541; 0,4853)	$\hat{u}_{37,2}$	-0,3935	(-0,6946; -0,09232)
$\hat{\sigma}_{\alpha}^2$	8,9571	(6,2597; 13,8770)	$\hat{\sigma}_{\epsilon}^2$	0,6360	(0,5429; 0,7553)
$\hat{\sigma}_{\beta}^2$	0,1206	(0,08076; 0,1997)			

Tabela 3.3: Alguns resultados do Proc MIXED do SAS para o exemplo *Weights* supondo estrutura VC

O intervalo de confiança para os componentes de variância apresentado na Tabela 3.3 foi feito através do Teste de Wald.

Por se pensar ser desnecessário mostrar aqui todos os interceptos e coeficientes angulares estimados para cada sujeito, a Tabela 3.3 apresenta apenas as estimativas para os interceptos e coeficientes angulares dos sujeitos 1, 21, e 37, que são o primeiro sujeito de cada tratamento.

As equações de regressão ajustadas para os indivíduos 1, 21 e 37 seriam, respectivamente:

$$\hat{y}_{1j1} = (80,0143 + 4,5882) + (-0,04821 + 0,3502)tempo_j = 84,6025 + 0,30199tempo_j;$$

$$\hat{y}_{21j2} = (79,9554 - 0,9988) + (0,2098 - 0,1250)tempo_j = 78,9566 + 0,0848tempo_j;$$

$$\hat{y}_{37j3} = (80,9456 + 3,0328) + (0,3197 - 0,3935)tempo_j = 83,9784 - 0,0738tempo_j.$$

Analisando a Tabela 3.3, pode-se concluir também que os interceptos de cada tratamento não são significativamente diferentes, já que os intervalos de confiança se sobrepõe. Este resultado parece razoável visto que os interceptos representam a média de força entre indivíduos de cada tratamento antes do início do estudo. Analisando as estimativas para os coeficientes angulares de cada tratamento conclui-se que, para o tratamento 1 (CONT), a tendência linear não é significativa, já que o intervalo de confiança contém o valor zero. Isto sugere que a força dos indivíduos se mantenha constante ao longo do tempo. Analogamente, conclui-se que a regressão linear é significativa para os tratamentos 2 e 3 (RI e WI), e aparentemente ambos os programas aumentam a força do indivíduo ao passar do tempo.

Para comparar a eficiência dos tratamentos WI e RI em aumentar a força dos indivíduos ao longo do tempo, deve-se comparar seus coeficientes angulares com o coeficiente angular do tratamento CONT. No tratamento CONT, os indivíduos não fazem levantamento de peso, ou seja, estão em seu estado natural. Os coeficientes angulares do tratamento WI e do tratamento CONT são significativamente diferentes ($\alpha=0,05$). Isto é visto pela não-sobreposição dos intervalos. Por isto, e pelo fato de que o coeficiente angular do tratamento WI é significativamente diferente de zero, este tratamento deve ser eficiente em aumentar a força do indivíduo ao longo do tempo. Já o coeficiente angular do tratamento RI não se diferencia significativamente do tratamento CONT. Portanto, não se pode considerar o tratamento RI como eficiente.

Para concluir-se sobre as diferenças entre os interceptos e coeficientes angulares dos sujeitos, deve-se olhar o intervalo de confiança do componente de variância correspondente e verificar se ele é significativamente diferente de zero. Porém, como mencionado na Seção 3.3, o teste de Wald não é válido para os componentes de variância e por isso não se pode concluir sobre seu resultado.

- Análise 2

Para este mesmo modelo misto, a análise foi procedida novamente, supondo uma estrutura de covariância AR(1) para a matriz \mathbf{R} . O objetivo desta análise é somente verificar quais as diferenças em termos de estimativas pontuais em relação à análise anterior. Por isso não serão apresentados intervalos de confiança e teste de hipóteses. Os resultados estão mostrados na Tabela 3.4.

As equações de regressão ajustadas para os indivíduos 1, 21 e 37 seriam, respectivamente:

$$\hat{y}_{1j1} = (79,8766 + 4,6769) + (-0,02975 + 0,2226)tempo_j = 84,5535 + 0,19285tempo_j ;$$

$$\hat{y}_{21j2} = (79,6914 - 0,9287) + (0,2335 - 0,04989)tempo_j = 78,7627 + 0,18361tempo_j ;$$

$$\hat{y}_{37j3} = (80,8104 + 1,8843) + (0,3369 - 0,1107)tempo_j = 82,6947 + 0,2262tempo_j .$$

Estimador	Estimativa	Estimador	Estimativa	Estimativa	Resultado
$\hat{\alpha}_{trat1}$	79,8766	$\hat{u}_{1,1}$	4,6769	$\hat{\alpha}_{suj.1} = \hat{\alpha}_{trat1} + \hat{u}_{1,1}$	84,5535
$\hat{\alpha}_{trat2}$	79,6914	$\hat{u}_{21,1}$	-0,9287	$\hat{\alpha}_{suj.21} = \hat{\alpha}_{trat2} + \hat{u}_{21,1}$	78,7627
$\hat{\alpha}_{trat3}$	80,8104	$\hat{u}_{37,1}$	1,8843	$\hat{\alpha}_{suj.37} = \hat{\alpha}_{trat3} + \hat{u}_{37,1}$	82,6947
$\hat{\beta}_{trat1}$	-0,02975	$\hat{u}_{1,2}$	0,2226	$\hat{\beta}_{suj.1} = \hat{\beta}_{trat1} + \hat{u}_{1,2}$	0,19285
$\hat{\beta}_{trat2}$	0,2335	$\hat{u}_{21,2}$	-0,04989	$\hat{\beta}_{suj.21} = \hat{\beta}_{trat2} + \hat{u}_{21,2}$	0,18361
$\hat{\beta}_{trat3}$	0,3369	$\hat{u}_{37,2}$	-0,1107	$\hat{\beta}_{suj.37} = \hat{\beta}_{trat3} + \hat{u}_{37,2}$	0,2262
$\hat{\sigma}_{\alpha}^2$	7,4597	$\hat{\sigma}_c^2$	1,8355		
$\hat{\sigma}_{\beta}^2$	0,05414	$\hat{\rho}$	0,7300		

Tabela 3.4: Alguns resultados do Proc MIXED do SAS para o exemplo *Weights* supondo estrutura AR(1)

Com o Proc MIXED do SAS, é possível que se peça que o programa imprima a matriz \mathbf{R} estimada. O resultado para o exemplo está a seguir.

$$\hat{\mathbf{R}} = \begin{bmatrix} 1,8355 & 1,3399 & 0,9781 & 0,7140 & 0,5212 & 0,3805 & 0,2777 \\ 1,3399 & 1,8355 & 1,3399 & 0,9781 & 0,7140 & 0,5212 & 0,3805 \\ 0,9781 & 1,3399 & 1,8355 & 1,3399 & 0,9781 & 0,7140 & 0,5212 \\ 0,7140 & 0,9781 & 1,3399 & 1,8355 & 1,3399 & 0,9781 & 0,7140 \\ 0,5212 & 0,7140 & 0,9781 & 1,3399 & 1,8355 & 1,3399 & 0,9781 \\ 0,3805 & 0,5212 & 0,7140 & 0,9781 & 1,3399 & 1,8355 & 1,3399 \\ 0,2777 & 0,3805 & 0,5212 & 0,7140 & 0,9781 & 1,3399 & 1,8355 \end{bmatrix}.$$

Para se ter uma melhor idéia da diferença nos ajustes das retas por sujeito e por tratamento ao se supor estruturas de covariância VC e AR(1) para a matriz \mathbf{R} , produziu-se os gráficos mostrados na Figura 3.2. Através destes gráficos, percebe-se que visualmente não está claro qual o modelo que estimou retas de melhor ajuste. Há na literatura medidas a serem calculadas para se comparar modelos e ajudar a decidir qual o melhor, mas por não fazer parte dos objetivos deste trabalho, este tópico não será abordado. Porém, nota-se que a maior diferença entre as retas ajustadas se encontra nos ajustes por sujeitos, o que era esperado já que se alterou a estrutura da matriz \mathbf{R} , que representa a variabilidade dentro dos sujeitos.

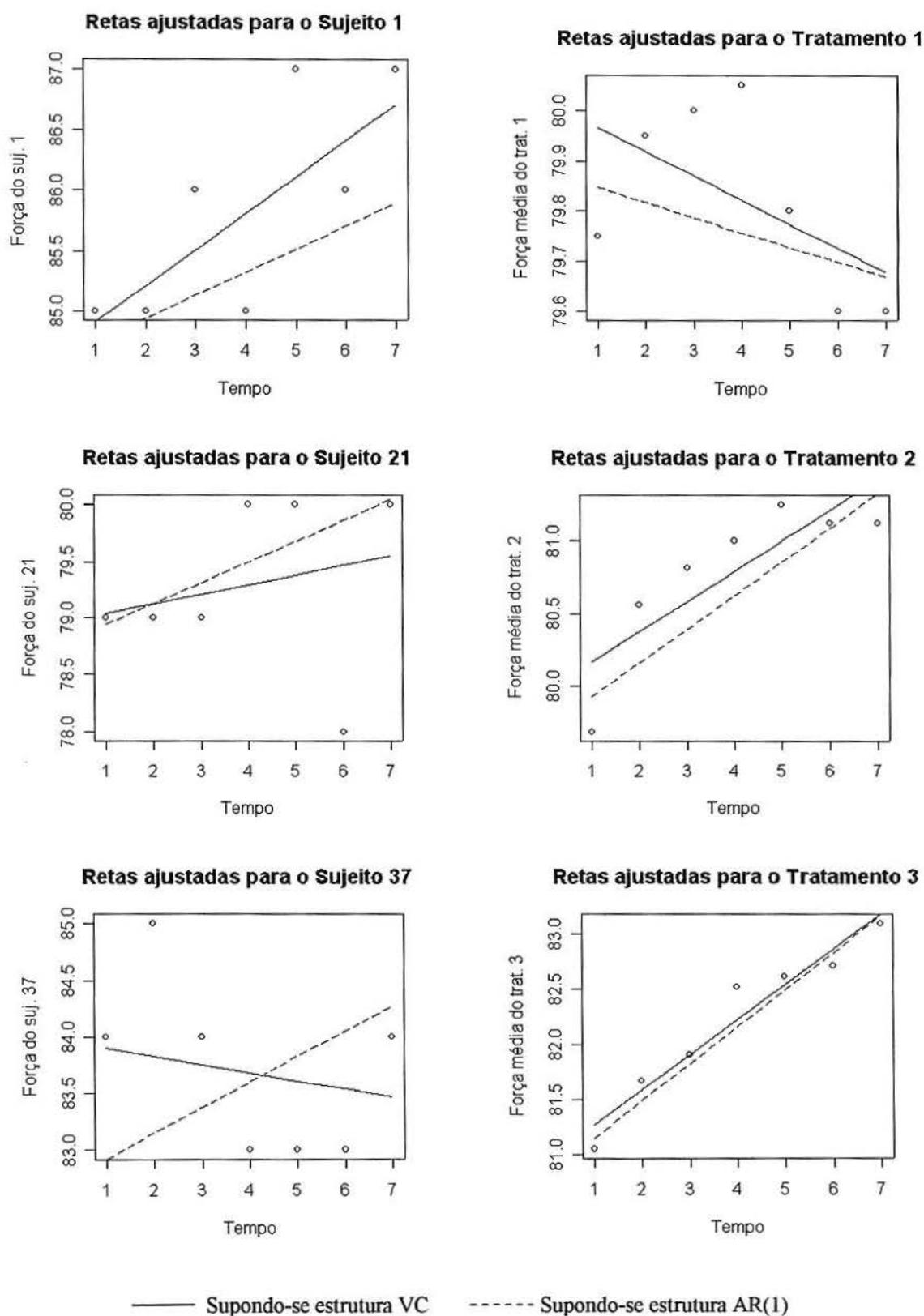


Figura 3.2: Retas ajustadas aos sujeitos 1, 21 e 37 e aos tratamentos, supondo-se estruturas VC e AR(1) para a matriz R e VC para D. Nos gráficos para os sujeitos, pontos representam a força medida naquele sujeito a cada tempo. Nos gráficos de cada tratamento, pontos representam a média das observações dos indivíduos daquele tratamento a cada tempo

4. ABORDAGEM BAYESIANA PARA MODELOS MISTOS NORMAIS

A principal diferença entre a inferência clássica e a inferência Bayesiana é que a primeira supõe que os parâmetros desconhecidos do modelo analisado são constantes, enquanto a segunda os considera variáveis aleatórias. Sendo assim, a abordagem Bayesiana baseia-se em distribuições de probabilidade para os parâmetros desconhecidos. Antes de observar uma amostra de dados estas distribuições são construídas baseadas em toda a informação existente sobre estes parâmetros e são chamadas de distribuição a priori. A idéia da inferência Bayesiana é combinar a informação a priori com a informação proveniente dos dados amostrais. Ou seja, combinar a distribuição a priori e a função de verossimilhança. Esta combinação é feita através do Teorema de Bayes, dando origem à distribuição a posteriori. Esta distribuição mostra o comportamento dos parâmetros após os dados amostrais serem observados e a partir dela são feitas as inferências sobre os parâmetros e suas devidas interpretações.

Todo conjunto de dados pode ser analisado utilizando-se as técnicas Bayesianas. Entretanto, alguns problemas, como modelos com muitos parâmetros, exigem cálculos muito difíceis, ou simplesmente impossíveis de serem executados. Este fato foi um empecilho para a utilização da inferência Bayesiana por muito tempo. Somente com o avanço dos métodos computacionais, especialmente dos métodos MCMC introduzidos no Capítulo 2, que os métodos Bayesianos puderam ser aplicados a muitos problemas estatísticos, como o caso dos modelos mistos, tratados neste trabalho.

Neste capítulo os principais conceitos da inferência Bayesiana serão brevemente apresentados na Seção 4.1. Define-se a distribuição a posteriori e discute-se sobre maneiras de especificar a distribuição a priori, além da apresentação de outros conceitos nesta seção. Na Seção 4.2 estes conceitos são aplicados ao caso dos modelos mistos normais, mostrando as prioris usualmente utilizadas e a dificuldade de obtenção da distribuição a posteriori de interesse. A maneira de se aplicar o algoritmo “Gibbs Sampling” nas análises Bayesianas para modelos mistos é mostrada na Seção 4.3, discutindo brevemente o uso dos softwares R e WinBUGS. Por fim, alguns exemplos são analisados na Seção 4.4 e um estudo de simulação procedido é mostrado na Seção 4.5.

4.1. Conceitos básicos de Inferência Bayesiana

A inferência Bayesiana, como já mencionado anteriormente, se fundamenta na obtenção da distribuição a posteriori dos parâmetros desconhecidos. Para obter esta distribuição, dois fatores são necessários: a distribuição a priori, denominada assim por ser a distribuição de probabilidade dos parâmetros antes de se observar qualquer dado amostral; e a função de verossimilhança, que fornece a informação dos dados observados. Tendo a distribuição a priori e a função de verossimilhança, a distribuição a posteriori resulta de uma aplicação direta do Teorema de Bayes, apresentando a seguinte expressão:

$$\pi(\boldsymbol{\theta} | \mathbf{y}) = \frac{\pi(\boldsymbol{\theta})L(\boldsymbol{\theta} | \mathbf{y})}{f(\mathbf{y})},$$

onde $\boldsymbol{\theta}$ é o vetor de parâmetros, \mathbf{y} é o vetor de dados observados, $\pi(\boldsymbol{\theta})$ é a distribuição a priori de $\boldsymbol{\theta}$, $L(\boldsymbol{\theta} | \mathbf{y})$ é a função de verossimilhança de $\boldsymbol{\theta}$ e $f(\mathbf{y}) = \int \cdots \int_{\Theta} \pi(\boldsymbol{\theta})L(\boldsymbol{\theta} | \mathbf{y})d\boldsymbol{\theta}$ é um valor constante em relação a $\boldsymbol{\theta}$, e em certos casos, não pode ser obtida analiticamente. Note que a denominação distribuição a posteriori vem do fato de que esta é a densidade de probabilidade de $\boldsymbol{\theta}$ condicional aos dados observados, ou seja, após a realização da amostra.

Um recurso muito utilizado na inferência Bayesiana para determinar a distribuição a posteriori é o da proporcionalidade. Já que $f(\mathbf{y})$ não depende de $\boldsymbol{\theta}$, pode-se estabelecer que

$$\pi(\boldsymbol{\theta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta})L(\boldsymbol{\theta} | \mathbf{y}). \quad (4.1)$$

Se através de (4.1) for possível identificar-se o núcleo de alguma distribuição conhecida, se saberá que a posteriori segue esta distribuição sem necessitar o cálculo de $f(\mathbf{y})$. Chama-se núcleo de uma distribuição de probabilidade para $\boldsymbol{\theta}$ a expressão resultante da retirada de todas as constantes em relação a $\boldsymbol{\theta}$ da densidade de probabilidade $f(\boldsymbol{\theta})$.

Em muitos casos, não se consegue determinar a distribuição a posteriori analiticamente nem mesmo através do caminho da proporcionalidade. Para estes casos, os métodos MCMC são soluções computacionais que permitem que a distribuição a posteriori de interesse seja aproximada.

A distribuição a priori deve ser escolhida pelo pesquisador, que deve especificar uma forma para esta distribuição e valores para seus parâmetros. Denomina-se hiperparâmetros aos

parâmetros da distribuição a priori, para diferenciá-los dos parâmetros desconhecidos que se quer analisar. A priori é determinada utilizando-se todo conhecimento disponível sobre o parâmetro que se está estudando antes de se obter qualquer dado. Este conhecimento pode ser obtido a partir da opinião de especialistas, posterioris de estudos anteriores, dentre outras fontes. Se não há absolutamente qualquer conhecimento sobre o parâmetro, pode-se construir uma priori não-informativa, e assim as estimativas obtidas a partir da distribuição a posteriori serão baseadas apenas na função de verossimilhança dos dados. A distribuição a priori não-informativa dependerá do parâmetro estudado e poderá muitas vezes não caracterizar uma distribuição de probabilidade. Isto porque a integral em todos os valores possíveis do parâmetro desconhecido pode não resultar em 1. A estas prioris, denomina-se prioris impróprias. Um exemplo de priori imprópria seria $\pi(\theta) \propto c$, onde c é uma constante. Um cuidado que se deve ter ao especificar priori imprópria é o de averiguar se a respectiva posteriori é realmente uma distribuição de probabilidade. No caso de usar métodos de simulação para estimar os parâmetros é comum não se utilizar prioris impróprias, visto que é complicado verificar se a distribuição a posteriori estimada realmente caracteriza uma distribuição de probabilidade.

Um conceito muito utilizado ao especificar-se distribuições a priori é o de prioris conjugadas. São prioris que têm a propriedade de “conjugar-se” com a função de verossimilhança, de tal forma que a distribuição a posteriori será da mesma família da distribuição a priori. O uso das prioris conjugadas traz vantagens aos procedimentos de inferência, pois facilita os cálculos analíticos. Prioris conjugadas são sempre prioris próprias. Dependendo dos valores especificados para os hiperparâmetros, uma priori conjugada pode ser não informativa. Por exemplo, uma distribuição normal pode caracterizar uma priori não-informativa se o valor estabelecido para a variância for muito grande.

Quando \mathbf{y} ainda não foi observado, $f(\mathbf{y}) = \int \pi(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}$ é uma distribuição de probabilidade, denominada distribuição preditiva a priori. Ela fornece a distribuição esperada para \mathbf{y} já que $f(\mathbf{y}) = E[L(\boldsymbol{\theta}|\mathbf{y})]$. Um conceito similar pode ser aplicado para a predição de uma futura observação \mathbf{x} após \mathbf{y} observado. Esta predição deve ser baseada na distribuição de \mathbf{x} dado \mathbf{y} , ou seja, na distribuição preditiva a posteriori, definida por

$$\pi(\mathbf{x}|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|\mathbf{y})L(\boldsymbol{\theta}|\mathbf{x})d\boldsymbol{\theta}.$$

Nota-se que, analogamente à distribuição preditiva a priori, a preditiva a posteriori fornece a distribuição esperada para $\mathbf{x} | \mathbf{y}$.

Com a distribuição a posteriori determinada, podem-se obter estimativas pontuais para cada componente de $\boldsymbol{\theta}$, digamos θ_i , através do cálculo da esperança, moda ou mediana deste parâmetro, segundo sua distribuição a posteriori marginal:

$$\pi(\theta_i | \mathbf{y}) = \int \pi(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta}_{-i}, \quad (4.2)$$

onde $\boldsymbol{\theta}_{-i}$ é o vetor $\boldsymbol{\theta}$ excluindo-se θ_i .

O estimador pontual mais utilizado, quando temos uma expressão fechada para a posteriori, é a moda, pela idéia análoga da maximização de funções de verossimilhança de se estimar um parâmetro pelo seu valor mais verossímil.

Aqui, pode-se notar uma das vantagens da inferência Bayesiana em relação à inferência clássica. Inferências sobre cada componente de $\boldsymbol{\theta}$ são feitas sem necessitar fazer suposições para os demais componentes, como é o caso da inferência clássica.

Pode-se obter também a estimativa da variância do parâmetro e construir intervalos de credibilidade. Estes intervalos são calculados diretamente da distribuição a posteriori, determinando-se os quantis de interesse. Um tipo especial de intervalo de credibilidade, conhecido como HDR (“Highest Density Region”), é construído de tal forma que a densidade de probabilidade dos pontos dentro do intervalo é maior que a de todos os pontos fora dele e a probabilidade do verdadeiro valor do parâmetro estar entre os limites definidos pelo intervalo é igual a $1 - \alpha$, onde α é escolhido arbitrariamente. Quando se está estimando a distribuição a posteriori por simulação, não é possível calcular o intervalo HDR. Por isso, estima-se um intervalo de credibilidade central, calculando-se os quantis $\alpha/2$ e $1 - \alpha/2$ da distribuição a posteriori. Entretanto, não é necessário obter-se a distribuição de estimadores como na inferência clássica, que muitas vezes recorre a aproximações destas distribuições.

Assim sendo, percebe-se que a inferência Bayesiana tem grandes vantagens em relação à clássica, mesmo quando não se tem nenhuma informação a priori sobre o parâmetro considerado, já que as estimativas pontuais e por intervalo são exatas, quando se tem uma

expressão fechada para a posteriori, e não necessitam de estimativas de erro padrão e cálculo de distribuições de estimadores.

A seguir, os conceitos Bayesianos apresentados serão aplicados aos modelos mistos normais.

4.2. Análise Bayesiana para modelos mistos normais

Para se fazer uma análise estatística através das técnicas Bayesianas para o modelo misto normal especificado em (3.1), primeiramente precisa-se definir a distribuição a priori conjunta para os parâmetros desconhecidos, que são $\boldsymbol{\beta}$, \mathbf{D} e \mathbf{R} . Ou seja, precisa-se determinar a priori conjunta $\pi(\boldsymbol{\beta}, \mathbf{D}, \mathbf{R})$. Usualmente, supõe-se independência a priori entre estes parâmetros. Conseqüentemente, $\pi(\boldsymbol{\beta}, \mathbf{D}, \mathbf{R}) = \pi(\boldsymbol{\beta})\pi(\mathbf{D})\pi(\mathbf{R})$. Assumindo a independência a priori, necessita-se especificar distribuições para os parâmetros $\boldsymbol{\beta}$, \mathbf{D} e \mathbf{R} separadamente. Deve-se notar que o vetor de efeitos aleatórios \mathbf{u} já tem sua distribuição de probabilidade especificada pelas suposições do modelo e por isso não se determina distribuição a priori para ele.

Segundo a formulação do modelo misto visto como um modelo hierárquico em (3.1) e (3.2), as distribuições a priori $\pi(\boldsymbol{\beta})$, $\pi(\mathbf{D})$ e $\pi(\mathbf{R})$ correspondem a um terceiro estágio da hierarquia.

A distribuição a priori usual para $\boldsymbol{\beta}$ é

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}^*, \mathbf{H}),$$

onde $\boldsymbol{\beta}^*$ é um vetor conhecido, de dimensão $p \times 1$, e \mathbf{H} é uma matriz de variâncias e covariâncias conhecida, de dimensão $p \times p$. Uma priori não-informativa para $\boldsymbol{\beta}$ pode ser obtida especificando-se $\mathbf{H} = \sigma^2 \mathbf{I}_p$, onde $\sigma^2 \rightarrow \infty$.

As distribuições a priori para \mathbf{D} e \mathbf{R} dependerão da estrutura de covariância adotada para cada uma delas. Em modelos normais, o mais usual é parametrizar componentes de variância e matrizes de variâncias e covariâncias como componentes e matrizes de precisão, pois isto facilita os cálculos das posteriores. Um componente de precisão é o recíproco do componente de variância correspondente, e uma matriz de precisão é a inversa da matriz de

variância correspondente. Assim, aqui serão especificadas distribuições a priori para os componentes e matrizes de precisão.

- Estrutura de Covariância VC

Se a estrutura adotada for VC, pode-se adotar a distribuição Gama como priori para cada um dos componentes de precisão da matriz, ou seja, $\sigma_w^{-2} \sim G(a_w, b_w)$. A priori não-informativa neste caso é obtida quando $a_w, b_w \rightarrow 0$. Outra distribuição a priori não-informativa que poderia ser obtida seria especificar $\sigma_w \sim U(0, c_w)$, onde $c_w \rightarrow \infty$. Neste

caso, a distribuição de probabilidade para σ_w^{-2} tem densidade $f(\sigma_w^{-2}) = \frac{(\sigma_w^{-2})^{-3/2}}{2c_w}$,

$$c_w^{-2} \leq \sigma_w^{-2} < +\infty.$$

- Estrutura de Covariância AR(1)

Se uma estrutura AR(1) for suposta para a matriz de variâncias e covariâncias, pode-se especificar uma distribuição Gama para o componente de precisão σ^{-2} , e uma priori para o parâmetro ρ poderia ser $\rho \sim U(a, b)$, onde $a < b$ e $a \geq -1$ e $b \leq 1$. Uma priori não-informativa para ρ resulta se for definido $a = -1$ e $b = 1$.

- Estrutura de Covariância UN

Já no caso da estrutura UN, não se precisa assumir uma priori para cada componente de variância para a matriz. As distribuições Wishart e Inversa-Wishart são distribuições de probabilidade para, respectivamente, as matrizes de precisão e de variâncias e covariâncias. São uma generalização multivariada da distribuição Gama. Além disso, são prioris conjugadas para, respectivamente, a matriz de precisão e de variâncias e covariâncias de uma distribuição normal multivariada. Segundo notação de Gelman et al. (1997), se $W \sim Wishart(\nu, S)$, em que W é uma matriz positiva-definida simétrica $k \times k$, ν é o número de graus de liberdade e S é uma matriz de escala positiva-definida simétrica $k \times k$, então a função densidade de probabilidade é dada por

$$f_w(w) = \left(2^{k/2} \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma\left(\frac{\nu+1-i}{2}\right) \right)^{-1} |S|^{-\nu/2} |w|^{(\nu-k)/2} \exp\left(-\frac{1}{2} tr(S^{-1}w)\right).$$

Da mesma forma, se $W \sim \text{Inv-Wishart}(\nu, S^{-1})$, em que W é uma matriz positiva-definida simétrica $k \times k$, ν é o número de graus de liberdade e S é uma matriz de escala positiva-definida simétrica $k \times k$, então a função densidade de probabilidade é dada por

$$f_W(w) = \left(2^{k/2} \pi^{k(k-1)/4} \prod_{i=1}^k \Gamma\left(\frac{\nu+1-i}{2}\right) \right)^{-1} |S|^{k/2} |w|^{-(\nu+k+1)/2} \exp\left(-\frac{1}{2} \text{tr}(Sw^{-1})\right).$$

Uma priori não-informativa tanto para a distribuição Wishart quanto para a Inversa-Wishart é obtida quando $\nu = k$.

Após a especificação das distribuições a priori apropriadas para os parâmetros β , D e R , levando em consideração as estruturas de covariância assumidas para as matrizes D e R , o próximo passo seria a derivação da distribuição a posteriori. Porém, independente da estrutura escolhida, nota-se ser complicado calcular-se analiticamente a distribuição a posteriori, ou reconhecer o núcleo de alguma distribuição multivariada conhecida através do caminho da proporcionalidade. O algoritmo de “Gibbs Sampling” pode ser aplicado neste caso, já que, para os modelos mistos, é possível derivar-se as distribuições condicionais completas.

4.3. Distribuições condicionais completas para modelos mistos normais

Como explicado na Seção 2.3, as distribuições condicionais completas de todos os parâmetros desconhecidos são necessárias para o “Gibbs Sampling”. Estas distribuições também dependem das distribuições a priori especificadas. A seguir, as distribuições condicionais para o modelo de componentes de variância apresentado na Subseção 3.4.1 e o modelo de dados longitudinais utilizado no banco de dados *Weights*, na Subseção 3.4.2, serão mostradas. As condicionais completas que serão mostradas foram derivadas a partir da formulação hierárquica destes dois modelos.

4.3.1. Modelo de Componentes de Variância

Para se aplicar as técnicas Bayesianas ao modelo de componentes de variância descrito pelos estágios em (3.17) e (3.18), ou seja,

$$\text{Estágio 1: } y_{ij} | \theta_i \sim N(\theta_i, \sigma_\epsilon^2)$$

$$\text{Estágio 2: } \theta_i \sim N(\mu_\theta, \sigma_\theta^2),$$

deve-se atribuir prioris para os parâmetros μ_θ , σ_θ^{-2} e σ_ε^{-2} . Usualmente, supõe-se independência a priori entre eles e atribui-se uma distribuição normal para μ_θ , isto é, $\mu_\theta \sim N(\mu_0, \sigma_0^2)$, e gama para σ_θ^{-2} e σ_ε^{-2} , isto é, $\sigma_\theta^{-2} \sim G(a_1, b_1)$ e $\sigma_\varepsilon^{-2} \sim G(a_2, b_2)$. Note que não é necessário atribuir-se distribuição a priori para os parâmetros $\theta_i, i = 1, \dots, K$, pois já há uma distribuição especificada para eles no Estágio 2.

As distribuições condicionais completas do modelo de componentes de variância com as distribuições a priori especificadas anteriormente são as que seguem:

$$\sigma_\theta^2 | y, \mu_\theta, \theta, \sigma_\varepsilon^2 \sim IG\left(a_1 + \frac{1}{2}K, b_1 + \frac{1}{2} \sum_{i=1}^K (\theta_i - \mu_\theta)^2\right);$$

$$\sigma_\varepsilon^2 | y, \mu_\theta, \theta, \sigma_\theta^2 \sim IG\left(a_2 + \frac{1}{2}KJ, b_2 + \frac{1}{2} \sum_{j=1}^J \sum_{i=1}^K (y_{ij} - \theta_i)^2\right);$$

$$\mu_\theta | y, \theta, \sigma_\theta^2, \sigma_\varepsilon^2 \sim N\left(\frac{\sigma_\theta^2 \mu_0 + \sigma_0^2 \sum_{i=1}^K \theta_i}{\sigma_\theta^2 + K\sigma_0^2}, \frac{\sigma_\theta^2 \sigma_0^2}{\sigma_\theta^2 + K\sigma_0^2}\right) e$$

$$\theta_i | y, \mu_\theta, \sigma_\theta^2, \sigma_\varepsilon^2 \sim N\left(\frac{J\sigma_\theta^2}{J\sigma_\theta^2 + \sigma_\varepsilon^2} \bar{y}_i + \frac{\sigma_\varepsilon^2}{J\sigma_\theta^2 + \sigma_\varepsilon^2} \mu_\theta, \frac{\sigma_\theta^2 \sigma_\varepsilon^2}{J\sigma_\theta^2 + \sigma_\varepsilon^2}\right),$$

onde $\bar{y}_i = \frac{1}{J} \sum_{j=1}^J y_{ij}$ e *IG* se refere à distribuição Inversa-Gama. Estas distribuições condicionais completas foram retiradas de Gelfand e Smith (1990).

4.3.2. Modelo para Dados Longitudinais

Pode-se analisar o modelo hierárquico definido em (3.20) e (3.21), ou seja,

$$\text{Estágio 1: } y_{ijk} | \alpha_{suj,i}, \beta_{suj,i} \stackrel{iid}{\sim} N(\alpha_{suj,i} + \beta_{suj,i} \text{tempo}_j, \sigma_\varepsilon^2)$$

$$\text{Estágio 2: } \begin{cases} \alpha_{suj,i} \stackrel{iid}{\sim} N(\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2 + \dots + \alpha_{(t-1)} t_{(t-1)}, \sigma_\alpha^2) \\ \beta_{suj,i} \stackrel{iid}{\sim} N(\beta_0 + \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_{(t-1)} t_{(t-1)}, \sigma_\beta^2) \end{cases},$$

através da abordagem Bayesiana determinando prioris para os parâmetros $\alpha_0, \dots, \alpha_{t-1}, \beta_0, \dots, \beta_{t-1}, \sigma_\varepsilon^{-2}, \sigma_\alpha^{-2}$ e σ_β^{-2} . Não é necessário atribuir-se distribuições a priori para os interceptos e coeficientes angulares de cada sujeito, pois eles já têm uma distribuição de probabilidade associada, inerente ao modelo. Para os componentes de precisão, pode-se especificar as mesmas distribuições a priori do que o caso da estrutura de covariância VC. Para os α_s e β_s , $s = 0, \dots, t-1$, pode-se especificar distribuições normais.

Assim, considere as distribuições a priori: $\alpha_h \sim N(\alpha_h, c_h^2)$ $h = 0, 1, 2$;
 $\beta_h \sim N(b_h, d_h^2)$ $h = 0, 1, 2$; $\sigma_\varepsilon^{-2} \sim G(e_1, e_2)$; $\sigma_\alpha^{-2} \sim f(\sigma_\alpha^{-2}) = \frac{(\sigma_\alpha^{-2})^{-3/2}}{2c_\alpha}$, $c_\alpha^{-2} \leq \sigma_\alpha^{-2} < +\infty$ e
 $\sigma_\beta^{-2} \sim f(\sigma_\beta^{-2}) = \frac{(\sigma_\beta^{-2})^{-3/2}}{2c_\beta}$, $c_\beta^{-2} \leq \sigma_\beta^{-2} < +\infty$. Note que já se está particularizando o modelo para o caso de 3 tratamentos, como o exemplo *Weights*. Com algumas manipulações algébricas chega-se às condicionais completas a seguir.

$$\pi(\alpha_0 | \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N\left(\frac{c_0^2 \left(\sum_{i=1}^n \alpha_{suj.i} - n_{trat1} \alpha_1 - n_{trat2} \alpha_2\right) + a_0 \sigma_\alpha^2}{nc_0^2 + \sigma_\alpha^2}, \frac{c_0^2 \sigma_\alpha^2}{nc_0^2 + \sigma_\alpha^2}\right);$$

$$\pi(\alpha_1 | \alpha_0, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N\left(\frac{c_1^2 \left(\sum_{i=1}^n t_1 \alpha_{suj.i} - n_{trat1} \alpha_0\right) + a_1 \sigma_\alpha^2}{n_{trat1} c_1^2 + \sigma_\alpha^2}, \frac{c_1^2 \sigma_\alpha^2}{n_{trat1} c_1^2 + \sigma_\alpha^2}\right);$$

$$\pi(\alpha_2 | \alpha_0, \alpha_1, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i = 1, \dots, n, \beta_{suj.i} \ i = 1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N\left(\frac{c_2^2 \left(\sum_{i=1}^n t_2 \alpha_{suj.i} - n_{trat2} \alpha_0\right) + a_2 \sigma_\alpha^2}{n_{trat2} c_2^2 + \sigma_\alpha^2}, \frac{c_2^2 \sigma_\alpha^2}{n_{trat2} c_2^2 + \sigma_\alpha^2}\right);$$

$$\pi(\beta_0 | \alpha_0, \alpha_1, \alpha_2, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left(\frac{d_0^2 \left(\sum_{i=1}^n \beta_{suj.i} - n_{trat1} \beta_1 - n_{trat2} \beta_2 \right) + b_0 \sigma_\beta^2}{nd_0^2 + \sigma_\beta^2}; \frac{d_0^2 \sigma_\beta^2}{nd_0^2 + \sigma_\beta^2} \right);$$

$$\pi(\beta_1 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left(\frac{d_1^2 \left(\sum_{i=1}^n t_1 \beta_{suj.i} - n_{trat1} \beta_0 \right) + b_1 \sigma_\beta^2}{n_{trat1} d_1^2 + \sigma_\beta^2}; \frac{d_1^2 \sigma_\beta^2}{n_{trat1} d_1^2 + \sigma_\beta^2} \right);$$

$$\pi(\beta_2 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left(\frac{d_2^2 \left(\sum_{i=1}^n t_2 \beta_{suj.i} - n_{trat2} \beta_0 \right) + b_2 \sigma_\beta^2}{n_{trat2} d_2^2 + \sigma_\beta^2}; \frac{d_2^2 \sigma_\beta^2}{n_{trat2} d_2^2 + \sigma_\beta^2} \right);$$

$$\pi(\alpha_{suj.i} | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, i-1, i+1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left(\frac{\sigma_\alpha^2 \sum_{i=1}^n (y_{ijk} - \beta_{suj.i} tempo_j) + \sigma_\varepsilon^2 (\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2)}{g \sigma_\alpha^2 + \sigma_\varepsilon^2}; \frac{\sigma_\alpha^2 \sigma_\varepsilon^2}{g \sigma_\alpha^2 + \sigma_\varepsilon^2} \right);$$

$$\pi(\beta_{suj.i} | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, i-1, i+1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$N \left(\frac{\sigma_\beta^2 \sum_{i=1}^n tempo_j (y_{ijk} - \alpha_{suj.i}) + \sigma_\varepsilon^2 (\beta_0 + \beta_1 t_1 + \beta_2 t_2)}{\sigma_\beta^2 \sum_{j=1}^g tempo_j^2 + \sigma_\varepsilon^2}; \frac{\sigma_\beta^2 \sigma_\varepsilon^2}{\sigma_\beta^2 \sum_{j=1}^g tempo_j^2 + \sigma_\varepsilon^2} \right);$$

$$\pi(\sigma_\alpha^2 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\beta^2, \sigma_\varepsilon^2) \sim$$

$$IG \left(\frac{n-1}{2}; \frac{\sum_{i=1}^n (\alpha_{suj.i} - (\alpha_0 + \alpha_1 t_1 + \alpha_2 t_2))^2}{2} \right);$$

$$\pi(\sigma_\beta^2 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\alpha^2, \sigma_\varepsilon^2) \sim$$

$$IG \left(\frac{n-1}{2}, \frac{\sum_{i=1}^n (\beta_{suj.i} - (\beta_0 + \beta_1 t_1 + \beta_2 t_2))^2}{2} \right) e$$

$$\pi(\sigma_\varepsilon^2 | \alpha_0, \alpha_1, \alpha_2, \beta_0, \beta_1, \beta_2, \alpha_{suj.i} \ i=1, \dots, n, \beta_{suj.i} \ i=1, \dots, n, \sigma_\alpha^2, \sigma_\beta^2) \sim$$

$$IG \left(\frac{ng}{2} + e_1, \frac{\sum_{i=1}^n \sum_{j=1}^g (y_{ijk} - (\alpha_{suj.i} + \beta_{suj.i} tempo_j))^2}{2} + e_2 \right).$$

Com estas distribuições condicionais, a aplicação do Gibbs Sampling pode ser feita em qualquer software que gere dados aleatórios das distribuições Normal e Gama, já que para gerar números de uma distribuição Inversa-Gama somente é necessário tomar o recíproco dos valores gerados de uma Gama (Gelman et al., 1997).

Entretanto, utilizando o software WinBUGS, pode-se analisar estes e outros problemas Bayesianos através do “Gibbs Sampling” sem necessitar-se conhecer as expressões das distribuições condicionais completas. No WinBUGS, somente precisa-se informar as distribuições a priori e a função de verossimilhança. Este software está disponível gratuitamente na página <http://www.mrc-bsu.cam.ac.uk/bugs/> e atualmente está na versão 1.4.1. Ao instalá-lo, é necessário fazer um registro e obter uma chave de acesso para poder se utilizar todos os recursos do software. Este registro é fácil de ser feito, seguindo as instruções da página, e tem apenas a finalidade de prover informações aos idealizadores e patrocinadores do software de quantas pessoas o estão utilizando e com qual objetivo.

O WinBUGS possui um Manual do Usuário, que o acompanha no pacote de instalação, e dois documentos com muitos exemplos. Para o usuário iniciante, recomenda-se como primeira leitura o Tutorial que está inserido no Manual do Usuário. Alguns problemas quanto à programação e mensagens de erro podem ser resolvidos enviando-se um *e-mail* à equipe que idealizou o software (bugs@mrc-bsu.cam.ac.uk). Há também uma lista de discussão de usuários do WinBUGS (<http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A0=bugs>), na qual as pessoas cadastradas podem enviar *e-mails* que serão recebidos por todos os demais cadastrados.

O software R também possui alguns recursos úteis para se utilizar o “Gibbs Sampling”. Dentre eles destaca-se o pacote *MCMCpack*, disponível na Internet, que possui funções como *rwish* que gera números aleatórios da distribuição Wishart, além de funções que aplicam MCMC a modelos de regressão linear, logística, entre outros.

4.4. Exemplos

4.4.1. Modelo de Componentes de Variância

A análise do exemplo introduzido na Subseção 3.4.1, agora sob a abordagem Bayesiana, foi procedida através do software R, sendo que a programação utilizada está apresentada no Apêndice B.

Adotaram-se distribuições a priori não-informativas para todos os parâmetros do modelo. Assim, os valores definidos para os hiperparâmetros (parâmetros das distribuições a priori) foram os seguintes: $\mu_0 = 0$; $\sigma_0^2 = 10000$; $a_1 = 0$; $b_1 = 0$; $a_2 = 0$ e $b_2 = 0$. Sabe-se que a distribuição Gama exige que seus dois parâmetros sejam maiores ou iguais a zero, porém, sabe-se que a priori não-informativa para esta distribuição tem parâmetros o mais próximo a zero quanto possível. Analisando-se as expressões para as condicionais completas de σ_θ^2 e σ_ε^2 , percebe-se que mesmo atribuindo-se valores nulos para a_1, b_1, a_2 e b_2 não conduzirá a uma distribuição condicional completa imprópria.

Simulou-se uma cadeia de tamanho $m = 10000$ iterações, com um período de “burn-in” de $b = 1000$ iterações. Para a cadeia gerada, desconsideravam-se os valores gerados no período de “burn-in” e produziam-se as estimativas para os parâmetros e para as distribuições a posteriori marginais da mesma maneira mostrada na Subseção 2.3.3.

Utilizando-se a equação (2.8), estimaram-se as distribuições a posteriori marginais de cada parâmetro, cujos resultados estão mostrados nos gráficos da Figura 4.1.

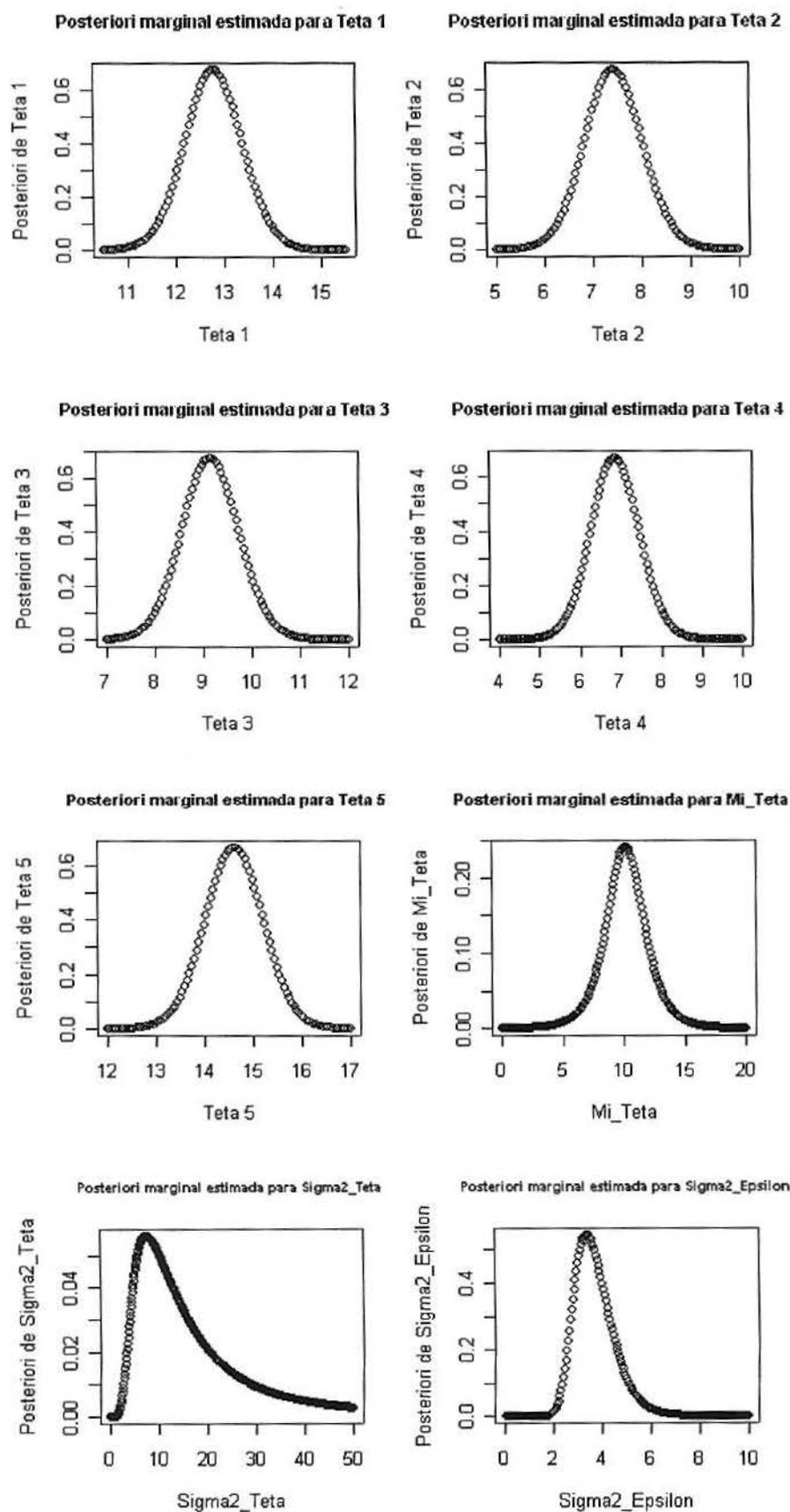


Figura 4.1: Posterioris marginais estimadas para o Modelo de Componentes de Variância

A apresentação do gráfico da posteriori traz muito mais informação sobre o parâmetro que se está querendo estimar do que medidas resumo tais como as que apresentaremos na Tabela 4.1. Por exemplo, com estes gráficos pode-se ter uma idéia da moda da distribuição, que não é possível de ser estimada através do valor mais freqüente da amostra, como já mencionado. Outras conclusões são possíveis, tais como, por exemplo, notar que há muito pouca probabilidade de que o verdadeiro valor do parâmetro σ_c^2 seja um valor em torno de zero, ou maior que oito.

Além das posteriores marginais estimadas, produziram-se medidas resumo que estão mostradas na Tabela 4.1. Nota-se, que o “Gibbs Sampling” produziu estimativas de média e mediana dos parâmetros μ_θ e θ_i , $i=1, \dots, K$, muito próximas entre si. Isto porque a distribuição marginal a posteriori destes parâmetros é simétrica, como se percebe na Figura 4.1. Além disso, estas estimativas foram próximas em relação ao verdadeiro valor simulado.

Param.	Valor Verdadeiro	Média	Mediana	Variância	2,5%	97,5%
μ_θ	10	10,15348	10,12321	4,462695	6,013227	14,444730
θ_1	12,961318	12,796235	12,800798	0,3622292	11,60976	13,97815
θ_2	6,371229	7,428587	7,421113	0,3632217	6,244806	8,623872
θ_3	9,998053	9,166459	9,158802	0,3612992	7,99605	10,35477
θ_4	6,660848	6,866930	6,855522	0,3646160	5,685802	8,083723
θ_5	13,973707	14,622859	14,624788	0,3701037	13,41785	15,82013
σ_θ^2	15	22,61218	14,16722	1549,902	3,997617	89,979414
σ_c^2	3	3,694488	3,583119	0,664745	2,430471	5,600511

Tabela 4.1: Resultados obtidos com o “Gibbs Sampling”

Brown e Prescott (1999) indicam como melhor estimador para os componentes de variância σ_θ^2 e σ_c^2 a mediana. Isto porque, em geral, as respectivas distribuições marginais a posteriori devem ser assimétricas. Para este exemplo específico, esta afirmação é válida, já que os gráficos das posteriores marginais estimadas de σ_θ^2 e σ_c^2 mostram duas distribuições assimétricas. Além disso, analisando-se a Tabela 4.1, verifica-se que a estimativa da mediana da posteriori é mais próxima ao verdadeiro valor que a média da posteriori para os componentes de variância.

O cálculo dos percentis 2,5% e 97,5% fornece uma estimativa para o intervalo de credibilidade central, como já mencionado na Seção 4.1. Para o parâmetro μ_θ , por exemplo,

tem-se que, com 95% de probabilidade a posteriori, o verdadeiro valor deste parâmetro está entre 6,013227 e 14,444730.

Para se comparar as estimativas Bayesianas com as estimativas clássicas, deveria-se ter estimativas das modas das distribuições a posteriori marginais. Isto porque o estimador de máxima verossimilhança, como o que se utilizou no caso dos modelos mistos, é análogo ao estimador da moda da distribuição posteriori marginal do parâmetro considerado. Na Subseção 2.3.3, comentou-se sobre maneiras de estimar a moda através dos valores gerados por “Gibbs Sampling”. Entretanto, no caso de distribuições posterioris marginais aproximadamente simétricas, as estimativas de média, moda e mediana serão aproximadamente iguais, logo, pode-se comparar as estimativas de média e mediana com as estimativas clássicas.

Comparando-se então as estimativas de média e mediana dos parâmetros μ_θ e θ_i , $i = 1, \dots, K$, produzidas pelo “Gibbs Sampling” e as estimativas clássicas apresentadas na Seção 3.4.1, verifica-se que os valores são muito próximos, independentemente do método de estimação de parâmetros de variância e covariância utilizado. Para os componentes de variância, aproximando a moda através da análise visual dos gráficos das posterioris marginais estimadas e comparando-se este valor com as estimativas clássicas, percebe-se que, para σ_ϵ^2 , os valores são muito próximos, independentemente do método de estimação de parâmetros de variância e covariância utilizado. Já para σ_θ^2 , os valores são mais próximos para o método de estimação dos componentes de variância ML. Esperava-se resultados similares já que se definiu prioris não-informativas.

4.4.2. Modelo para Dados Longitudinais – Banco de dados *Weights*

Analisou-se o conjunto de dados longitudinais *Weights*, já introduzido na Subseção 3.4.2, utilizando-se “Gibbs sampling” através do software WinBUGS. Toda a análise foi feita através da formulação do problema segundo o modelo misto apresentado em (3.15). Os programas utilizados estão apresentados no Apêndice C.

Consideraram-se cinco situações em relação às estruturas de covariância assumidas para as matrizes **D** e **R**:

1. \mathbf{D} VC e \mathbf{R} VC;
2. \mathbf{D} UN e \mathbf{R} VC ;
3. \mathbf{D} VC e \mathbf{R} UN;
4. \mathbf{D} VC e \mathbf{R} AR(1);
5. \mathbf{D} UN e \mathbf{R} UN.

Aqui somente serão apresentados os resultados para as situações 1, 4 e 5. Optou-se por isto para evitar a repetitividade e porque, com a programação das três situações que serão apresentadas, a adaptação desta para as duas situações que não serão abordadas não é difícil.

- Análise 1:

Supõe estruturas de covariância VC para as matrizes \mathbf{D} e \mathbf{R} , ou seja, está se supondo que

$$\mathbf{D} = \begin{bmatrix} \sigma_{\alpha}^2 & 0 \\ 0 & \sigma_{\beta}^2 \end{bmatrix} \text{ e } \mathbf{R} = \sigma_{\epsilon}^2 \mathbf{I}_7.$$

A função de verossimilhança para este caso é

$$L(\boldsymbol{\beta}, \mathbf{u}, \mathbf{D}, \mathbf{R} | \mathbf{y}) = \prod_{i=1}^n \prod_{j=1}^g N(\mathbf{X}_{i(j)} \boldsymbol{\beta} + \mathbf{Z}_{i(j)} \mathbf{u}_i; \sigma_{\epsilon}^2) \times \prod_{i=1}^n N(0; \sigma_{\alpha}^2) \times \prod_{i=1}^n N(0; \sigma_{\beta}^2),$$

onde $\mathbf{X}_{i(j)}$ é a j-ésima linha da matriz \mathbf{X}_i e $\mathbf{Z}_{i(j)}$ é a j-ésima linha da matriz \mathbf{Z}_i .

Adotaram-se as seguintes distribuições a priori para os parâmetros desconhecidos do modelo:

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}^*; \mathbf{H}), \text{ onde } (\boldsymbol{\beta}^*)' = [0 \ 0 \ 0 \ 0 \ 0 \ 0] \text{ e } \mathbf{H} = 100000 \mathbf{I}_6;$$

$$\sigma_{\epsilon}^{-2} \sim G(0,005; 0,001);$$

$$\sigma_{\alpha}^{-2} \sim G(0,005; 0,001)$$

e

$$\sigma_{\beta}^{-2} \sim G(0,005; 0,001).$$

Observando-se os valores escolhidos para os hiperparâmetros das prioris é possível notar que prioris não-informativas para todos os parâmetros do modelo estão sendo

assumidas. Ao observar a programação mostrada no Apêndice C, pode-se observar que a matriz \mathbf{H} está especificada como $\mathbf{H} = 0,00001\mathbf{I}_6$. Isto acontece porque o WinBUGS parametriza a distribuição normal através da precisão (recíproco da variância) e não através da variância, como está sendo apresentado.

Gerou-se então uma cadeia de tamanho 15000, com um “burn-in” de 5000 iterações. O tamanho do “burn-in” foi escolhido através de análise visual dos gráficos de trajetória que o WinBUGS fornece, como o mostrado na Figura 4.2.

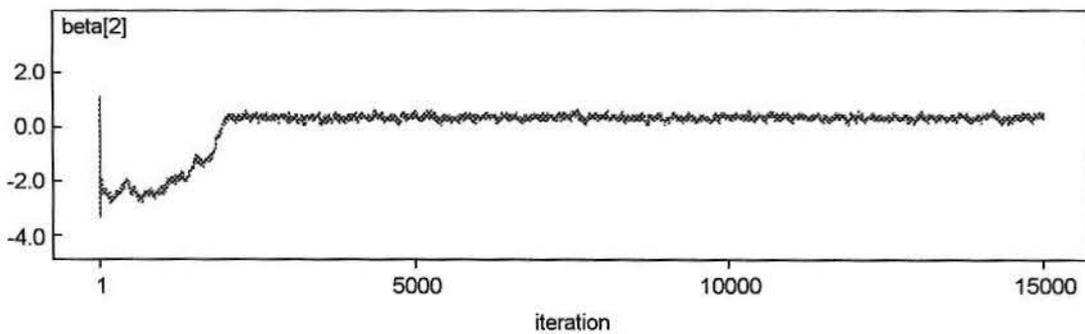


Figura 4.2: Exemplo de gráfico de trajetória disponível no WinBUGS

É interessante sempre produzirmos um gráfico de trajetória excluindo-se o período do “burn-in”, para se conferir se o tamanho do “burn-in” escolhido realmente é suficiente. A Figura 4.3 mostra o mesmo gráfico apresentado na Figura 4.2, apenas excluindo-se o período de “burn-in” de 5000 iterações, que se nota ser aparentemente suficiente para convergência.

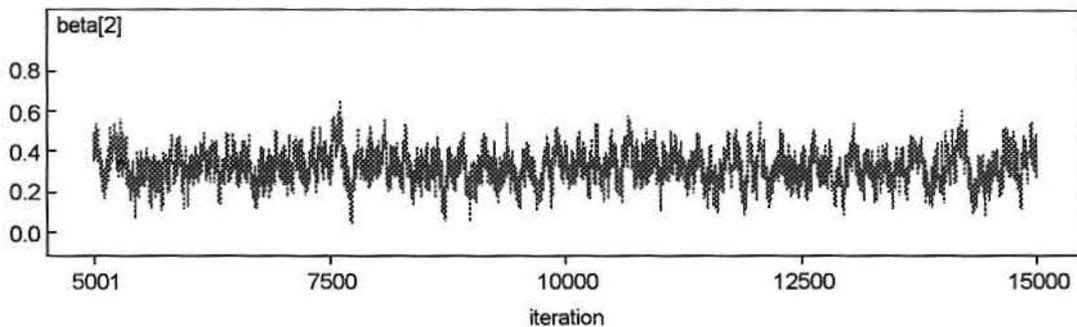


Figura 4.3: Exemplo de gráfico de trajetória excluindo-se o período de “burn-in”

Quanto aos valores iniciais, o software disponibiliza a opção de se gerar valores iniciais aleatórios a partir da distribuição a priori especificada para cada parâmetro. Esta opção foi utilizada para todos os modelos analisados através do WinBUGS.

O WinBUGS estima de maneira não-paramétrica (através de “Kernel Smoothing”) as distribuições a posteriori de cada parâmetro, produzindo gráficos como os da Figura 4.4.

Percebe-se que, com exceção aos componentes de variância dos efeitos aleatórios, as demais posteriores parecem ser aproximadamente simétricas.

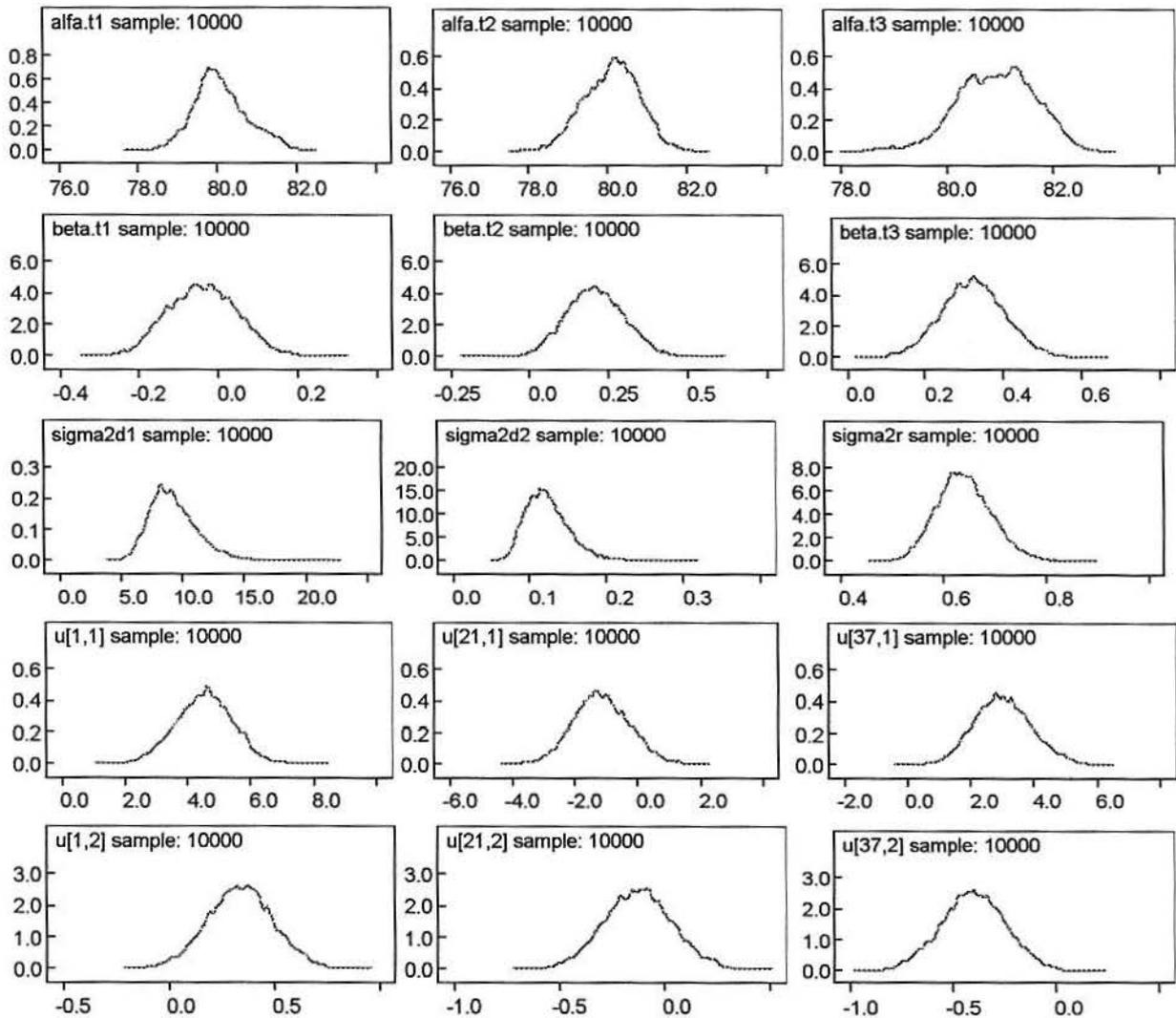


Figura 4.4: Algumas posteriores estimadas não-parametricamente pelo WinBUGS

Na Figura 4.4, “alfa.t1” corresponde ao parâmetro α_{trat1} , “beta.t1” corresponde ao parâmetro β_{trat1} , e para os outros tratamentos a notação é análoga. A notação “u[1,1]” corresponde ao parâmetro $u_{1,1}$, “u[1,2]” corresponde a $u_{1,2}$, e assim por diante. A notação “sigma2d1” corresponde ao parâmetro σ_{α}^2 , “sigma2d2” corresponde a σ_{β}^2 e “sigma2r” corresponde a σ_{ε}^2 .

Medidas resumo foram produzidas e os resultados estão apresentados na Tabela 4.2.

Parâmetro	Média	Mediana	Variância	Percentil 2,5%	Percentil 97,5%
α_{trat1}	80,08	80,01	0,4595	78,84	81,54
α_{trat2}	80,1	80,15	0,5112	78,66	81,43
α_{trat3}	80,94	80,97	0,5633	79,35	82,29
β_{trat1}	-0,03977	-0,04003	0,0075	-0,2008	0,1276
β_{trat2}	0,2078	0,2058	0,0088	0,03386	0,3916
β_{trat3}	0,3235	0,3231	0,0067	0,1636	0,4886
$u_{1,1}$	4,535	4,557	0,8051	2,754	6,258
$u_{21,1}$	-1,142	-1,179	0,8257	-2,873	0,6996
$u_{37,1}$	3,043	3,009	0,8718	1,303	4,971
$u_{1,2}$	0,3395	0,3397	0,0237	0,03239	0,6448
$u_{21,2}$	-0,1239	-0,1234	0,0256	-0,436	0,1667
$u_{37,2}$	-0,3985	-0,3979	0,0249	-0,711	-0,08959
σ_{α}^2	9,287	9,035	3,7752	6,193	13,77
σ_{β}^2	0,1239	0,1205	0,0008	0,07822	0,1908
σ_{ϵ}^2	0,6426	0,6397	0,0029	0,5439	0,7571

Tabela 4.2: Alguns resultados da análise do modelo misto com R VC e D VC através do WinBUGS

As equações de regressão ajustadas para os indivíduos 1, 21 e 37 seriam, respectivamente:

$$\hat{y}_{1j1} = (80,08 + 4,535) + (-0,03977 + 0,3395)tempo_j = 84,615 + 0,29973tempo_j;$$

$$\hat{y}_{21j2} = (80,1 - 1,142) + (0,2078 - 0,1239)tempo_j = 78,958 + 0,0839tempo_j;$$

$$\hat{y}_{37j3} = (80,94 + 3,043) + (0,3235 - 0,3985)tempo_j = 83,983 - 0,075tempo_j.$$

Note que o indivíduo 1 é do tratamento 1, o indivíduo 21 é do tratamento 2 e o indivíduo 37 é do tratamento 3.

As conclusões que podem ser tomadas a partir da análise dos resultados da Tabela 4.2 são idênticas às obtidas pela análise clássica em (3.4.2). Porém, a maneira de interpretar um problema analisado por abordagem Bayesiana é diferente. Para que esta diferença seja evidenciada, se repetirá a interpretação dos resultados.

A partir dos intervalos de credibilidade apresentados para os interceptos de cada tratamento, pode-se concluir que estes não devem ser considerados diferentes, já que os intervalos se sobrepõem. Este resultado parece razoável visto que os interceptos representam a

média de força entre indivíduos de cada tratamento antes do início do estudo. Analisando as estimativas para os coeficientes angulares de cada tratamento conclui-se que, para o tratamento 1 (CONT), a tendência linear não deve existir, já que o intervalo de credibilidade contém o valor zero. Isto sugere que a força dos indivíduos se mantenha constante ao longo do tempo neste tratamento. Analogamente, conclui-se que a regressão linear deve existir para os tratamentos 2 e 3 (RI e WI).

O coeficiente angular do tratamento 3 (WI) pode ser considerado diferente do coeficiente angular do tratamento 1 (CONT), o que sugere que este tratamento seja eficiente em aumentar a força do sujeito à medida que ele se exercita. Já no caso do tratamento 2 (RI), seu coeficiente angular não deve ser considerado diferente do coeficiente angular do tratamento 1 (RI), o que sugere que ele não seja eficiente em proporcionar um aumento de força ao indivíduo que pratica este programa.

Pelos gráficos das posteriores estimadas mostrados na Figura 4.4 e pela proximidade dos valores estimados da média e mediana apresentados na Tabela 4.2 para todos os parâmetros, pode-se assumir que as distribuições a posteriori marginais de todos os parâmetros são aproximadamente simétricas. Por isto, pode-se comparar os valores de média e mediana da Tabela 4.2 com os resultados clássicos apresentados na Tabela 3.3. Fazendo isso, percebe-se que os resultados são razoavelmente semelhantes, como era esperado já que se utilizaram prioris não-informativas.

- Análise 2:

Supõe estruturas de covariância VC para a matriz \mathbf{D} e estrutura AR(1) para a matriz \mathbf{R} , ou seja:

$$\mathbf{D} = \begin{bmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{bmatrix}; \quad \mathbf{R} = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^6 \\ \rho & 1 & \rho & \dots & \rho^5 \\ \rho^2 & \rho & 1 & \dots & \rho^4 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^6 & \rho^5 & \rho^4 & \dots & 1 \end{bmatrix}.$$

A função de verossimilhança para este caso é

$$L(\boldsymbol{\beta}, \mathbf{u}, \mathbf{D}, \mathbf{R} | \mathbf{y}) = \prod_{i=1}^n N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i; \mathbf{R}) \times \prod_{i=1}^n N(0; \sigma_\alpha^2) \times \prod_{i=1}^n N(0; \sigma_\beta^2).$$

As distribuições a priori determinadas aos parâmetros foram as mesmas adotadas na análise anterior, adicionando-se a priori para o parâmetro de correlação ρ . Assumiu-se que $\rho \sim U(-1;1)$, ou seja, uma priori não-informativa.

Gerou-se uma cadeia de tamanho 11000, com um “burn-in” de 1000 iterações. Os valores iniciais foram especificados através da opção do WinBUGS de gerar números aleatórios das distribuições a priori.

Os resultados obtidos nesta análise estão sintetizados na Tabela 4.3:

Parâmetro	Média	Mediana	Variância	Percentil 2,5%	Percentil 97,5%
α_{trat1}	79,86	79,86	0,5820	78,35	81,36
α_{trat2}	79,62	79,62	0,7415	77,92	81,32
α_{trat3}	80,76	80,76	0,5294	79,35	82,19
β_{trat1}	-0,02767	-0,02727	0,00757	-0,1999	0,1435
β_{trat2}	0,2371	0,2371	0,0095	0,04736	0,4304
β_{trat3}	0,3374	0,338	0,0073	0,1728	0,5051
$u_{1,1}$	2,654	2,781	5,3453	-0,505	6,777
$u_{21,1}$	-0,5294	-0,2725	1,5252	-3,248	1,764
$u_{37,1}$	1,019	0,7462	1,9404	-1,31	3,929
$u_{1,2}$	0,1551	0,06897	0,0276	-0,1231	0,5188
$u_{21,2}$	-0,02582	-0,01406	0,0187	-0,3354	0,2424
$u_{37,2}$	-0,05133	-0,02517	0,0206	-0,4115	0,1966
σ_{α}^2	4,425	4,754	13,381	0,002752	11,07
σ_{β}^2	0,02825	0,01725	0,0009	0,000724	0,09922
ρ	0,8463	0,8979	0,0139	0,5766	0,9621
σ_{ϵ}^2	5,908	5,083	15,476	1,155	13,57

Tabela 4.3: Alguns resultados da análise do modelo misto com R AR(1) e D VC através do WinBUGS

As equações de regressão ajustadas para os indivíduos 1, 21 e 37 seriam, respectivamente:

$$\hat{y}_{1j1} = (79,86 + 2,654) + (-0,02767 + 0,1551)tempo_j = 82,514 + 0,12743tempo_j;$$

$$\hat{y}_{21j2} = (79,62 - 0,5294) + (0,2371 - 0,02582)tempo_j = 79,0906 + 0,21128tempo_j;$$

$$\hat{y}_{37j3} = (80,76 + 1,019) + (0,3374 - 0,05133)tempo_j = 81,779 + 0,28607tempo_j.$$

Ao se analisar os resultados da Tabela 4.3, percebe-se, mesmo alterando-se a estrutura de covariância da matriz \mathbf{R} , as conclusões continuam as mesmas da Análise 1. Os valores estimados para a média e a mediana apresentados na Tabela 4.3 somente são próximos para os efeitos fixos. Assim sendo, somente para estes parâmetros se fez comparação dos resultados do “Gibbs Sampling” com os resultados clássicos mostrados na Tabela 3.4. Fazendo-se isto, percebe-se bastante similaridade nos valores encontrados. Isto era esperado novamente, pois as priors especificadas eram não-informativas.

- Análise 3:

Supõe estruturas UN para ambas as matrizes \mathbf{D} e \mathbf{R} , ou seja:

$$\mathbf{D} = \begin{bmatrix} \sigma_{\alpha}^2 & \sigma_{\alpha\beta} \\ \sigma_{\alpha\beta} & \sigma_{\beta}^2 \end{bmatrix} \text{ e } \mathbf{R} = \begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \cdots & \sigma_{71} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \cdots & \sigma_{72} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \cdots & \sigma_{73} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{71} & \sigma_{72} & \sigma_{73} & \cdots & \sigma_7^2 \end{bmatrix}.$$

A função de verossimilhança para este caso é

$$L(\boldsymbol{\beta}, \mathbf{u}, \mathbf{D}, \mathbf{R} | \mathbf{y}) = \prod_{i=1}^n N(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i; \mathbf{R}) \times \prod_{i=1}^n N(\mathbf{0}; \mathbf{D}).$$

Adotaram-se as seguintes distribuições a priori para os parâmetros desconhecidos do modelo:

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}^*; \mathbf{H}), \text{ onde } (\boldsymbol{\beta}^*)' = [0 \ 0 \ 0 \ 0 \ 0 \ 0] \text{ e } \mathbf{H} = 100000\mathbf{I}_6;$$

$$\mathbf{D} \sim \text{Wishart}(2, \mathbf{I}_2); \mathbf{R} \sim \text{Wishart}(7, \mathbf{I}_7).$$

Gerou-se uma cadeia de tamanho 11000, com um “burn-in” de 1000 iterações. Os valores iniciais foram especificados da mesma maneira que nas duas análises anteriores.

As estimativas produzidas pelo WinBUGS estão apresentadas na Tabela 4.4. Como ficaria muito extenso colocar na Tabela 4.4 todos os 28 parâmetros da matriz \mathbf{R} , somente será apresentada a estimativa da mediana destes componentes:

$$\mathbf{R}_{Med} = \begin{bmatrix} 0,3908 & 0,1069 & -0,05674 & -0,1563 & -0,1305 & -0,00595 & 0,192 \\ 0,1069 & 0,7022 & 0,179 & -0,05738 & -0,01994 & -0,04981 & -0,1248 \\ -0,05674 & 0,179 & 0,6602 & 0,3433 & 0,2717 & -0,03082 & -0,0766 \\ -0,1563 & -0,05738 & 0,3433 & 0,9174 & 0,4664 & -0,05579 & -0,1431 \\ -0,1305 & -0,01994 & 0,2717 & 0,4664 & 0,852 & 0,1681 & -0,1705 \\ -0,00595 & -0,04981 & -0,03082 & -0,05579 & 0,1681 & 0,6053 & 0,05069 \\ 0,192 & -0,1248 & -0,0766 & -0,1431 & -0,1705 & 0,05069 & 0,5925 \end{bmatrix}$$

Parâmetro	Média	Mediana	Variância	Percentil 2,5%	Percentil 97,5%
α_{trat1}	79,64	79,7	0,5498	78,03	80,88
α_{trat2}	79,55	79,56	0,4204	78,22	80,78
α_{trat3}	80,99	81,01	0,3888	79,7	82,2
β_{trat1}	-0,03512	-0,03413	0,0075	-0,2061	0,1326
β_{trat2}	0,2472	0,25	0,0096	0,04855	0,434
β_{trat3}	0,3504	0,351	0,0089	0,1695	0,533
$u_{1,1}$	4,83	4,787	0,8782	3,135	6,818
$u_{21,1}$	-0,8617	-0,8765	0,7223	-2,51	0,8417
$u_{37,1}$	2,75	2,76	0,7377	0,9814	4,418
$u_{1,2}$	0,3468	0,3489	0,0199	0,06789	0,6214
$u_{21,2}$	-0,113	-0,1157	0,0225	-0,4076	0,1829
$u_{37,2}$	-0,3391	-0,3374	0,0205	-0,6152	-0,06228
σ_{α}^2	9,296	9,057	3,7326	6,213	13,81
σ_{β}^2	0,1473	0,1436	0,0010	0,09632	0,221
$\sigma_{\alpha\beta}$	-0,2039	-0,1945	0,0319	-0,5811	0,1242

Tabela 4.4: Alguns resultados da análise do modelo misto com R UN e D UN através do WinBUGS

As conclusões obtidas por este modelo são idênticas às obtidas na Análise 1.

As equações de regressão ajustadas para os indivíduos 1, 21 e 37 seriam, respectivamente:

$$\hat{y}_{1j1} = (79,64 + 4,83) + (-0,03512 + 0,3468)tempo_j = 84,47 + 0,31168tempo_j;$$

$$\hat{y}_{21j2} = (79,55 - 0,8617) + (0,2472 - 0,113)tempo_j = 78,6883 + 0,1342tempo_j;$$

$$\hat{y}_{37j3} = (80,99 + 2,75) + (0,3504 - 0,3391)tempo_j = 83,74 + 0,0113tempo_j.$$

Para se comparar as retas ajustadas nas três análises, produziram-se os gráficos mostrados na Figura 4.5.

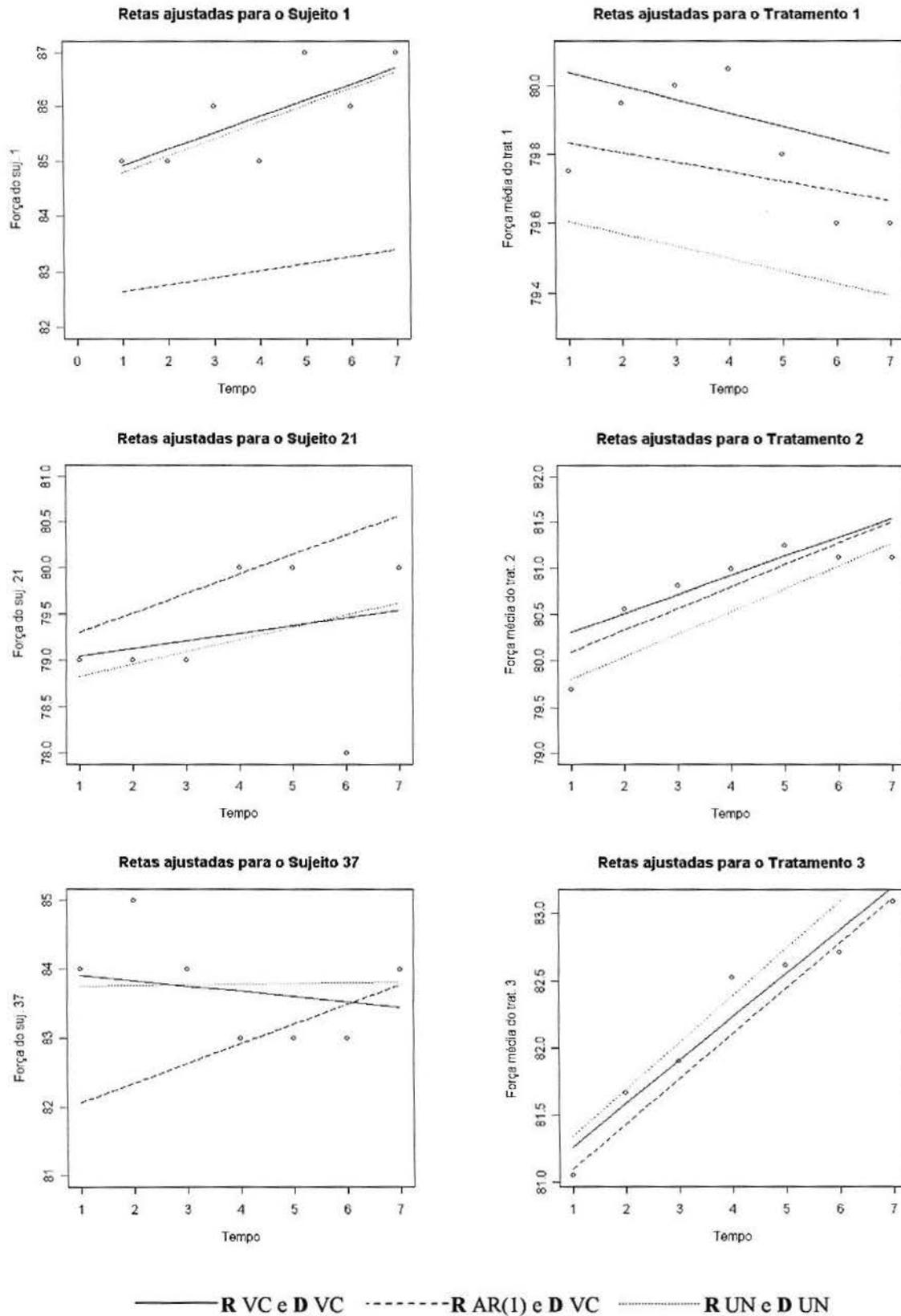


Figura 4.5: Retas ajustadas aos sujeitos 1, 21 e 37 e aos tratamentos 1, 2 e 3, supondo-se diferentes estruturas para as matrizes R e D. Nos gráficos para os sujeitos, pontos representam a força medida naquele sujeito a cada tempo. Nos gráficos de cada tratamento, pontos representam a média das observações dos indivíduos daquele tratamento a cada tempo

Somente pela análise dos gráficos da Figura 4.5, não se consegue decidir qual modelo ajusta melhores retas aos indivíduos e aos tratamentos. Assim como na inferência clássica, a inferência Bayesiana também possui técnicas para se comparar modelos e decidir qual o melhor. Como este tópico não faz parte dos objetivos deste trabalho, ele não será abordado.

4.5. Estudo de Simulação de Dados Longitudinais

Para se verificar algumas propriedades dos estimadores produzidos pelo “Gibbs Sampling” procedeu-se um estudo de simulação no software R. A programação utilizada está apresentada no Apêndice D.

Simularam-se conjuntos de dados semelhantes com o *Weights*, com 3 tratamentos, porém com 5 tempos e 27 sujeitos (8 no tratamento 1, 10 no tratamento 2 e 9 no tratamento 3). Especificaram-se os seguintes valores para os efeitos fixos: $\alpha_0 = 1$; $\alpha_1 = 5$; $\alpha_2 = 3$; $\beta_0 = 2$; $\beta_1 = -4$ e $\beta_2 = -6$. Ou seja, especificou-se $\alpha_{trat.1} = 6$, $\alpha_{trat.2} = 4$, $\alpha_{trat.3} = 1$, $\beta_{trat.1} = -2$, $\beta_{trat.2} = -4$ e $\beta_{trat.3} = 2$. A partir destes valores e dos especificados para os componentes de variância, geraram-se os efeitos aleatórios (interceptos e coeficientes angulares dos sujeitos), segundo o modelo hierárquico apresentado em (3.4.2).

Os valores simulados para os interceptos e coeficientes angulares destes sujeitos originalmente foram: $\alpha_{suj.1} = 3,894378$; $\alpha_{suj.9} = 5,152967$; $\alpha_{suj.19} = 0,4110964$; $\beta_{suj.1} = -4,920952$; $\beta_{suj.9} = -5,477128$ e $\beta_{suj.19} = 1,771896$.

Em um primeiro momento, simularam-se dados especificando estruturas de covariância VC para ambas as matrizes **D** e **R**, com $\sigma_\alpha^2 = 1$, $\sigma_\beta^2 = 4$ e $\sigma_\varepsilon^2 = 9$, e utilizaram-se as distribuições condicionais completas do modelo hierárquico apresentadas na Seção 4.3 para implementar o “Gibbs Sampling”. Geraram-se então cadeias de 11000 iterações, com “burn-in” de 1000. Foram gravados os valores de média, mediana e moda (estimada não-parametricamente) para cada um dos parâmetros desconhecidos do modelo. O processo de gerar um conjunto de dados e aplicar o “Gibbs Sampling” foi replicado 500 vezes, para que se pudesse analisar a distribuição dos estimadores média, moda e mediana.

Após isto, simularam-se dados especificando estrutura de covariância VC para a matriz **D** e UN para a matriz **R**, e utilizaram-se as mesmas condicionais completas anteriores, que consideram uma estrutura VC para **R**. Isto foi feito para se verificar como ficaria o

comportamento dos estimadores quando a suposição de erros independentes não é válida. O restante do processo permaneceu inalterado. A matriz \mathbf{R} definida para a simulação está apresentada abaixo:

$$\mathbf{R} = \begin{bmatrix} 3 & 0,95 & 0,9 & 0,8 & 0,7 \\ 0,95 & 1 & 0,9 & 0,8 & 0,6 \\ 0,9 & 0,9 & 2 & 0,8 & 0,65 \\ 0,8 & 0,8 & 0,8 & 1,5 & 0,75 \\ 0,7 & 0,6 & 0,65 & 0,75 & 0,5 \end{bmatrix}.$$

Após as 500 replicações terem sido concluídas, fez-se Box-plots das estimativas de cada parâmetro para cada estimador. Assim, pode-se estudar vício e variabilidade dos estimadores. Alguns resultados para o primeiro caso simulado estão nas Figuras 4.6, 4.7 e 4.8. A linha pontilhada horizontal representa o valor verdadeiro.

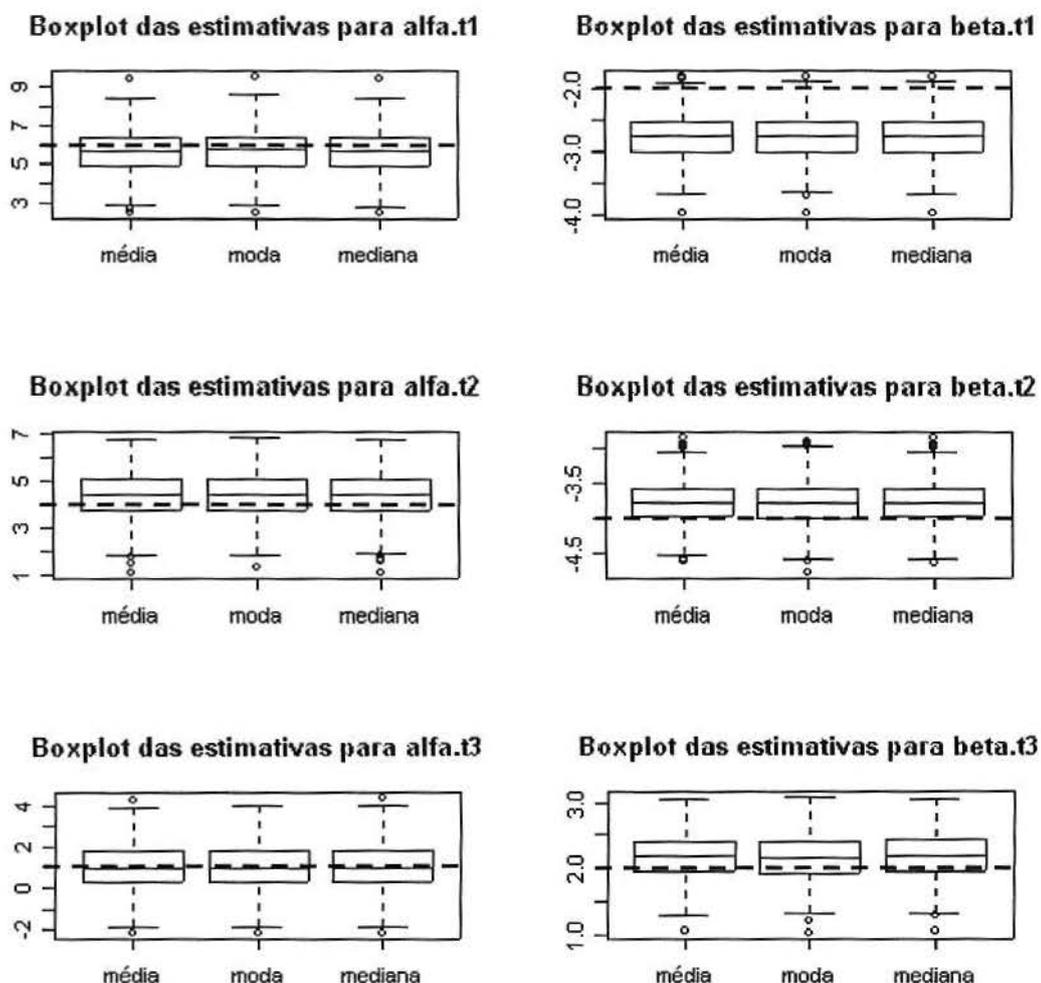


Figura 4.6: Box-plot dos interceptos e coeficientes angulares de cada tratamento na situação de suposições atendidas

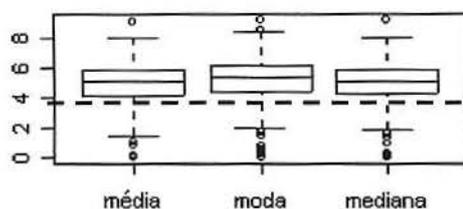
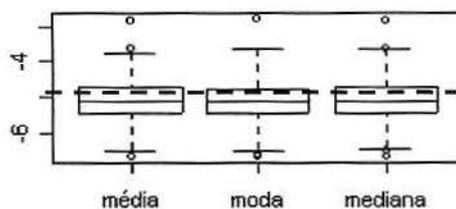
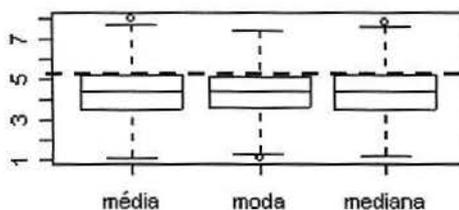
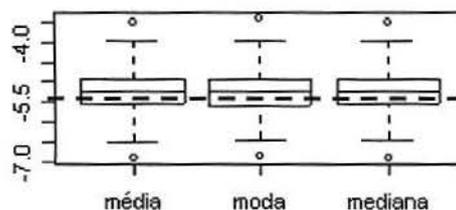
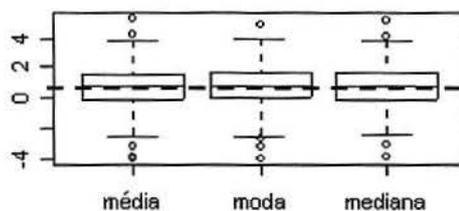
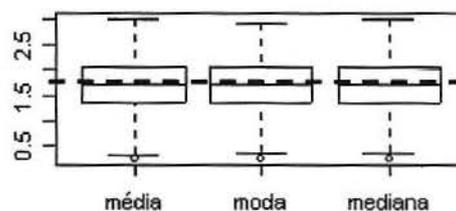
Boxplot das estimativas para alfa.suj.1**Boxplot das estimativas para beta.suj.1****Boxplot das estimativas para alfa.suj.9****Boxplot das estimativas para beta.suj.9****Boxplot das estimativas para alfa.suj.19****Boxplot das estimativas para beta.suj.19**

Figura 4.7: Box-plot dos interceptos e coeficientes angulares dos sujeitos 1, 9 e 19 na situação de suposições atendidas

Novamente, só estamos apresentando os resultados para um sujeito de cada tratamento, por questão de resumo.

Percebe-se que, para os efeitos fixos e aleatórios, os estimadores média, moda (estimada não-parametricamente) e mediana têm propriedades muito parecidas em termos de vício e variabilidade. Sendo assim, somente por estes resultados, não podemos concluir se há algum estimador melhor que outro, em relação aos efeitos fixos e aleatórios.

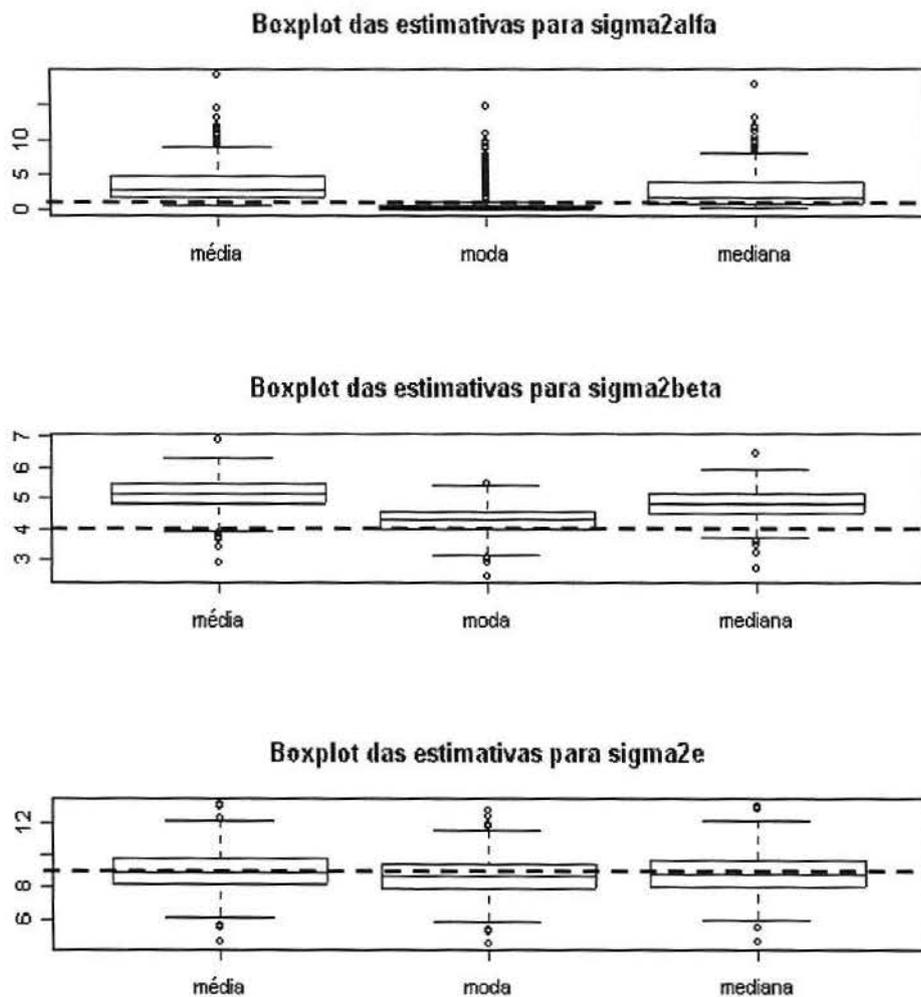


Figura 4.8: Box-plot dos componentes de variância na situação de suposições atendidas

Para os componentes de variância, a moda se apresentou como estimador de menor vício, comparando-se com os valores originalmente definidos para os componentes de variância, e também com menor variabilidade. Estas propriedades foram mais acentuadas nos casos dos componentes de variância da matriz **D**.

A seguir, será apresentada a análise feita a partir da situação onde simulou-se um conjunto de dados com estrutura de covariância UN para a matriz **R** e procedeu-se o “Gibbs Sampling” ignorando esta suposição. As Figuras 4.9, 4.10 e 4.11 mostram alguns resultados. A linha pontilhada horizontal representa o valor verdadeiro.

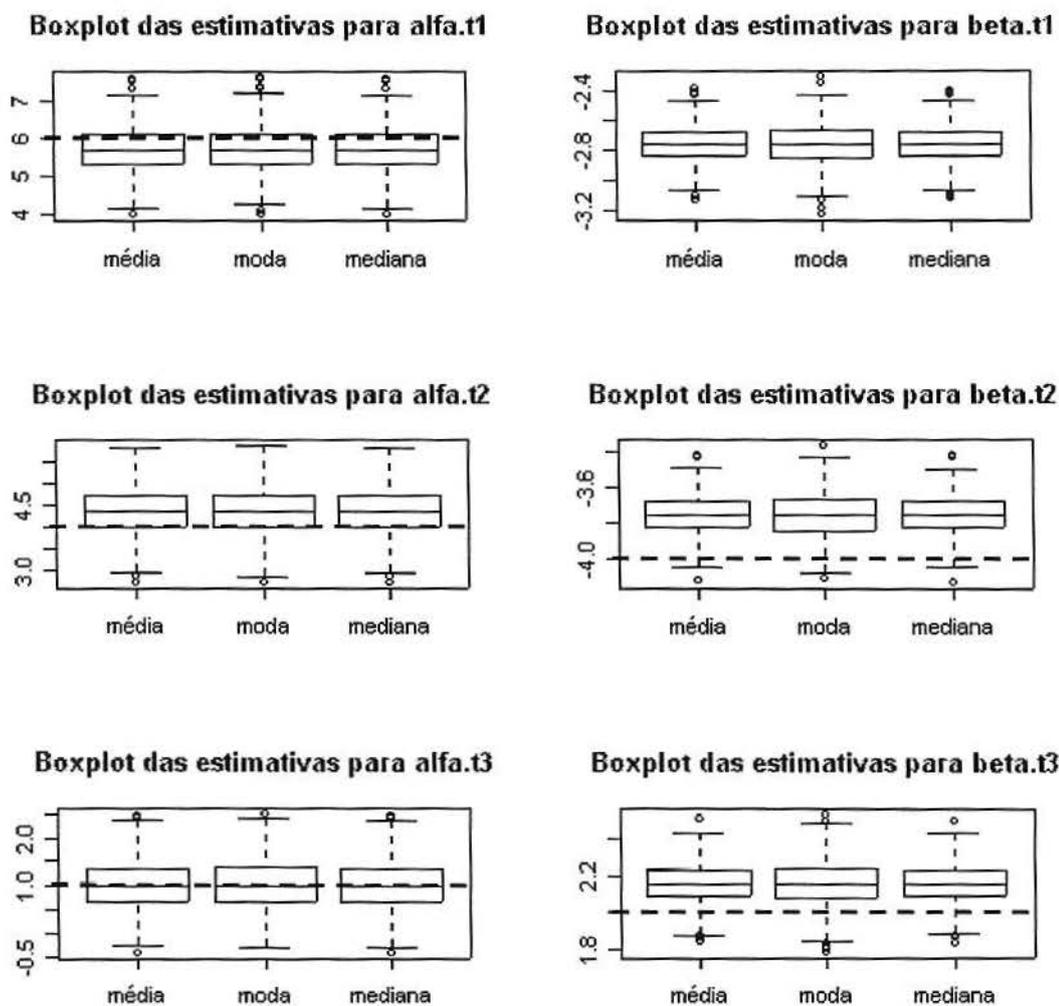


Figura 4.9: Box-plot dos interceptos e coeficientes angulares de cada tratamento na situação de suposição de erros independentes não atendida

Para o coeficiente angular do tratamento 1, não foi possível plotar o valor verdadeiro pois ele está fora da escala do gráfico.

Novamente, percebe-se a grande semelhança na distribuição dos estimadores média, moda e mediana para os efeitos fixos e aleatórios. Sendo assim, somente por estes resultados, não se pode concluir se há algum estimador melhor que outro, em relação aos efeitos fixos e aleatórios.

É possível notar, entretanto, que o vício dos estimadores aumentou ao se supor uma estrutura VC para a matriz \mathbf{R} , quando a verdadeira estrutura era UN.

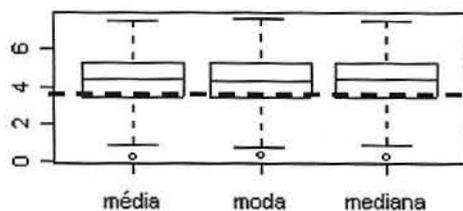
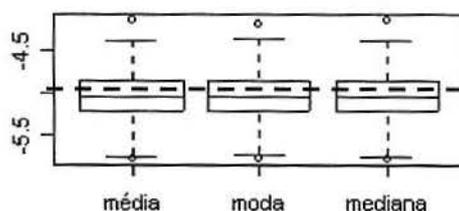
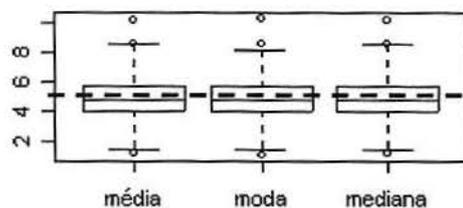
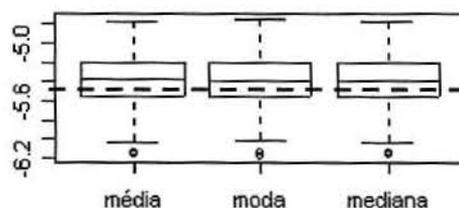
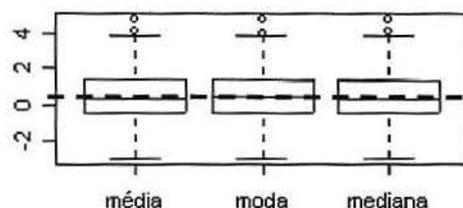
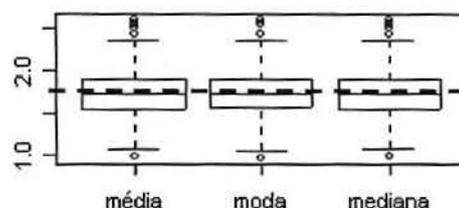
Boxplot das estimativas para alfa.suj.1**Boxplot das estimativas para beta.suj.1****Boxplot das estimativas para alfa.suj.9****Boxplot das estimativas para beta.suj.9****Boxplot das estimativas para alfa.suj.19****Boxplot das estimativas para beta.suj.19**

Figura 4.10: Box-plot dos interceptos e coeficientes angulares para os sujeitos 1, 9 e 19 na suposição de erros independentes não atendida

Também para os interceptos e coeficientes angulares dos sujeitos, evidenciou-se a grande semelhança das distribuições dos estimadores média, moda e mediana.

Em oposição ao que ocorreu para os efeitos fixos, o vício dos estimadores média, moda e mediana para os efeitos aleatórios não aumentou ao se supor uma estrutura errada para a matriz \mathbf{R} .

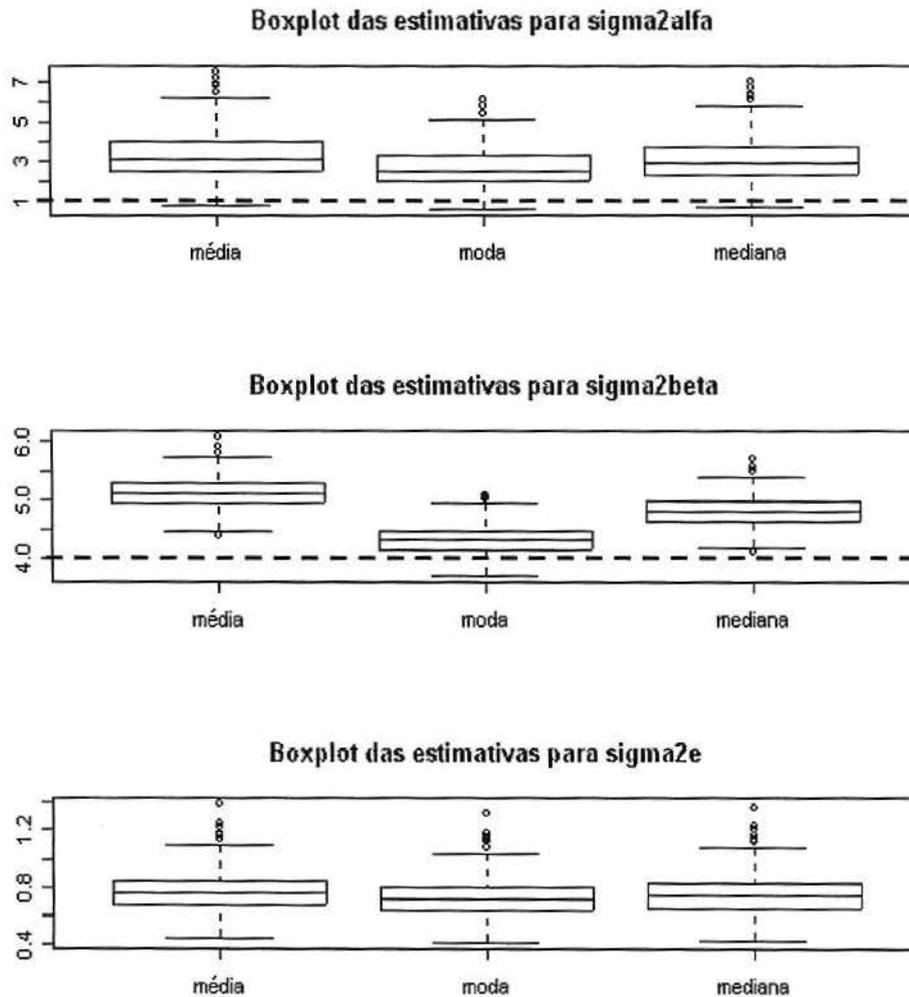


Figura 4.11: Box-plot dos componentes de variância na situação de suposição de erros independentes não atendida

Para os componentes de variância σ_{α}^2 e σ_{β}^2 , a moda se apresentou como estimador de menor vício, comparando-se com os valores originalmente definidos para os componentes de variância, e também com menor variabilidade. Entretanto, o vício aumentou em relação a situação onde a suposição quanto a estrutura da matriz \mathbf{R} estava atendida. Quanto ao componente de variância σ_{ϵ}^2 , analisando-se a diagonal da matriz \mathbf{R} originalmente definida, percebe-se que os três estimadores subestimaram a variabilidade dentro dos indivíduos.

Com este estudo conclui-se que, ao suporem-se erros independentes, é importante avaliar se esta suposição não é muito restritiva, pois aparentemente o não atendimento desta aumenta o vício dos estimadores dos efeitos fixos, e dos componentes de variância. Além disto, conclui-se que a moda estimada não-parametricamente é um estimador com boas propriedades no caso do modelo estudado.

5. CONSIDERAÇÕES FINAIS

Um dos objetivos deste trabalho foi introduzir os métodos MCMC, destacando sua eficiência para resolver problemas de modelagem Bayesianos. Enfoque maior foi dado ao algoritmo de “Gibbs Sampling” e percebeu-se que este algoritmo é simples e muito versátil para se analisar vários problemas estatísticos, em especial, modelos mistos.

A abordagem clássica para os modelos mistos também foi revisada, podendo-se notar que este modelo é muito útil para problemas estatísticos de modelagens mais sofisticadas, como o caso dos dados longitudinais. Além disso, ele unifica a análise de vários modelos segundo uma única abordagem.

Conceitos Bayesianos foram apresentados e aplicados ao caso de modelo misto. Foi visto que o grande empecilho da abordagem Bayesiana para os modelos mistos é que a distribuição a posteriori não pode ser derivada analiticamente. Entretanto, este empecilho é facilmente superado com o uso do “Gibbs Sampling”, que se mostrou um algoritmo sem grandes dificuldades de ser aplicado tanto nos softwares R como WinBUGS. Os resultados encontrados, utilizando prioris não-informativas, foram comparados com os obtidos por métodos clássicos, mostrando-se, em geral, muito parecidos. Percebe-se, porém, que a abordagem Bayesiana para modelos mistos apresenta vantagens sobre a abordagem clássica, já que as únicas suposições que foram feitas, além das usuais do modelo, eram em respeito às distribuições a priori. Já na abordagem clássica, além das suposições usuais do modelo misto, precisa-se supor que a matriz de variâncias é conhecida para estimar os efeitos fixos e aleatórios. Também é necessário supor uma distribuição aproximada para a estatística de teste de hipóteses dos efeitos fixos e aleatórios e basear-se também nesta suposição para poder construir intervalos de confiança. Comparando-se estas suposições, percebe-se que a suposição de uma distribuição a priori é uma suposição bastante realista, pois é feita com base em toda a informação existente sobre o parâmetro e totalmente flexível à escolha do pesquisador. Já a suposição de que se conhecem as variâncias para se poderem estimar os efeitos fixos e aleatórios raramente é atendida.

Ao longo do trabalho foi notado que o assunto é um tema muito rico para ser pesquisado. Por restrições de tempo, muitos tópicos interessantes de serem abordados não puderam ser contemplados. Um dos assuntos não abordados neste trabalho foi o uso de prioris informativas e seu impacto sobre as estimativas produzidas pelo “Gibbs Sampling”. Outro

tópico seria adicionar no estudo de simulação procedido as estimativas clássicas dos parâmetros, comparando-se os resultados dos três métodos de estimação de componentes de variância clássicos (REML, ML e ANOVA) com os resultados do “Gibbs Sampling”. Neste mesmo sentido, poderia-se tomar exemplos de dados em que as estimativas clássicas dos componentes de variância fossem problemáticas (variâncias negativas produzidas pelo método da ANOVA, que geralmente fazem com que os métodos REML e ML trunquem sua estimativa no valor zero) e verificar qual o comportamento do “Gibbs Sampling” frente a estas situações.

Em nenhum momento deste trabalho foram feitas análises de adequabilidade dos modelos ajustados, tanto na parte clássica como na parte Bayesiana. Também não foram estudados métodos formais de comparar os modelos ajustados, em busca do melhor modelo. Pelo que se pôde perceber na revisão de literatura, há bastante material disponível sobre este assunto.

Outro tópico que poderia ser pesquisado seria como fazer predição de novas observações e de observações faltantes no contexto Bayesiano. Sabe-se que há como fazer isto utilizando-se também o “Gibbs Sampling” e que o WinBUGS tem alguns recursos implementados neste sentido.

Há alguns trabalhos que enfocam maneiras de se parametrizar o modelo misto de tal forma que torne o “Gibbs Sampling” mais eficiente, ou seja, de convergência mais rápida. Duas referências que tratam deste assunto são Gelfand et al. (1995) e o Capítulo 6 de Gilks et al. (1996).

Enfim, espera-se que este trabalho seja considerado uma referência para aqueles que desejam obter uma introdução aos métodos MCMC, em especial ao “Gibbs Sampling”, e também à aplicação das técnicas Bayesianas aos modelos mistos. Espera-se que se tenha definido de maneira satisfatória os conceitos teóricos envolvidos e que o trabalho sirva como base para aqueles interessados em analisar problemas práticos similares aos analisados nesta monografia.

REFERÊNCIAS BIBLIOGRÁFICAS

- Best, N., Cowles, M. e Vines, S. (1995) *CODA: Convergence Diagnosis and Output Analysis Software for Gibbs sampling output, Version 0.3*. Relatório técnico, Biostatistics Unit-MRC, Cambridge, Inglaterra.
- Brooks, S. e Roberts, G. (1998) Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing*, 8, 319-335.
- Brown, H. e Prescott, R. (1999). *Applied Mixed Models in Medicine*. Chichester: John Wiley & Sons, Inc.
- Casella, G. e George, E. (1992). Explaining the Gibbs Sampler. *The American Statistician*, 46, 167-174.
- Cowles, M. e Carlin, B. (1994) Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91, 883-904.
- Davidian, M. e Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. São Francisco: Chapman & Hall.
- Gamerman, D. (1997). *Markov Chain Monte Carlo*. Londres: Chapman & Hall.
- Gelfand, A. E. e Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.
- Gelfand, A., Sahu, S. e Carlin, B. (1995) Efficient parametrizations for normal linear mixed models. *Biometrika*, 82, 479-488.
- Gelman, A., Carlin, J., Stern, H., Rubin, D. (1997). *Bayesian Data Analysis*. Boca Raton: Chapman & Hall.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.
- Gilks, W., Wang, C., Yvonnet, B. e Coursaget, P. (1993) Random-effects models for longitudinal data using Gibbs sampling. *Biometrics*, 49, 441-453.
- Gilks, W.R., Richardson, S. e Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Londres: Chapman & Hall.
- Hastings, W. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.
- Henderson, C. (1984) *Applications of Linear Models in Animal Breeding*, University of Guelph, Ontario.
- Laird, N. e Ware, J. (1982) Random-effects models for longitudinal data. *Biometrics*, 38, 963-974.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. e Teller, E. (1953) Equation of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, 1087-1091.

Ross, S. (1996) *Stochastic Processes*. Nova York: John Wiley & Sons, Inc.

SAS Institute Inc. (2002) *SAS Online Doc v.9*. SAS Institute Inc., Cary, EUA.

Satterthwaite, F. (1946) An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, 2, 110-114.

Searle, S., Casella, G., McCulloch, C. (1992). *Variance Components*. Nova York: John Wiley & Sons, Inc.

Spiegelhalter, D., Thomas, A., Best, N. e Gilks, W. (1995) *Bayesian Inference Using Gibbs Sampling: Version 0.5*. Relatório técnico, Biostatistics Unit-MRC, Cambridge, Inglaterra.

Tierney, L. (1994) Markov Chains for exploring posterior distributions (with discussion). *Annals of Statistics*, 22, 1701-1762.

APÊNDICE A – Programação em R para o Exemplo 2.3.4

```

s <- 5000 # Tamanho da cadeia

mix <- 3
sigmax <- 1
miy <- 0
sigmay <- 5
ro <- 0.8
aux1 <- ro*sigmax/sigmay
aux2 <- sigmax*sqrt(1-ro^2)
aux3 <- ro*sigmay/sigmax
aux4 <- sigmay*sqrt(1-ro^2)

y <- rep(0,s)
x <- rep(0,s)

# Valores iniciais das cadeias
y[1] <- 30
x[1] <- -5

t <- seq(1,s,1)

for (i in 1:(s-1)) {
  x[i+1] <- rnorm( 1, (mix+aux1*(y[i]-miy)), aux2 )
  y[i+1] <- rnorm( 1, (miy+aux3*(x[i+1]-mix)), aux4 )
}

plot(t,x,type="l")
lines(t, rep(mix+2*sigmax, s), lty="dotted")
lines(t, rep(mix-2*sigmax, s), lty="dotted")
title(main="Trajetória da simulação de x")

windows()
plot(t, y, type="l")
lines(t, rep(miy+2*sigmay, s), lty="dotted")
lines(t, rep(miy-2*sigmay, s), lty="dotted")
title(main="Trajetória da simulação de y")

burn <- 500 # Valores a serem desprezados

# Estimando a média
exest <- mean(x[(burn+1):s])
eyest <- mean(y[(burn+1):s])

# Estimando a variância
varxest <- var(x[(burn+1):s])
varyest <- var(y[(burn+1):s])

# Estimando a correlação
roest <- cor(x[(burn+1):s], y[(burn+1):s])

# Densidade verdadeira de x
windows()
plot(c(0,6), c(0, 0.4), type="n")
curve(exp(-(x-3)^2/2)/(sqrt(2*pi)), from=0, to=6, xlab="x", ylab="f(x)")
title(main="Marginais verdadeira e estimada de x")

# Estimando a densidade marginal de x
pxdadoy <- rep(0,(s-burn))
for(i in 1:(s-burn)) {
  for (j in 1:(s-burn)) {
    k <- dnorm(x[i+burn], (mix+aux1*(y[j+burn]-miy)), aux2 )/(s-burn)
    pxdadoy[i] <- pxdadoy[i] + k
  }
}
points(x[(burn+1):s], pxdadoy)

```

```
# Densidade verdadeira de y
windows()
plot(c(-15,15), c(0, 0.9), type="n")
curve(exp(-(x^2)/50)/(sqrt(2*pi)*5), from=-15, to=15, xlab="y", ylab="f(y)")
title(main="Marginais verdadeira e estimada de y")

# Estimando a densidade marginal de y
pydadox <- rep(0, (s-burn))
for(i in 1:(s-burn)) {
  for (j in 1:(s-burn)) {
    k <- dnorm(y[i+burn], (miy+aux3*(x[j+burn]-mix)), aux4 )/(s-burn)
    pydadox[i] <- pydadox[i] + k
  }
}
points(y[(burn+1):s], pydadox)
```

APÊNDICE B – Programação em R para os Exemplos 3.4.1 e 4.4.1

```

J <- 10
K <- 5

y <- matrix(rep(0,J*K), nrow=K, ncol=J)

miteta <- 10
sigma2teta <- 15
sigma2e <- 3
teta <- rnorm(K,miteta, sqrt(sigma2teta))

# Simulando os dados
for( i in 1:K) {
  y[i,] <- rnorm(J, teta[i], sqrt(sigma2e)) }

# Parâmetros das prioris
mi0 <- 0
sigma20 <- 100000
a1 <- 0
b1 <- 0
a2 <- 0
b2 <- 0

# Gibbs-sampling
m <- 10000 # Tamanho da cadeia
burn <- 1000 # Valores a serem desprezados

teta.sim <- matrix(rep(0, K*m), nrow=m, ncol=K)
sigma2e.sim <- rep(0, m)
sigma2teta.sim <- rep(0, m)
miteta.sim <- rep(0, m)
sumsq <- matrix(rep(0, K*m), nrow=m, ncol=K)

# Valores iniciais - para miteta e teta's os valores iniciais são iguais a zero
sigma2e.sim[1] = 0.1
sigma2teta.sim[1] = 0.1

for (i in 1:(m-1)) {
  miteta.sim[i+1] <- rnorm(1, (sigma2teta.sim[i]*mi0 +
sigma20*sum(teta.sim[i,]))/(sigma2teta.sim[i]+K*sigma20),
sqrt(sigma2teta.sim[i]*sigma20/(sigma2teta.sim[i]+K*sigma20)))
  for(j in 1:K) {
    teta.sim[i+1,j] <- rnorm(1,
(J*sigma2teta.sim[i]*ybarra[j]/(J*sigma2teta.sim[i]+sigma2e.sim[i])+sigma2e.sim[i]*
miteta.sim[i+1]/(J*sigma2teta.sim[i]+sigma2e.sim[i])),
sqrt(sigma2teta.sim[i]*sigma2e.sim[i]/(J*sigma2teta.sim[i]+sigma2e.sim[i])))
  }
  sigma2teta.sim[i+1] <- 1/(rgamma(1, shape = (a1+0.5*K), rate=(b1 +
0.5*sum((teta.sim[i+1,]-miteta.sim[i+1])^2))))
  for (l in 1:K) {
    sumsq[i+1,l] <- sum((y[l,]-teta.sim[i+1,l])^2)}
  sigma2e.sim[i+1] <- 1/(rgamma(1, shape = (a2+0.5*K*J), rate=(b2 +
0.5*sum(sumsq[i+1,])))
}

# Resultados do Gibbs-Sampling
teta.estm <- rep(0, K)
for (g in 1:K) {
  teta.estm[g] <- mean(teta.sim[(burn+1):m,g])}

teta.estme <- rep(0, K)
for (g in 1:K) {
  teta.estme[g] <- median(teta.sim[(burn+1):m,g])}

miteta.estm <- mean(miteta.sim[(burn+1):m])

```

```

miteta.estme <- median(miteta.sim[(burn+1):m])

sigma2teta.estm <- mean(sigma2teta.sim[(burn+1):m])

sigma2teta.estme <- median(sigma2teta.sim[(burn+1):m])

sigma2e.estm <- mean(sigma2e.sim[(burn+1):m])

sigma2e.estme <- median(sigma2e.sim[(burn+1):m])

varteta.est <- rep(0, K)
for (g in 1:K) {
  varteta.est[g] <- var(teta.sim[(burn+1):m,g])}

varmiteta.est <- var(miteta.sim[(burn+1):m])

varsigma2teta.est <- var(sigma2teta.sim[(burn+1):m])

varsigma2e.est <- var(sigma2e.sim[(burn+1):m])

cred <- 95 # Percentual de credibilidade
alfa <- (1+cred/100)/2

icteta1 <- quantile(teta.sim[(burn+1):m,1], probs=c((1-alfa), alfa))
icteta2 <- quantile(teta.sim[(burn+1):m,2], probs=c((1-alfa), alfa))
icteta3 <- quantile(teta.sim[(burn+1):m,3], probs=c((1-alfa), alfa))
icteta4 <- quantile(teta.sim[(burn+1):m,4], probs=c((1-alfa), alfa))
icteta5 <- quantile(teta.sim[(burn+1):m,5], probs=c((1-alfa), alfa))
icmiteta <- quantile(miteta.sim[(burn+1):m], probs=c((1-alfa), alfa))
icsigma2teta <- quantile(sigma2teta.sim[(burn+1):m], probs=c((1-alfa), alfa))
icsigma2e <- quantile(sigma2e.sim[(burn+1):m], probs=c((1-alfa), alfa))

# Estimando as posteriores marginais

amplteta1 <- seq(10.5, 15.5, 0.05)
fteta1 <- rep(0, length(amplteta1))
for(i in 1:length(amplteta1)) {
  for(j in 1:(m-burn)) {
    k <- dnorm(amplteta1[i],
(J*sigma2teta.sim[j+burn]*ybarra[1]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn])+
sigma2e.sim[j+burn]*miteta.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn]
))),
sqrt(sigma2teta.sim[j+burn]*sigma2e.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.s
im[j+burn]))/(m-burn)
    fteta1[i] <- fteta1[i] + k
  }
}

amplteta2 <- seq(5, 10, 0.05)
fteta2 <- rep(0, length(amplteta2))
for(i in 1:length(amplteta2)) {
  for(j in 1:(m-burn)) {
    k <- dnorm(amplteta2[i],
(J*sigma2teta.sim[j+burn]*ybarra[2]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn])+
sigma2e.sim[j+burn]*miteta.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn]
))),
sqrt(sigma2teta.sim[j+burn]*sigma2e.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.s
im[j+burn]))/(m-burn)
    fteta2[i] <- fteta2[i] + k
  }
}

amplteta3 <- seq(7, 12, 0.05)
fteta3 <- rep(0, length(amplteta3))
for(i in 1:length(amplteta3)) {
  for(j in 1:(m-burn)) {
    k <- dnorm(amplteta3[i],
(J*sigma2teta.sim[j+burn]*ybarra[3]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn]))+

```

```

sigma2e.sim[j+burn]*miteta.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn
])),
sqrt(sigma2teta.sim[j+burn]*sigma2e.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.s
im[j+burn]))/(m-burn)
      fteta3[i] <- fteta3[i] + k
    }
  }

amplteta4 <- seq(4, 10, 0.05)
fteta4 <- rep(0, length(amplteta4))
for(i in 1:length(amplteta4)) {
  for(j in 1:(m-burn)) {
    k <- dnorm(amplteta4[i],
(J*sigma2teta.sim[j+burn]*ybarra[4]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn])+
sigma2e.sim[j+burn]*miteta.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn
])),
sqrt(sigma2teta.sim[j+burn]*sigma2e.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.s
im[j+burn]))/(m-burn)
      fteta4[i] <- fteta4[i] + k
    }
  }

amplteta5 <- seq(12, 17, 0.05)
fteta5 <- rep(0, length(amplteta5))
for(i in 1:length(amplteta5)) {
  for(j in 1:(m-burn)) {
    k <- dnorm(amplteta5[i],
(J*sigma2teta.sim[j+burn]*ybarra[5]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn])+
sigma2e.sim[j+burn]*miteta.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.sim[j+burn
])),
sqrt(sigma2teta.sim[j+burn]*sigma2e.sim[j+burn]/(J*sigma2teta.sim[j+burn]+sigma2e.s
im[j+burn]))/(m-burn)
      fteta5[i] <- fteta5[i] + k
    }
  }

amplmiteta <- seq(0, 20, 0.1)
fmiteta <- rep(0, length(amplmiteta))
for(i in 1:length(amplmiteta)) {
  for(j in 1:(m-burn)) {
    k <- dnorm(amplmiteta[i], (sigma2teta.sim[j+burn]*mi0 +
sigma20*sum(teta.sim[j+burn,]))/(sigma2teta.sim[j+burn]+K*sigma20),
sqrt(sigma2teta.sim[j+burn]*sigma20/(sigma2teta.sim[j+burn]+K*sigma20)))/(m-burn)
      fmiteta[i] <- fmiteta[i] + k
    }
  }

amplsigma2teta <- seq(0, 50, 0.1)
fsigma2teta <- rep(0, length(amplsigma2teta))
for(i in 1:length(amplsigma2teta)) {
  for(j in 1:(m-burn)) {
    shape = a1+0.5*K
    rate = b1 + 0.5*sum((teta.sim[j+burn,]-miteta.sim[j+burn])^2)
    k = (rate^shape)*amplsigma2teta[i]^(-shape-1)*exp(-
rate/amplsigma2teta[i])/gamma(shape)
    fsigma2teta[i] <- fsigma2teta[i] + k/(m-burn)
  }
}

amplsigma2e <- seq(0, 10, 0.05)
fsigma2e <- rep(0, length(amplsigma2e))
for(i in 1:length(amplsigma2e)) {
  for(j in 1:(m-burn)) {

```

```

        shape = a2+0.5*K*J
        rate = b2 + 0.5*sum(sumsq[j+burn,])
        k = (rate^shape)*amplsigma2e[i]^(-shape-1)*exp(-
rate/amplsigma2e[i])/gamma(shape)
        fsigma2e[i] <- fsigma2e[i] + k/(m-burn)
    }
}

windows()
par(mfrow=c(2,2))

plot(amplteta1, fteta1, xlab="Teta 1", ylab="Posteriori de Teta 1")
title(main="Posteriori marginal estimada para Teta 1", cex.main=0.9)

plot(amplteta2, fteta2, xlab="Teta 2", ylab="Posteriori de Teta 2")
title(main="Posteriori marginal estimada para Teta 2", cex.main=0.9)

plot(amplteta3, fteta3, xlab="Teta 3", ylab="Posteriori de Teta 3")
title(main="Posteriori marginal estimada para Teta 3", cex.main=0.9)

plot(amplteta4, fteta4, xlab="Teta 4", ylab="Posteriori de Teta 4")
title(main="Posteriori marginal estimada para Teta 4", cex.main=0.9)

windows()
par(mfrow=c(2,2))

plot(amplteta5, fteta5, xlab="Teta 5", ylab="Posteriori de Teta 5")
title(main="Posteriori marginal estimada para Teta 5", cex.main=0.8)

plot(amplmiteta, fmiteta, xlab="Mi_Teta", ylab="Posteriori de Mi_Teta")
title(main="Posteriori marginal estimada para Mi_Teta", cex.main=0.8)

plot(amplsigma2teta, fsigma2teta, xlab="Sigma2_Teta", ylab="Posteriori de
Sigma2_Teta")
title(main="Posteriori marginal estimada para Sigma2_Teta", cex.main=0.7)

plot(amplsigma2e, fsigma2e, xlab="Sigma2_Epsilon", ylab="Posteriori de
Sigma2_Epsilon")
title(main="Posteriori marginal estimada para Sigma2_Epsilon", cex.main=0.7)

```

APÊNCIDE C – Programação no WinBUGS para o Exemplo 4.4.2

Análise 1: Estrutura VC para as matrizes R e D

Arquivo de Dados:

```
list(
  n=57, p = 6, q = 2, g = 7,

  mibeta = c(0,0,0,0,0,0),

  H = structure( .Data =c(
    0.00001, 0, 0, 0, 0, 0,
    0, 0.00001, 0, 0, 0, 0,
    0, 0, 0.00001, 0, 0, 0,
    0, 0, 0, 0.00001, 0, 0,
    0, 0, 0, 0, 0.00001, 0,
    0, 0, 0, 0, 0, 0.00001), .Dim=c(6,6) ),

  Y = structure( .Data = c(
    85, 85, 86, 85, 87, 86, 87,
    80, 79, 79, 78, 78, 79, 78,
    78, 77, 77, 77, 76, 76, 77,
    84, 84, 85, 84, 83, 84, 85,
    80, 81, 80, 80, 79, 79, 80,
    76, 78, 77, 78, 78, 77, 74,
    79, 79, 80, 79, 80, 79, 81,
    76, 76, 76, 75, 75, 74, 74,
    77, 78, 78, 80, 80, 81, 80,
    79, 79, 79, 79, 77, 78, 79,
    81, 81, 80, 80, 80, 81, 82,
    77, 76, 77, 78, 77, 77, 77,
    82, 83, 83, 83, 84, 83, 83,
    84, 84, 83, 82, 81, 79, 78,
    79, 81, 81, 82, 82, 82, 80,
    79, 79, 78, 77, 77, 78, 78,
    83, 82, 83, 85, 84, 83, 82,
    78, 78, 79, 79, 78, 77, 77,
    80, 80, 79, 79, 80, 80, 80,
    78, 79, 80, 81, 80, 79, 80,
    79, 79, 79, 80, 80, 78, 80,
    83, 83, 85, 85, 86, 87, 87,
    81, 83, 82, 82, 83, 83, 82,
    81, 81, 81, 82, 82, 83, 81,
    80, 81, 82, 82, 82, 84, 86,
    76, 76, 76, 76, 76, 76, 75,
    81, 84, 83, 83, 85, 85, 85,
    77, 78, 79, 79, 81, 82, 81,
    84, 85, 87, 89, 88, 85, 86,
    74, 75, 78, 78, 79, 78, 78,
    76, 77, 77, 77, 77, 76, 76,
    84, 84, 86, 85, 86, 86, 86,
    79, 80, 79, 80, 80, 82, 82,
    78, 78, 77, 76, 75, 75, 76,
    78, 80, 77, 77, 75, 75, 75,
    84, 85, 85, 85, 85, 83, 82,
    84, 85, 84, 83, 83, 83, 84,
    74, 75, 75, 76, 75, 76, 76,
    83, 84, 82, 81, 83, 83, 82,
    86, 87, 87, 87, 87, 87, 86,
    82, 83, 84, 85, 84, 85, 86,
    79, 80, 79, 79, 80, 79, 80,
    79, 79, 79, 81, 81, 83, 83,
    87, 89, 91, 90, 91, 92, 92,
    81, 81, 81, 82, 82, 83, 83,
    82, 82, 82, 84, 86, 85, 87,
```


Análise 2: Estrutura AR(1) para a matriz R e VC para a matriz DArquivo de Dados:

Idêntico ao arquivo de dados utilizado para a Análise 1.

Arquivo do Modelo:

```

model{
  beta[1:p] ~ dmnorm(mibeta[ ], H[ , ])

  for (i in 1:n) {
    for (j in 1:g) {
      for (c in 1:q) {
        X[i, j, c] <- Z[j, c]
        X[i, j, (c+q)] <- Z[j, c]*t1[i]
        X[i, j, (c+2*q)] <- Z[j, c]*t2[i]
      }
      mi[i, j] <- inprod(X[i, j, ], beta[ ]) + inprod(Z[j, ], u[i, ])
    }
    Y[i, 1:g] ~ dmnorm(mi[i, ], invR[ , ])

    u[i, 1] ~ dnorm(0, taud1)
    u[i, 2] ~ dnorm(0, taud2)
  }

  rho ~ dunif(-1, 1)
  taur ~ dgamma(0.005, 0.001)
  sigma2r <- 1 / taur

  for(m in 1:g) {
    for (l in 1:g) {
      R[m, l] <- sigma2r*pow(rho, abs(m - l))
    }
  }
  invR[1:g, 1:g] <- inverse(R[ , ])

  sigma2d1 <- 1 / taud1
  taud1 ~ dgamma(0.005, 0.001)

  sigma2d2 <- 1 / taud2
  taud2 ~ dgamma(0.005, 0.001)

  alfa.t1 <- beta[1] + beta[3]
  alfa.t2 <- beta[1] + beta[5]
  alfa.t3 <- beta[1]
  beta.t1 <- beta[2] + beta[4]
  beta.t2 <- beta[2] + beta[6]
  beta.t3 <- beta[2]
}

```

Análise 3: Estrutura UN para as matrizes R e DArquivo de Dados:

```

list(
  n=57, p = 6, q = 2, g = 7,

  mi.u = c(0,0),

  Sigma2D = structure( .Data = c(
    1,0,
    0,1), .Dim=c(2,2) ),

  Sigma2R = structure( .Data = c(
    1,0,0,0,0,0,0,
    0,1,0,0,0,0,0,
    0,0,1,0,0,0,0,

```

```

0,0,0,1,0,0,0,
0,0,0,0,1,0,0,
0,0,0,0,0,1,0,
0,0,0,0,0,0,1), .Dim=c(7,7) ),

```

```

mibeta = c(0,0,0,0,0,0),

```

```

H = structure( .Data =c(
0.00001, 0, 0, 0, 0, 0,
0, 0.00001, 0, 0, 0, 0,
0, 0, 0.00001, 0, 0, 0,
0, 0, 0, 0.00001, 0, 0,
0, 0, 0, 0, 0.00001, 0,
0, 0, 0, 0, 0, 0.00001), .Dim=c(6,6) ),

```

```

Y = structure( .Data = c(
85, 85, 86, 85, 87, 86, 87,
80, 79, 79, 78, 78, 79, 78,
78, 77, 77, 77, 76, 76, 77,
84, 84, 85, 84, 83, 84, 85,
80, 81, 80, 80, 79, 79, 80,
76, 78, 77, 78, 78, 77, 74,
79, 79, 80, 79, 80, 79, 81,
76, 76, 76, 75, 75, 74, 74,
77, 78, 78, 80, 80, 81, 80,
79, 79, 79, 79, 77, 78, 79,
81, 81, 80, 80, 80, 81, 82,
77, 76, 77, 78, 77, 77, 77,
82, 83, 83, 83, 84, 83, 83,
84, 84, 83, 82, 81, 79, 78,
79, 81, 81, 82, 82, 82, 80,
79, 79, 78, 77, 77, 78, 78,
83, 82, 83, 85, 84, 83, 82,
78, 78, 79, 79, 78, 77, 77,
80, 80, 79, 79, 80, 80, 80,
78, 79, 80, 81, 80, 79, 80,
79, 79, 79, 80, 80, 78, 80,
83, 83, 85, 85, 86, 87, 87,
81, 83, 82, 82, 83, 83, 82,
81, 81, 81, 82, 82, 83, 81,
80, 81, 82, 82, 82, 84, 86,
76, 76, 76, 76, 76, 76, 75,
81, 84, 83, 83, 85, 85, 85,
77, 78, 79, 79, 81, 82, 81,
84, 85, 87, 89, 88, 85, 86,
74, 75, 78, 78, 79, 78, 78,
76, 77, 77, 77, 77, 76, 76,
84, 84, 86, 85, 86, 86, 86,
79, 80, 79, 80, 80, 82, 82,
78, 78, 77, 76, 75, 75, 76,
78, 80, 77, 77, 75, 75, 75,
84, 85, 85, 85, 85, 83, 82,
84, 85, 84, 83, 83, 83, 84,
74, 75, 75, 76, 75, 76, 76,
83, 84, 82, 81, 83, 83, 82,
86, 87, 87, 87, 87, 87, 86,
82, 83, 84, 85, 84, 85, 86,
79, 80, 79, 79, 80, 79, 80,
79, 79, 79, 81, 81, 83, 83,
87, 89, 91, 90, 91, 92, 92,
81, 81, 81, 82, 82, 83, 83,
82, 82, 82, 84, 86, 85, 87,
79, 79, 80, 81, 81, 81, 81,
79, 80, 81, 82, 83, 82, 82,
83, 84, 84, 84, 84, 83, 83,
81, 81, 82, 84, 83, 82, 85,
78, 78, 79, 79, 78, 79, 79,
83, 82, 82, 84, 84, 83, 84,

```


APÊNDICE D – Programação em R para o Estudo de Simulação em 4.5

Estudo 1: Suposição de estrutura de covariância VC para a matriz R atendida

```
# Precisa carregar o pacote nnet!

local({pkg <- select.list(sort(.packages(all.available = TRUE)))
if(nchar(pkg)) library(pkg, character.only=TRUE)})

n = 27 #Número de sujeitos
g = 5 #Número de tempos
t = 3 #Número de tratamentos
q = 2 #número de efeitos aleatórios
p = 6 # número de efeitos fixos
cvar = 3 #Número de componentes de variância
nparms=p+q*n+cvar #Número de parâmetros

n1 = 8
n2 = 10
n3 = 9

trat = matrix(c(rep(1, n1), rep(0, (n-n1)), rep(0, n1), rep(1, n2), rep(0, n3)),
nrow=n, ncol=(t-1))

tempo = seq(1, 5, 1)
sum2tempo = sum(tempo^2)

alfa0.def = 1
alfa1.def = 5
alfa2.def = 3
beta0.def = 2
beta1.def = -4
beta2.def = -6

sigma2alfa.def = 1
sigma2beta.def = 4
sigma2e.def = 9

alfa.suj.def = rep(0,n)
beta.suj.def = rep(0,n)
for(i in 1:n) {
  alfa.suj.def[i] = rnorm(1,
(alfa0.def+alfa1.def*trat[i,1]+alfa2.def*trat[i,2]), sqrt(sigma2alfa.def))
  beta.suj.def[i] = rnorm(1,
(beta0.def+beta1.def*trat[i,1]+beta2.def*trat[i,2]), sqrt(sigma2beta.def))
}

# Parâmetros das prioris
a0 = 0
c02 = 10000
a1 = 0
c12 = 10000
a2 = 0
c22 = 10000
b0 = 0
d02 = 10000
b1 = 0
d12 = 10000
b2 = 0
d22 = 10000
e1 = 0
e2 = 0

# Gibbs Sampling
m = 11000 # Tamanho da cadeia
burn = 1000 # Valores a serem desprezados
```

```

rep = 500 # Replicações da simulação

cred <- 95 # Percentual de credibilidade
alfa <- (1+cred/100)/2

media = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
mediana = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
moda = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
var = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
perc1 = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
perc2 = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)

for(r in 1:rep){

  y = matrix(rep(0, n*g), nrow=n, ncol=g)
  for(i in 1:n) {
    for(j in 1:g){
      y[i,j] = rnorm(1, (alfa.suj.def[i]+beta.suj.def[i]*tempo[j]),
sqrt(sigma2e.def))
    }
  }

  alfa.suj = matrix(rep(0.01, m*n), nrow=m, ncol=n)
  beta.suj = matrix(rep(0.01, m*n), nrow=m, ncol=n)
  alfa0 = rep(0.01,m)
  alfa1 = rep(0.01,m)
  alfa2 = rep(0.01,m)
  beta0 = rep(0.01,m)
  beta1 = rep(0.01,m)
  beta2 = rep(0.01,m)
  sigma2e = rep(0.001, m)
  sigma2alfa = rep(0.001, m)
  sigma2beta = rep(0.001, m)

  for (l in 1:(m-1)) {

    aux1 = (n-1)/2
    aux2 = sum((alfa.suj[l, ]-(alfa0[l] + alfa1[l]*trat[ ,1] +
alfa2[l]*trat[ ,2]))^2)/2
    sigma2alfa[l+1] = 1/rgamma(1, shape=aux1, rate=aux2)

    aux3 = sum((beta.suj[l, ]-(beta0[l] + beta1[l]*trat[ ,1] +
beta2[l]*trat[ ,2]))^2)/2
    sigma2beta[l+1] = 1/rgamma(1, shape=aux1, rate=aux3)

    aux4 = n*g/2 + e1
    aux5 = rep(0, n)
    for(i in 1:n){
      aux5[i] = sum((y[i, ] - (alfa.suj[l, i] + beta.suj[l,
i]*tempo))^2)
    }
    aux6 = sum(aux5)/2 + e2
    sigma2e[l+1] = 1/rgamma(1, shape=aux4, rate=aux6)

    aux7 = (c02*(sum(alfa.suj[l, ])-alfa1[l]*n1-
alfa2[l]*n2)+a0*sigma2alfa[l+1])/(sigma2alfa[l+1] + n*c02)
    aux8 = (sigma2alfa[l+1]*c02)/(sigma2alfa[l+1] + n*c02)
    alfa0[l+1] = rnorm(1, aux7, sqrt(aux8))

    aux9 = (c12*(sum(trat[ ,1]*alfa.suj[l, ])-
n1*alfa0[l+1])+sigma2alfa[l+1]*a1)/(n1*c12+sigma2alfa[l+1])
    aux10 = (sigma2alfa[l+1]*c12)/(sigma2alfa[l+1] + n1*c12)
    alfa1[l+1] = rnorm(1, aux9, sqrt(aux10))

    aux11 = (c22*(sum(trat[ ,2]*alfa.suj[l, ])-
n2*alfa0[l+1])+sigma2alfa[l+1]*a2)/(n2*c22+sigma2alfa[l+1])
    aux12 = (sigma2alfa[l+1]*c22)/(sigma2alfa[l+1] + n2*c22)
    alfa2[l+1] = rnorm(1, aux11, sqrt(aux12))
  }
}

```

```

aux13 = (d02*(sum(beta.suj[l, ])-beta1[l]*n1-
beta2[l]*n2)+b0*sigma2beta[l+1])/(sigma2beta[l+1] + n*d02)
aux14 = (sigma2beta[l+1]*d02)/(sigma2beta[l+1] + n*d02)
beta0[l+1] = rnorm(1, aux13, sqrt(aux14))

aux15 = (d12*(sum(trat[,1]*beta.suj[l, ])-
n1*beta0[l+1])+sigma2beta[l+1]*b1)/(n1*d12+sigma2beta[l+1])
aux16 = (sigma2beta[l+1]*d12)/(sigma2beta[l+1] + n1*d12)
beta1[l+1] = rnorm(1, aux15, sqrt(aux16))

aux17 = (d22*(sum(trat[,2]*beta.suj[l, ])-
n2*beta0[l+1])+sigma2beta[l+1]*b2)/(n2*d22+sigma2beta[l+1])
aux18 = (sigma2beta[l+1]*d22)/(sigma2beta[l+1] + n2*d22)
beta2[l+1] = rnorm(1, aux17, sqrt(aux18))

aux19 =
(sigma2alfa[l+1]*sigma2e[l+1])/(g*sigma2alfa[l+1]+sigma2e[l+1])
aux20 =
(sigma2beta[l+1]*sigma2e[l+1])/(sum2tempo*sigma2beta[l+1]+sigma2e[l+1])
aux21 = rep(0, n)
aux22 = rep(0, n)
for(i in 1:n){
  aux21[i] = (sigma2alfa[l+1]*sum(y[i, ]-beta.suj[l,
i]*tempo)+sigma2e[l+1]*(alfa0[l+1]+alfa1[l+1]*trat[i,1]+alfa2[l+1]*trat[i,2]))/(g*s
igma2alfa[l+1]+sigma2e[l+1])
  alfa.suj[l+1, i] = rnorm(1, aux21[i], sqrt(aux19))
  aux22[i] = (sigma2beta[l+1]*sum(tempo*(y[i, ]-alfa.suj[l+1,
i]))+sigma2e[l+1]*(beta0[l+1]+beta1[l+1]*trat[i,1]+beta2[l+1]*trat[i,2]))/(sum2temp
o*sigma2beta[l+1]+sigma2e[l+1])
  beta.suj[l+1, i] = rnorm(1, aux22[i], sqrt(aux20))
}

)

alfa.t1 = alfa0+alfa1
alfa.t2 = alfa0+alfa2
alfa.t3 = alfa0
beta.t1 = beta0+beta1
beta.t2 = beta0+beta2
beta.t3 = beta0

# Resultados do Gibbs Sampling

for(i in 1:n) {
  media[r, i] = mean(alfa.suj[(burn+1):m, i])
  media[r, (i+n)] = mean(beta.suj[(burn+1):m, i])
  mediana[r, i] = median(alfa.suj[(burn+1):m, i])
  mediana[r, (i+n)] = median(beta.suj[(burn+1):m, i])
  perc1[r, i] = quantile(alfa.suj[(burn+1):m, i], probs=c(1-alfa))
  perc2[r, i] = quantile(alfa.suj[(burn+1):m, i], probs=c(alfa))
  perc1[r, (i+n)] = quantile(beta.suj[(burn+1):m, i], probs=c(1-alfa))
  perc2[r, (i+n)] = quantile(beta.suj[(burn+1):m, i], probs=c(alfa))
  var[r, i] = var(alfa.suj[(burn+1):m, i])
  var[r, (i+n)] = var(beta.suj[(burn+1):m, i])
  kernell=density(alfa.suj[(burn+1):m, i], bw="nrd0", n = 512,
from=min(alfa.suj[(burn+1):m, i]), to=max(alfa.suj[(burn+1):m, i]))
  posmo1 = which.is.max(kernell$y)
  moda[r, i] = kernell$x[posmo1]
  kernel2=density(beta.suj[(burn+1):m, i], bw="nrd0", n = 512,
from=min(beta.suj[(burn+1):m, i]), to=max(beta.suj[(burn+1):m, i]))
  posmo2 = which.is.max(kernel2$y)
  moda[r, (i+n)] = kernel2$x[posmo2]
}

)

media[r, 55] = mean(alfa.t1[(burn+1):m])
media[r, 56] = mean(alfa.t2[(burn+1):m])
media[r, 57] = mean(alfa.t3[(burn+1):m])
media[r, 58] = mean(beta.t1[(burn+1):m])

```

```

media[r, 59] = mean(beta.t2[(burn+1):m])
media[r, 60] = mean(beta.t3[(burn+1):m])
media[r, 61] = mean(sigma2alfa[(burn+1):m])
media[r, 62] = mean(sigma2beta[(burn+1):m])
media[r, 63] = mean(sigma2e[(burn+1):m])

mediana[r, 55] = median(alfa.t1[(burn+1):m])
mediana[r, 56] = median(alfa.t2[(burn+1):m])
mediana[r, 57] = median(alfa.t3[(burn+1):m])
mediana[r, 58] = median(beta.t1[(burn+1):m])
mediana[r, 59] = median(beta.t2[(burn+1):m])
mediana[r, 60] = median(beta.t3[(burn+1):m])
mediana[r, 61] = median(sigma2alfa[(burn+1):m])
mediana[r, 62] = median(sigma2beta[(burn+1):m])
mediana[r, 63] = median(sigma2e[(burn+1):m])

perc1[r, 55] <- quantile(alfa.t1[(burn+1):m], probs=c(1-alfa))
perc1[r, 56] <- quantile(alfa.t2[(burn+1):m], probs=c(1-alfa))
perc1[r, 57] <- quantile(alfa.t3[(burn+1):m], probs=c(1-alfa))
perc1[r, 58] <- quantile(beta.t1[(burn+1):m], probs=c(1-alfa))
perc1[r, 59] <- quantile(beta.t2[(burn+1):m], probs=c(1-alfa))
perc1[r, 60] <- quantile(beta.t3[(burn+1):m], probs=c(1-alfa))
perc1[r, 61] <- quantile(sigma2alfa[(burn+1):m], probs=c(1-alfa))
perc1[r, 62] <- quantile(sigma2beta[(burn+1):m], probs=c(1-alfa))
perc1[r, 63] <- quantile(sigma2e[(burn+1):m], probs=c(1-alfa))

perc2[r, 55] <- quantile(alfa.t1[(burn+1):m], probs=c(alfa))
perc2[r, 56] <- quantile(alfa.t2[(burn+1):m], probs=c(alfa))
perc2[r, 57] <- quantile(alfa.t3[(burn+1):m], probs=c(alfa))
perc2[r, 58] <- quantile(beta.t1[(burn+1):m], probs=c(alfa))
perc2[r, 59] <- quantile(beta.t2[(burn+1):m], probs=c(alfa))
perc2[r, 60] <- quantile(beta.t3[(burn+1):m], probs=c(alfa))
perc2[r, 61] <- quantile(sigma2alfa[(burn+1):m], probs=c(alfa))
perc2[r, 62] <- quantile(sigma2beta[(burn+1):m], probs=c(alfa))
perc2[r, 63] <- quantile(sigma2e[(burn+1):m], probs=c(alfa))

var[r, 55] = var(alfa.t1[(burn+1):m])
var[r, 56] = var(alfa.t2[(burn+1):m])
var[r, 57] = var(alfa.t3[(burn+1):m])
var[r, 58] = var(beta.t1[(burn+1):m])
var[r, 59] = var(beta.t2[(burn+1):m])
var[r, 60] = var(beta.t3[(burn+1):m])
var[r, 61] = var(sigma2alfa[(burn+1):m])
var[r, 62] = var(sigma2beta[(burn+1):m])
var[r, 63] = var(sigma2e[(burn+1):m])

kernel3 = density(alfa.t1[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t1[(burn+1):m]), to=max(alfa.t1[(burn+1):m]))
posmo3 = which.is.max(kernel3$y)
moda[r, 55] = kernel3$x[posmo3]

kernel4 = density(alfa.t2[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t2[(burn+1):m]), to=max(alfa.t2[(burn+1):m]))
posmo4 = which.is.max(kernel4$y)
moda[r, 56] = kernel4$x[posmo4]

kernel5 = density(alfa.t3[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t3[(burn+1):m]), to=max(alfa.t3[(burn+1):m]))
posmo5 = which.is.max(kernel5$y)
moda[r, 57] = kernel5$x[posmo5]

kernel6 = density(beta.t1[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t1[(burn+1):m]), to=max(beta.t1[(burn+1):m]))
posmo6 = which.is.max(kernel6$y)
moda[r, 58] = kernel6$x[posmo6]

kernel7 = density(beta.t2[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t2[(burn+1):m]), to=max(beta.t2[(burn+1):m]))

```

```

posmo7 = which.is.max(kernel7$y)
moda[r, 59] = kernel7$x[posmo7]

kernel8 = density(beta.t3[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t3[(burn+1):m]), to=max(beta.t3[(burn+1):m]))
posmo8 = which.is.max(kernel8$y)
moda[r, 60] = kernel8$x[posmo8]

kernel9 = density(sigma2alfa[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2alfa[(burn+1):m]), to=max(sigma2alfa[(burn+1):m]))
posmo9 = which.is.max(kernel9$y)
moda[r, 61] = kernel9$x[posmo9]

kernel10 = density(sigma2beta[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2beta[(burn+1):m]), to=max(sigma2beta[(burn+1):m]))
posmo10 = which.is.max(kernel10$y)
moda[r, 62] = kernel10$x[posmo10]

kernel11 = density(sigma2e[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2e[(burn+1):m]), to=max(sigma2e[(burn+1):m]))
posmo11 = which.is.max(kernel11$y)
moda[r, 63] = kernel11$x[posmo11]
}

windows()
par(mfrow=c(1,3))
ts.plot(sigma2alfa)
ts.plot(sigma2beta)
ts.plot(sigma2e)

windows()
par(mfrow=c(2,3))
ts.plot(alfa.t1)
ts.plot(alfa.t2)
ts.plot(alfa.t3)
ts.plot(beta.t1)
ts.plot(beta.t2)
ts.plot(beta.t3)

windows()
par(mfrow=c(2,2))
ts.plot(alfa.suj[,1])
ts.plot(beta.suj[,1])
ts.plot(alfa.suj[,2])
ts.plot(beta.suj[,2])

windows()
par(mfrow=c(3,2))

boxplot(media[,1], moda[,1], mediana[,1], names=c("média", "moda", "mediana"))
title(main="Boxplot das estimativas para alfa.suj.1")

boxplot(media[,28], moda[,28], mediana[,28], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.suj.1")

boxplot(media[,9], moda[,9], mediana[,9], names=c("média", "moda", "mediana"))
title(main="Boxplot das estimativas para alfa.suj.9")

boxplot(media[,36], moda[,36], mediana[,36], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.suj.9")

boxplot(media[,19], moda[,19], mediana[,19], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para alfa.suj.19")

```

```

boxplot(media[,46], moda[,46], mediana[,46], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.suj.19")

windows()
par(mfrow=c(3,2))

boxplot(media[,55], moda[,55], mediana[,55], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para alfa.t1")

boxplot(media[,58], moda[,58], mediana[,58], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.t1")

boxplot(media[,56], moda[,56], mediana[,56], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para alfa.t2")

boxplot(media[,59], moda[,59], mediana[,59], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.t2")

boxplot(media[,57], moda[,57], mediana[,57], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para alfa.t3")

boxplot(media[,60], moda[,60], mediana[,60], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.t3")

windows()
par(mfrow=c(3,1))

boxplot(media[,61], moda[,61], mediana[,61], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para sigma2alfa")
boxplot(media[,62], moda[,62], mediana[,62], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para sigma2beta")
boxplot(media[,63], moda[,63], mediana[,63], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para sigma2e")

```

Estudo 2: Suposição de estrutura de covariância VC para a matriz R não atendida

```

# Precisa carregar os pacotes MASS e nnet

local({pkg <- select.list(sort(.packages(all.available = TRUE)))
if(nchar(pkg)) library(pkg, character.only=TRUE)})

local({pkg <- select.list(sort(.packages(all.available = TRUE)))
if(nchar(pkg)) library(pkg, character.only=TRUE)})

n = 27 #Número de sujeitos
g = 5 #Número de tempos
t = 3 #Número de tratamentos
q = 2 #número de efeitos aleatórios
p = 6 # número de efeitos fixos
cvar = 3 #Número de componentes de variância
nparms=p+q*n+cvar #Número de parâmetros

n1 = 8
n2 = 10
n3 = 9

trat = matrix(c(rep(1, n1), rep(0, (n-n1)), rep(0, n1), rep(1, n2), rep(0, n3)),
nrow=n, ncol=(t-1))

```

```

tempo = seq(1, g, 1)
sum2tempo = sum(tempo^2)

alfa0.def = 1
alfa1.def = 5
alfa2.def = 3
beta0.def = 2
beta1.def = -4
beta2.def = -6

sigma2alfa.def = 1
sigma2beta.def = 4

alfa.suj.def = rep(0,n)
beta.suj.def = rep(0,n)
miy = matrix(0, nrow=n, ncol=g)
for(i in 1:n) {
  alfa.suj.def[i] = rnorm(1,
(alfa0.def+alfa1.def*trat[i,1]+alfa2.def*trat[i,2]), sqrt(sigma2alfa.def))
  beta.suj.def[i] = rnorm(1,
(beta0.def+beta1.def*trat[i,1]+beta2.def*trat[i,2]), sqrt(sigma2beta.def))
  for(j in 1:g){
    miy[i,j] = alfa.suj.def[i]+beta.suj.def[i]*tempo[j]
  }
}

R.def = matrix(0, nrow=g, ncol=g)
R.def[1,1] = 3
R.def[2,2] = 1
R.def[3,3] = 2
R.def[4,4] = 1.5
R.def[5,5] = 0.5

R.def[1,2] = 0.95
R.def[2,1] = R.def[1,2]
R.def[1,3] = 0.9
R.def[3,1] = R.def[1,3]
R.def[1,4] = 0.8
R.def[4,1] = R.def[1,4]
R.def[1,5] = 0.7
R.def[5,1] = R.def[1,5]

R.def[2,3] = 0.9
R.def[3,2] = R.def[2,3]
R.def[2,4] = 0.8
R.def[4,2] = R.def[2,4]
R.def[2,5] = 0.6
R.def[5,2] = R.def[2,5]

R.def[3,4] = 0.8
R.def[4,3] = R.def[3,4]
R.def[3,5] = 0.65
R.def[5,3] = R.def[3,5]

R.def[4,5] = 0.75
R.def[5,4] = R.def[4,5]

# Parâmetros das prioris
a0 = 0
c02 = 10000
a1 = 0
c12 = 10000
a2 = 0
c22 = 10000
b0 = 0
d02 = 10000
b1 = 0
d12 = 10000

```

```

b2 = 0
d22 = 10000
e1 = 0
e2 = 0

# Gibbs Sampling

m = 11000 # Tamanho da cadeia
burn = 1000 # Valores a serem desprezados
rep = 500 # Replicações da simulação

cred <- 95 # Percentual de credibilidade
alfa <- (1+cred/100)/2

media = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
mediana = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
moda = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
var = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
perc1 = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)
perc2 = matrix(rep(0, rep*nparms), nrow=rep, ncol=nparms)

for(r in 1:rep){

  y = matrix(0, nrow=n, ncol=g)

  for(i in 1:n) {
    y[i, ] = mvrnorm(1, miy[i, ], R.def)}

  alfa.suj = matrix(rep(0.01, m*n), nrow=m, ncol=n)
  beta.suj = matrix(rep(0.01, m*n), nrow=m, ncol=n)
  alfa0 = rep(0.01,m)
  alfa1 = rep(0.01,m)
  alfa2 = rep(0.01,m)
  beta0 = rep(0.01,m)
  beta1 = rep(0.01,m)
  beta2 = rep(0.01,m)
  sigma2e = rep(0.001, m)
  sigma2alfa = rep(0.001, m)
  sigma2beta = rep(0.001, m)

  for (l in 1:(m-1)) {

    aux1 = (n-1)/2
    aux2 = sum((alfa.suj[l, ]-(alfa0[l] + alfa1[l]*trat[ ,1] +
alfa2[l]*trat[ ,2]))^2)/2
    sigma2alfa[l+1] = 1/rgamma(1, shape=aux1, rate=aux2)

    aux3 = sum((beta.suj[l, ]-(beta0[l] + beta1[l]*trat[ ,1] +
beta2[l]*trat[ ,2]))^2)/2
    sigma2beta[l+1] = 1/rgamma(1, shape=aux1, rate=aux3)

    aux4 = n*g/2 + e1
    aux5 = rep(0, n)
    for(i in 1:n){
      aux5[i] = sum((y[i, ] - (alfa.suj[l, i] + beta.suj[l,
i]*tempo))^2)
    }
    aux6 = sum(aux5)/2 + e2
    sigma2e[l+1] = 1/rgamma(1, shape=aux4, rate=aux6)

    aux7 = (c02*(sum(alfa.suj[l, ])-alfa1[l]*n1-
alfa2[l]*n2)+a0*sigma2alfa[l+1])/(sigma2alfa[l+1] + n*c02)
    aux8 = (sigma2alfa[l+1]*c02)/(sigma2alfa[l+1] + n*c02)
    alfa0[l+1] = rnorm(1, aux7, sqrt(aux8))

    aux9 = (c12*(sum(trat[ ,1]*alfa.suj[l, ])-
n1*alfa0[l+1])+sigma2alfa[l+1]*a1)/(n1*c12+sigma2alfa[l+1])
    aux10 = (sigma2alfa[l+1]*c12)/(sigma2alfa[l+1] + n1*c12)

```

```

    alfa[l+1] = rnorm(1, aux9, sqrt(aux10))

    aux11 = (c22*(sum(trat[,2]*alfa.suj[l,])-
n2*alfa0[l+1])+sigma2alfa[l+1]*a2)/(n2*c22+sigma2alfa[l+1])
    aux12 = (sigma2alfa[l+1]*c22)/(sigma2alfa[l+1] + n2*c22)
    alfa2[l+1] = rnorm(1, aux11, sqrt(aux12))

    aux13 = (d02*(sum(beta.suj[l,])-beta1[l]*n1-
beta2[l]*n2)+b0*sigma2beta[l+1])/(sigma2beta[l+1] + n*d02)
    aux14 = (sigma2beta[l+1]*d02)/(sigma2beta[l+1] + n*d02)
    beta0[l+1] = rnorm(1, aux13, sqrt(aux14))

    aux15 = (d12*(sum(trat[,1]*beta.suj[l,])-
n1*beta0[l+1])+sigma2beta[l+1]*b1)/(n1*d12+sigma2beta[l+1])
    aux16 = (sigma2beta[l+1]*d12)/(sigma2beta[l+1] + n1*d12)
    beta1[l+1] = rnorm(1, aux15, sqrt(aux16))

    aux17 = (d22*(sum(trat[,2]*beta.suj[l,])-
n2*beta0[l+1])+sigma2beta[l+1]*b2)/(n2*d22+sigma2beta[l+1])
    aux18 = (sigma2beta[l+1]*d22)/(sigma2beta[l+1] + n2*d22)
    beta2[l+1] = rnorm(1, aux17, sqrt(aux18))

    aux19 =
(sigma2alfa[l+1]*sigma2e[l+1])/(g*sigma2alfa[l+1]+sigma2e[l+1])
    aux20 =
(sigma2beta[l+1]*sigma2e[l+1])/(sum2tempo*sigma2beta[l+1]+sigma2e[l+1])
    aux21 = rep(0, n)
    aux22 = rep(0, n)
    for(i in 1:n){
        aux21[i] = (sigma2alfa[l+1]*sum(y[i,]-beta.suj[l,
i]*tempo)+sigma2e[l+1]*(alfa0[l+1]+alfa1[l+1]*trat[i,1]+alfa2[l+1]*trat[i,2]))/(g*s
igma2alfa[l+1]+sigma2e[l+1])
        alfa.suj[l+1, i] = rnorm(1, aux21[i], sqrt(aux19))
        aux22[i] = (sigma2beta[l+1]*sum(tempo*(y[i,]-alfa.suj[l+1,
i]))+sigma2e[l+1]*(beta0[l+1]+beta1[l+1]*trat[i,1]+beta2[l+1]*trat[i,2]))/(sum2temp
o*sigma2beta[l+1]+sigma2e[l+1])
        beta.suj[l+1, i] = rnorm(1, aux22[i], sqrt(aux20))
    }
}

alfa.t1 = alfa0+alfa1
alfa.t2 = alfa0+alfa2
alfa.t3 = alfa0
beta.t1 = beta0+beta1
beta.t2 = beta0+beta2
beta.t3 = beta0

# Resultados do Gibbs Sampling
for(i in 1:n) {
    media[r, i] = mean(alfa.suj[(burn+1):m, i])
    media[r, (i+n)] = mean(beta.suj[(burn+1):m, i])
    mediana[r, i] = median(alfa.suj[(burn+1):m, i])
    mediana[r, (i+n)] = median(beta.suj[(burn+1):m, i])
    perc1[r, i] = quantile(alfa.suj[(burn+1):m, i], probs=c(1-alfa))
    perc2[r, i] = quantile(alfa.suj[(burn+1):m, i], probs=c(alfa))
    perc1[r, (i+n)] = quantile(beta.suj[(burn+1):m, i], probs=c(1-alfa))
    perc2[r, (i+n)] = quantile(beta.suj[(burn+1):m, i], probs=c(alfa))
    var[r, i] = var(alfa.suj[(burn+1):m, i])
    var[r, (i+n)] = var(beta.suj[(burn+1):m, i])
    kernell=density(alfa.suj[(burn+1):m, i], bw="nrd0", n = 512,
from=min(alfa.suj[(burn+1):m, i]), to=max(alfa.suj[(burn+1):m, i]))
    posmol = which.is.max(kernell$y)
    moda[r, i] = kernell$x[posmol]
    kernell2=density(beta.suj[(burn+1):m, i], bw="nrd0", n = 512,
from=min(beta.suj[(burn+1):m, i]), to=max(beta.suj[(burn+1):m, i]))
    posmo2 = which.is.max(kernell2$y)
    moda[r, (i+n)] = kernell2$x[posmo2]
}

```

```

media[r, 55] = mean(alfa.t1[(burn+1):m])
media[r, 56] = mean(alfa.t2[(burn+1):m])
media[r, 57] = mean(alfa.t3[(burn+1):m])
media[r, 58] = mean(beta.t1[(burn+1):m])
media[r, 59] = mean(beta.t2[(burn+1):m])
media[r, 60] = mean(beta.t3[(burn+1):m])
media[r, 61] = mean(sigma2alfa[(burn+1):m])
media[r, 62] = mean(sigma2beta[(burn+1):m])
media[r, 63] = mean(sigma2e[(burn+1):m])

mediana[r, 55] = median(alfa.t1[(burn+1):m])
mediana[r, 56] = median(alfa.t2[(burn+1):m])
mediana[r, 57] = median(alfa.t3[(burn+1):m])
mediana[r, 58] = median(beta.t1[(burn+1):m])
mediana[r, 59] = median(beta.t2[(burn+1):m])
mediana[r, 60] = median(beta.t3[(burn+1):m])
mediana[r, 61] = median(sigma2alfa[(burn+1):m])
mediana[r, 62] = median(sigma2beta[(burn+1):m])
mediana[r, 63] = median(sigma2e[(burn+1):m])

percl[r, 55] <- quantile(alfa.t1[(burn+1):m], probs=c(1-alfa))
percl[r, 56] <- quantile(alfa.t2[(burn+1):m], probs=c(1-alfa))
percl[r, 57] <- quantile(alfa.t3[(burn+1):m], probs=c(1-alfa))
percl[r, 58] <- quantile(beta.t1[(burn+1):m], probs=c(1-alfa))
percl[r, 59] <- quantile(beta.t2[(burn+1):m], probs=c(1-alfa))
percl[r, 60] <- quantile(beta.t3[(burn+1):m], probs=c(1-alfa))
percl[r, 61] <- quantile(sigma2alfa[(burn+1):m], probs=c(1-alfa))
percl[r, 62] <- quantile(sigma2beta[(burn+1):m], probs=c(1-alfa))
percl[r, 63] <- quantile(sigma2e[(burn+1):m], probs=c(1-alfa))

perc2[r, 55] <- quantile(alfa.t1[(burn+1):m], probs=c(alfa))
perc2[r, 56] <- quantile(alfa.t2[(burn+1):m], probs=c(alfa))
perc2[r, 57] <- quantile(alfa.t3[(burn+1):m], probs=c(alfa))
perc2[r, 58] <- quantile(beta.t1[(burn+1):m], probs=c(alfa))
perc2[r, 59] <- quantile(beta.t2[(burn+1):m], probs=c(alfa))
perc2[r, 60] <- quantile(beta.t3[(burn+1):m], probs=c(alfa))
perc2[r, 61] <- quantile(sigma2alfa[(burn+1):m], probs=c(alfa))
perc2[r, 62] <- quantile(sigma2beta[(burn+1):m], probs=c(alfa))
perc2[r, 63] <- quantile(sigma2e[(burn+1):m], probs=c(alfa))

var[r, 55] = var(alfa.t1[(burn+1):m])
var[r, 56] = var(alfa.t2[(burn+1):m])
var[r, 57] = var(alfa.t3[(burn+1):m])
var[r, 58] = var(beta.t1[(burn+1):m])
var[r, 59] = var(beta.t2[(burn+1):m])
var[r, 60] = var(beta.t3[(burn+1):m])
var[r, 61] = var(sigma2alfa[(burn+1):m])
var[r, 62] = var(sigma2beta[(burn+1):m])
var[r, 63] = var(sigma2e[(burn+1):m])

kernel3 = density(alfa.t1[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t1[(burn+1):m]), to=max(alfa.t1[(burn+1):m]))
posmo3 = which.is.max(kernel3$y)
moda[r, 55] = kernel3$x[posmo3]

kernel4 = density(alfa.t2[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t2[(burn+1):m]), to=max(alfa.t2[(burn+1):m]))
posmo4 = which.is.max(kernel4$y)
moda[r, 56] = kernel4$x[posmo4]

kernel5 = density(alfa.t3[(burn+1):m], bw="nrd0", n = 512,
from=min(alfa.t3[(burn+1):m]), to=max(alfa.t3[(burn+1):m]))
posmo5 = which.is.max(kernel5$y)
moda[r, 57] = kernel5$x[posmo5]

kernel6 = density(beta.t1[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t1[(burn+1):m]), to=max(beta.t1[(burn+1):m]))
posmo6 = which.is.max(kernel6$y)

```

```

moda[r, 58] = kernel6$x[posmo6]

kernel7 = density(beta.t2[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t2[(burn+1):m]), to=max(beta.t2[(burn+1):m]))
posmo7 = which.is.max(kernel7$y)
moda[r, 59] = kernel7$x[posmo7]

kernel8 = density(beta.t3[(burn+1):m], bw="nrd0", n = 512,
from=min(beta.t3[(burn+1):m]), to=max(beta.t3[(burn+1):m]))
posmo8 = which.is.max(kernel8$y)
moda[r, 60] = kernel8$x[posmo8]

kernel9 = density(sigma2alfa[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2alfa[(burn+1):m]), to=max(sigma2alfa[(burn+1):m]))
posmo9 = which.is.max(kernel9$y)
moda[r, 61] = kernel9$x[posmo9]

kernel10 = density(sigma2beta[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2beta[(burn+1):m]), to=max(sigma2beta[(burn+1):m]))
posmo10 = which.is.max(kernel10$y)
moda[r, 62] = kernel10$x[posmo10]

kernel11 = density(sigma2e[(burn+1):m], bw="nrd0", n = 512,
from=min(sigma2e[(burn+1):m]), to=max(sigma2e[(burn+1):m]))
posmo11 = which.is.max(kernel11$y)
moda[r, 63] = kernel11$x[posmo11]
}

windows()
par(mfrow=c(1,3))
ts.plot(sigma2alfa)
ts.plot(sigma2beta)
ts.plot(sigma2e)

windows()
par(mfrow=c(2,3))
ts.plot(alfa.t1)
ts.plot(alfa.t2)
ts.plot(alfa.t3)
ts.plot(beta.t1)
ts.plot(beta.t2)
ts.plot(beta.t3)

windows()
par(mfrow=c(2,2))
ts.plot(alfa.suj[,1])
ts.plot(beta.suj[,1])
ts.plot(alfa.suj[,2])
ts.plot(beta.suj[,2])
windows()
par(mfrow=c(3,2))

boxplot(media[,1], moda[,1], mediana[,1], names=c("média", "moda", "mediana"))
title(main="Boxplot das estimativas para alfa.suj.1")

boxplot(media[,28], moda[,28], mediana[,28], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.suj.1")

boxplot(media[,9], moda[,9], mediana[,9], names=c("média", "moda", "mediana"))
title(main="Boxplot das estimativas para alfa.suj.9")

boxplot(media[,36], moda[,36], mediana[,36], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.suj.9")

boxplot(media[,19], moda[,19], mediana[,19], names=c("média", "moda",
"mediana"))

```

```

title(main="Boxplot das estimativas para alfa.suj.19")

boxplot(media[ ,46], moda[ ,46], mediana[ ,46], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.suj.19")

windows()
par(mfrow=c(3,2))

boxplot(media[ ,55], moda[ ,55], mediana[ ,55], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para alfa.t1")

boxplot(media[ ,58], moda[ ,58], mediana[ ,58], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.t1")

boxplot(media[ ,56], moda[ ,56], mediana[ ,56], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para alfa.t2")

boxplot(media[ ,59], moda[ ,59], mediana[ ,59], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.t2")

boxplot(media[ ,57], moda[ ,57], mediana[ ,57], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para alfa.t3")

boxplot(media[ ,60], moda[ ,60], mediana[ ,60], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para beta.t3")

windows()
par(mfrow=c(3,1))

boxplot(media[ ,61], moda[ ,61], mediana[ ,61], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para sigma2alfa")
boxplot(media[ ,62], moda[ ,62], mediana[ ,62], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para sigma2beta")
boxplot(media[ ,63], moda[ ,63], mediana[ ,63], names=c("média", "moda",
"mediana"))
title(main="Boxplot das estimativas para sigma2e")

```