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

Modelos Hierárquicos Bayesianos aplicados ao Marketing

Autor: Bárbara Pederiva
Orientador: Professora Dra. Patrícia KlarmannZiegelmann

Porto Alegre, 11 de Dezembro de 2013.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

Modelos Hierárquicos Bayesianos aplicados ao Marketing

Autor: Bárbara Pederiva

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professora Dra. Patrícia KlarmannZiegelmann (orientadora)
Professora Dra. Vanessa BielefeldtLeottiTorman

Porto Alegre, 11 de Dezembro de 2013.

Agradecimentos:

Foram muitas as mãos que ajudaram na construção deste trabalho e no conhecimento adquirido ao longo destes quatro anos dentro desta instituição, por isto, neste momento, tenho muito a agradecer.

Agradeço então aos meus pais, por todos os ensinamentos transmitidos ao longo da minha vida, sempre lembrando que uma das poucas coisas que ninguém pode nos tirar é o conhecimento; da mesma forma, não posso deixar de agradecer a minha irmã, Daniela, que mais do que uma irmã, é uma segunda mãe. Já que estamos falando em família, agradeço também as tias, pelos chimarrões em domingos a tarde e pelas palavras de apoio.

Pelos conhecimentos transmitidos, a todos os professores por quem passei desde minha formação básica, em especial, aos de matemática, por terem despertado o interesse pela área.

A estatística Karina Pretto pelo tempo dedicado em 2009, explicando um pouco sobre a Gradução em Estatística, bem como contar um pouco de sua carreira profissional, fazendo com que meu interesse fosse ainda maior.

Aos professores da Universidade Federal do Rio Grande do Sul por todos os conhecimentos, em especial aos professores do Departamento de Estatística, que além de cumprirem muito bem com seus papéis de professores, serviram também como apoio, foram compreensivos e deram-me oportunidades na carreira acadêmica e profissional; gostaria de citar aqui a Prof. Dra. Patrícia Ziegelmann, minha orientadora por toda dedicação e disponibilidade incessante.

Aos meus amigos, tendo surgido antes ou durante a graduação, pelos momentos de descontração, desabafo, compreensão e até pelos "acende a luz". Em especial, a Natália, Amanda, Flávia e Maiara, minhas irmãs porto alegrenses, meu geólogo e melhor amigo dos almoços pelo vale, Daniel, e meus companheiros de curso: Tiago, Douglas e Suziane, por todos intervalos ao sol. A Natália Giordani, primeira companheira de trabalho e aos colegas do Sicredi por todo apoio durante o estágio. A Marília e Danielle pelos finais de semana descontraídos e a Rafaela e a Luiza com suas palavras de apoio mesmo que a distância.

Enfim, a todos que de uma forma ou outra passaram pela minha vida auxiliando no meu crescimento pessoal ou profissional, muito obrigada!

Resumo

Em marketing é comum utilizar dados de painel, ou seja, dados em que se têm mais de uma observação para cada unidade amostral os estudos aplicados a esta área possuem objetivos ligados a analisar duas variações de forma conjunta: dentro e entre as unidades amostrais. Deste modo, modelos mais complexos são necessários para melhorar compreensão do comportamento cliente/produto. Assim, vem se tornando comum em marketing o uso de Estatística Bayesiana.

Modelos Hierárquicos Bayesianos conseguem trabalhar com os dois principais objetivos dos estudos de marketing: em um primeiro nível são modeladas as variações de produto e no segundo nível da hierarquia as variações entre consumidores. Outro benefício ao utilizar modelos bayesianos em estudos de marketing está em utilizar prioris de mistura, que conseguem captar melhor as variações de dados heterogêneos ou multimodais, por exemplo.

O objetivo deste trabalho é apresentar a abordagem bayesiana de modelos hierárquicos aplicados ao marketing. Pretende-se que este texto sirva como material introdutório voltado aos profissionais da área de marketing.

Para ilustrar os conceitos bayesianos aplicados ao marketing foram apresentados dois estudos de caso: o primeiro o uso de um modelo linear hierárquico bayesiano e o segundo um modelo multinomial hierárquico bayesiano. Estes são usados para auxiliar na compreensão do comportamento de volume de vendas e na probabilidade de compra.

Os modelos foram ajustado utilizando o pacote *bayesm* do software R, que é um pacote específico para modelagem bayesiana na área de marketng. Ao longo deste trabalho todos os comandos do *bayesm* necessários para o ajuste do modelo são apresentados.

Palavras-chave: Marketing, modelo linear hierárquico bayesiano, modelo multinomial hierárquico bayesiano, priori de mistura, dados de painel de mistura, dados de painel.

Abstract

In marketing is common to use panel data, in other words, data that have more than one observation for each sample unit. The studies applied to this area have objectives linked to analyze two variants jointly: within and between sampling units. Thus, more complex models are needed to improve understanding of customer/product behavior. So, it has become common in marketing the use of Bayesian Statistics.

Hierarchical Bayesian models can work with the two main objectives of marketing research: on a first level are modeled the product variations and on the second level of the hierarchy, the changes between consumers. Another benefit to using Bayesian models in marketing research is to use mixture models for the prior knowledge, which can better capture changes in heterogeneous or multi-modal data, for example.

The objective of this paper is to present Bayesian hierarchical models applied approach to marketing. It is intended that this text will serve as introductory material geared to professionals in the field of marketing.

To illustrate the Bayesian concepts applied to marketing were presented two case studies: The first use of a Bayesian hierarchical linear model and the second a Bayesian hierarchical multinomial model. These are used to assist in understanding the behavior of sales and purchase likelihood.

The models were adjusted using the R software package Bayesian , which is a specific package for Bayesian modeling in the area of marketing. Throughout this paper all Bayesian commands needed to fit the model are presented.

Keywords : Marketing , Bayesian hierarchical linear model , Bayesian hierarchical multinomial model , priori mixing , blending panel data , panel data .

Sumário

1. Introdução

Os estudos em marketing tem se desenvolvido muito ao longo dos anos, por exemplo, Schmittleinet al. (1989), pretendia estabelecer a probabilidade de uma base de clientes permanecer ativa tendo, para isto, informações consolidadas de toda a base de clientes. Porém, nestes estudos, embora já fosse de conhecimento que clientes t,eriam diferentes comportamentos devido suas características individuais estas informações ainda não eram consideradas. Nos anos 2000, Fader et al (2004), já consideravam diferenças entre os clientes fazendo com que a complexidade dos modelos fosse aumentada, em contra ponto, era também aumentada sua aplicabilidade. Kumar e Shah (2009) passaram a considerar além das diferenças entre clientes, as diferenças de um mesmo cliente.

Atualmente, de forma geral, pode-se dizer que os principais objetivos dos estudos de marketing estão em compreender as variações de produto e de consumidor de forma conjunta. Desta maneira modelos mais complexos são necessários para melhorar a compreensão do comportamento cliente/produto. Assim, vem se tornando comum em marketing o uso de Estatística Bayesiana (Rossi, 2005).

Modelos hierárquicos bayesianos tem grande importância para o marketing considerando os dados a serem utilizados são dados de painel, ou seja, têm-se mais de uma observação para cada unidade amostral e, deste modo, deseja-se analisar duas variações: dentro e entre as unidades amostrais de forma conjunta.O modelo hierárquico, através de dois níveis de hierarquia, consegue atingir os dois principais objetivos dos estudos de marketing, em um primeiro nível são modeladas as variações de produto e no segundo nível da hierarquia as variações entre consumidores.

Modelos bayesianos utilizam informações provenientes de estudos anteriores ou de conhecimentos de *experts* do assunto para a modelagem, este conhecimento é definido como priori e juntamente com a verossimilhança das informações amostradas é então encontrada a posteriori, resultado da inferência bayesiana. Como nem sempre é possível encontrar a distribuição a posteriori, métodos de simulação como o MCMC – Método de Monte Carlo via Cadeias de Markov – são utilizados a fim de encontrar uma amostra da distribuição a posteriori e, com base nesta amostra estimar os parâmetros desconhecidos do modelo. Existem diferentes formas de definir a informação a priori, inclusive considerando situações em que esta informação não é conhecida. A mistura de distribuições também pode ser considerada como uma forma de definição da priori (Congdon, 2001), embora não seja muito usual. Prioris de mistura auxiliam na modelagem de dados multimodais e provenientes de populações heterogêneas; diferente de modelos que contam com prioris sem mistura, estes permitem uma maior variabilidade de resultados, expressando assim de melhor forma o que acontece na prática.

O presente estudo tem como objetivo apresentar a abordagem bayesiana de modelos hierárquicos aplicados ao marketing. Pretende-se que este texto sirva como material introdutório voltado aos profissionais desta área. Para exemplificar como estes conceitos podem ser aplicados ao marketing serão expostos dois estudos de caso que serão implementados utilizando o pacote *bayesm* do software R. Os comandos utilizados na implementação são também apresentados e a incorporação de prioris de mistura é ilustrada.

O primeiro estudo de caso exemplifica a utilização de um modelo linear hierárquico bayesiano com a finalidade de entender a variabilidade de queijo vendido em diversos estabelecimentos através das informações de quanto do produto é exposto e seu respectivo preço. O segundo estudo de caso ilustra a utilização de um modelo logístico multinomial hierárquico bayesiano para entender a probabilidade de compra de diversas marcas de margarina por diferentes famílias dadas algumas informações demográficas delas e os preços das margarinas nas diferentes situações de compra.

A abordagem bayesiana ainda não é muito conhecida e difundida em diversas áreas do conhecimento, sua aplicação envolve métodos iterativos e métodos de simulação, MCMC. Desta forma, diversos conceitos da estatística bayesiana precisam ser compreendidos para melhor entendimento dos estudos de caso propostos ao longo deste trabalho.

Este trabalho está estruturado da seguinte forma: o Capitulo 2 traz algumas informações a respeito de estudos em marketing, o Capitulo 3 apresenta conceitos fundamentais de inferência bayesiana com especial enfoque a modelos hierárquicos, breve discussão a respeito de prioris e entendimento do MCMC. O Capitulo 4 apresenta o pacote B*ayesm* e no Capítulo 5 e no Capítulo 6 são apresentados os dois estudos de caso. No Capitulo 7, são apresentadas as considerações finais do trabalho.

2. Estudos na área de Marketing

2.1. Dados de Painel

Define-se por dados de painel, as situações onde são coletadas diversas observações de uma mesma unidade amostral ordenadas ou não no tempo.

Devido a natureza dos dados de painel é comum que exista heterogeneidade entre as unidades amostrais e, dependência entre as observações de uma mesma unidade amostral.

2.2. Exemplo 1: Volume de vendas dequeijo

Este estudo tem por objetivo principal entender o comportamento do volume de vendas de queijo baseados no percentual de queijo exposto e seu preço de venda. Os dados observados compreendem 5.555 observações de vendas oriundas de 88 diferentes estabelecimentos, conforme ilustra o quadro abaixo.

Quadro 1: Exemplificação dos dados apresentados com informações de: ID, identificação da observação; Estabelecimento, indicação de unidade amostral, Volume de venda de queijo, Percentual de queijo exposto e Preço.

ID	Estabelecimento	Volume de venda	Percentual exposto	Preço
1	1	778	0	3,052699
2	1	695	0	3,053237
⋮	⋮	⋮	⋮	⋮
61	1	794	0,1695	2,623426
62	2	4.099	0,0236	2,893145
⋮	⋮	⋮	⋮	⋮
5.554	88	1.067	0	3,186504
5.555	88	1.053	0	3,244065

Como informações para entender este comportamento tem-se a informação de quanto queijo é vendido em cada observação, em unidades, de quanto queijo do estabelecimento é exposto para venda naquela situação, dada em percentual, e qual o seu respectivo preço, dado em dólares.

Dado que se tem mais de uma informação para cada estabelecimento pode-se então dizer que estes dados são considerados como dados de painel, já que serão utilizadas diversas observações de cada uma das unidades experimentais.

2.3. Exemplo 2: Compra de diversas marcas de margarina

Este estudo tem por objetivo estudar a probabilidade de compra de algumas marcas de margarina, para isto, foram observadas 4.470 operações de compra de dez diferentes marcas de margarina realizadas por 516 diferentes famílias. Os dados observados estão exemplificados nos dois quadros abaixo.

Quadro 2: Exemplificação dos dados de primeiro nível, apresentando informações de: ID, identificação da observação; Família, indicação de unidade amostral, Compra, indicação de qual a marca comprada, Marca1 a Marca10, respectivo preço de cada uma das 10 marcas de margarina.

ID	Família	Compra	Marca1	Marca2	Marca3	Marca4	...	Marca10
1	2100016	1	0,58	0,67	1,09	0,57	...	0,33
2	2100016	1	0,19	0,67	0,99	0,57	...	0,37
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮
987	2112243	1	0,99	0,5	0,99	0,57	...	0,67
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮
4470	2157248	2	0,58	0,19	0,99	0,29	...	0,56

Quadro 3: Exemplificação dos dados de segundo nível, apresentando informações de Família, Renda e Número de pessoas.

Família	Renda	N º pessoas
2100016	32,5	2
2100024	17,5	3
⋮	⋮	⋮
2157248	12,5	2

O Quadro 2, ilustra as 4.470 observações de compra, é obtida a informação de qual família fê-la, qual a margarina escolhida, bem como o preço de cada uma das dez margarinas avaliadas. O Quadro 3 as informações a respeito das famílias que participaram do estudo, têm-se então, informações demográficas, de renda, número de pessoas, se há estudantes, empregados e aposentados, além de variáveis para famílias de mais de 5 pessoas e famílias de 3 ou 4 pessoas.

Pelo Quadro 2, se sabe que existe mais de uma informação para cada família, deste modo, considera-se estes dados como dados de painel, já que foram analisadas diversas observações de cada uma das unidades experimentais.

3. Inferência Estatística Bayesiana

A ideia básica da inferência estatística é pensar que os dados provenientes de uma amostra são resultados de um experimento aleatório e então utilizá-los para inferir sobre medidas desconhecidas de uma população. Existem duas escolas de inferência, a clássica e a bayesiana, a principal diferença entre elas é que, do ponto de vista clássico, o parâmetro desconhecido θ- escalar ou vetor – é uma quantidade fixa, enquanto, do ponto de vista bayesiano, o parâmetro θ, por ser desconhecido é considerado quantidade aleatória, ou seja, por ser não observável, possui uma distribuição de probabilidade associada.

A inferência bayesiana utiliza informações a Priori, que indicam o conhecimento prévio do parâmetro em questão e a verossimilhança, informação obtida através dos dados amostrados, para encontrar a distribuição a Posteriori.

A distribuição a posteriori para o parâmetro desconhecido θ, é obtida através do Teorema de Bayes da seguinte forma:

$$\pi(\theta|x) = \frac{P(X = x|\theta) \times P(\theta)}{P(X = x)} = \frac{L(\theta|x) \times \pi(\theta)}{f(x)} = \frac{L(\theta|x) \times \pi(\theta)}{\int L(\theta|x) \times \pi(\theta) d\theta}.$$

Note que, na expressão acima, $f(x)$ é uma constante em relação aθ. Desta forma:

$$\pi(\theta|x) \propto L(\theta|x) \times \pi(\theta),$$

$$Posteriori \propto Verossimilhança \times Priori.$$

Ou seja, a distribuição a posteriori é proporcional a verossimilhança dos dados e a informação a priori. Neste caso, fazendo uso sucessivo da técnica, ao serem amostrados novos dados, a informação da posteriori obtida anteriormente pode ser utilizada como informação a priori nesta nova etapa.

3.1. Priori

Em estatística Bayesiana, priori é a informação que se tem a respeito do parâmetro θa ser estudado antes da observação de dados de uma amostra. A informação a priori pode ser obtida através de estudos anteriores ou conhecimento de profissionais da área, por exemplo.

3.1.1. Priori Conjugada:

O modo mais comum de especificar informações a priori para os parâmetros é utilizar distribuições de probabilidade pertencentes a famílias de distribuições conhecidas. Utilizando-as é necessário definir todos os seus parâmetros, deixando-a inteiramente especificada já que trata-se de uma informação conhecida.

Por exemplo, a priori, tem-se que, um determinado θ, referente a média de uma população, segue distribuição Normal, deste modo, é necessário, definir uma média e uma variância para esta priori, estes valores são conhecidos como hiperparâmetros.

Deve-se levar em conta que a função da priori deve ser combinada com a função de verossimilhança para encontrar a distribuição a posteriori dos parâmetros. Uma forma de tornar matematicamente possível esta combinação é através do uso de prioris conjugadas.

Uma priori é chamada de conjugadaa um determinado modelo se a distribuição a priori e a posteriori pertencema mesma classe de distribuições. Assim a atualização do conhecimento que se tem de θ envolve apenas uma mudança nos hiperparâmetros.

3.1.2. Priori Não-informativa:

O não conhecimento de informações a priori a respeito do parâmetro desconhecido θ não impossibilita o uso de inferência bayesiana, nestes casos, prioris não informativas são utilizadas.

Um exemplo clássico para quando não existe informação a priori é o uso de uma distribuição Uniforme, desta forma, todos os possíveis valores para o parâmetro são igualmente prováveis. Porém, se o intervalo de definição de θ for ilimitado teremos uma distribuição imprópria, já que sua distribuição acumulada poderá ter valor maior de 1.

3.1.3. Priori de Mistura:

Modelos de mistura típicos são aqueles em que a função de probabilidade é descrita como uma combinação linear de várias distribuições. Esta classe de modelos surge em dois contextos:

- População heterogênea, constituída por grupos predefinidos, porém desconhecidos,
- Representação conveniente da densidade de probabilidade, ou seja, quando as distribuições standard não são capazes de se ajustar suficientemente bem aos dados, um exemplo claro é quando a distribuição dos dados não é unimodal.

Para exemplificar as prioris de mistura, a Figura 1 mostranos histogramas os dados a priori para a média de altura de duas populações distintas, o primeiro histograma mostra a especificação da primeira componente da Priori: Normal(1.9,0.05), correspondente a altura média de Americanos, o segundo, a segunda componente: Normal (1.6,0.07), representando a altura média de orientais. O terceiro histograma mostra a mistura das duas distribuições.

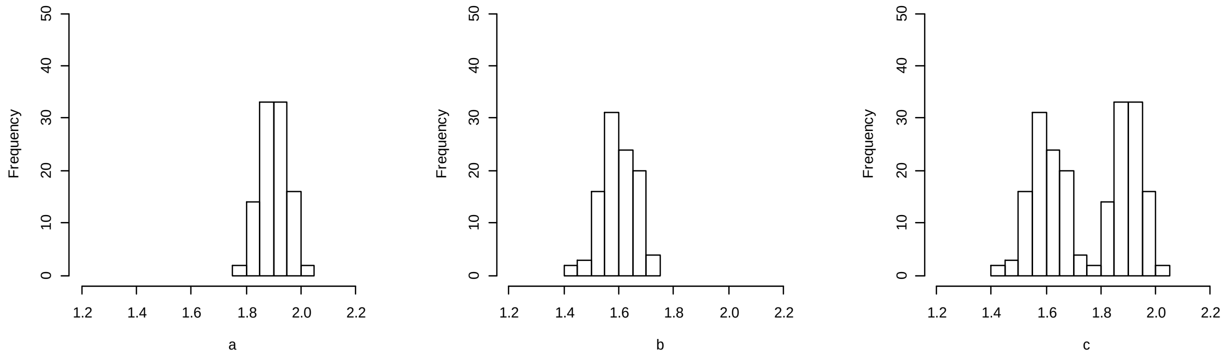


Figura 1: Mistura de duas Normais

3.2. Modelos hierárquicos:

Modelos bayesianos hierárquicos são apropriados para modelar problemas com estruturas complexas de dependência, tais como os dados apresentados nos casos de Queijo e Margarina.

Modelos hierárquicos podem ser generalizados como modelos em que os parâmetros do modelo que descrevea variabilidade da variável de interesse são considerados quantidade aleatória oriundas de uma distribuição de probabilidade cujos parâmetros são chamados de hiperparâmetros, a distribuição dos parâmetros a priori possui hiperparâmetros, que são estimados pelos dados. Um modelo hierárquico bayesiano pode ser representado considerando, pelo menos,3 níveis de hierarquia:

- 1º nível: modelo das observações: de X com os parâmetros;
- 2º nível: modelo dos parâmetros com seus hiperparâmetros;
- 3º nível modelo dos hiperparâmetros.

Para Rossi (2005), os três níveis são descritos da forma similar, apresentada abaixo:

- 1º nível: especificação da função de verossimilhança: $p(y_i|\theta_i)$;
- 2º nível: especificação das funções de ligação: $p(\theta_i|\tau)$;
- 3º nível: especificação das prioris dos parâmetros: $p(\tau|h)$.

Neste caso, a posteriori é dada por:

$$p(\theta_1, \dots, \theta_m | y_1, \dots, y_m, h) \propto \left[\prod_i p(y_i|\theta_i)\, p(\theta_i|\tau) \right] \times p(\tau|h).$$

3.3. Monte Carlo via cadeias de Markov:

Os métodos mais conhecidos e utilizados em bayesiana são conhecidos por simulação de Monte Carlo via cadeias de Markov (MCMC). A ideia é obteruma amostra da distribuição a posteriori sem conhecer exatamente qual é esta distribuição, utilizando para isto, a informação a priori e a verossimilhança, e calcular estimativas amostrais de características desta distribuição.

O MCMC é um método de simulação interativa, baseado em cadeias de Markov. Devido a esta característica os valores simulados não são independentes. Deste modo, podem ser utilizadosalguns artifícios a fim de obter valores independentes e identicamente distribuídos (iid).

Uma cadeia de Markov é um processo estocástico, tal que a distribuição de um dado X_t dados todos os valores anteriores, depende apenas do último anterior (X_{t-1}). Os métodos de MCMC exigem que a cadeia seja:

- Homogênea: ou seja, as probabilidades de transição de um estado para outro são invariantes;

- Irredutível, ou seja, que cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações;

- Aperiódica, que não haja estados absorventes.

Como resultado do MCMC tem-se uma amostra da posteriori, sob a qual é possível realizar inferências.

3.3.1. Amostrador de Gibbs:

Um dos métodos de MCMC é o Amostrador de Gibbs, que é o método adequado para os modelos hierárquicos apresentados neste trabalho.

A idéia básica do amostrador de Gibbs é tornar um problema multivariado numa sequência de problemas univariados; entre os quais itera-se para produzir uma cadeia de Markov. A distribuição de equilíbrio é a distribuição *a posteriori* desejada (Cassella e George, 1992).

O método de GibbsSampling cria uma cadeia de Markov a partir de uma distribuição conjunta. Se duas variáveis aleatórias possuem uma determinada densidade conjunta, a partir das densidades condicionais é gerada a cadeia de Markov.

O algoritmo torna-se simples quando são conhecidas as distribuições condicionais e a distribuição conjunta é estacionária. A convergência da cadeia é assegurada desde que os suportes das condicionais estejam ligados.

GibbsSampling só pode ser aplicado quando todas as distribuições condicionais completas são conhecidas, que é o caso dos estudos apresentados neste trabalho.

O amostrador de Gibbs é considerado um complemento computacional para modelos hierárquicos em que a distribuição conjunta é especificada através de sucessivos estratos de distribuições condicionais. A Figura 2 exemplifica o algoritmo usado para a implementação do GibbsSampling.

1º Gerar um valor inicial de cada um dos parâmetros desejados.

$\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_d^{(0)})'$

2º Obter um novo valor de $\theta^{(t)}$ *a partir de* $\theta^{(t-1)}$ *através de sucessivos valores gerados.*

$$\theta_1^{(t)} \sim \pi\ (\theta_1 | \theta_2^{(t-1)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)})$$
$$\theta_2^{(t)} \sim \pi\ (\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)})$$
$$\vdots$$
$$\theta_d^{(t)} \sim \pi\ (\theta_d | \theta_1^{(t)}, \theta_2^{(t)}, \ldots, \theta_{d-1}^{(t)})$$

3º Incrementar o contador t para t +1 e realizar sucessivas vezes o passo 2 até obtenção da convergência.

Figura 2: Algoritmo de Giibs Sampling

3.3.2. Diagnóstico de validade dos resultados:

Como os valores gerados são resultado de uma simulação, algumas características precisam ser verificadas para validar os resultados encontrados. Para o diagnóstico dos resultados, devem ser monitoradas três questões:

- Amostragem aproximadamente independente e identicamente distribuída (i.i.d.).
- Convergência para a estacionariedade;
- Convergência ergódica para as médias.

Para isto então devem ser avaliados três pontos: *Burn in, thin*e *número de iterações*.

Burn in: “período de aquecimento” da cadeia, as primeiras simulações da cadeia ainda não atingiram convergência d modo que não podem ser consideradas observações iid da posteriori. Um certo número de simulações iniciais precisa ser descartado.

Thin: é o espaçamento entre valores simulados que será utilizada para formar a amostra da posteriori. Este espaçamento é necessário para eliminar a correlação entre os valores a serem utilizados e desta forma assegurar independência entre os valores gerados.

Número de iterações: sabendo que será desprezado um determinado número de valores iniciais e que os dados a serem utilizados são espaçados uns dos outros, precisa-se assegurar também que existirá um número razoável de valores amostrados aptos para análise a fim de conseguir obter certa precisão de resultados.

4. O Pacote: Bayesm

O pacote *bayesm* – Inferência Bayesiana para Marketing e Microeconomia – foi desenvolvido e é mantido por Peter Rossi, professor de Marketing, Economia e Estatística na Anderson Schoolof Management, UCLA.

Todo o pacote de R, software livre, disponível em http://www.r-project.org/, é um conjunto de funções, conjuntos de dados e documentação que pode ser facilmente instalado e atualizado a partir do R. O Pacote em questão apresenta 66 tópicos, entre eles, funções e conjuntos de dados.

As funções apresentadas pelo pacote são as mais variadas que vão desde a preparação e formatação de dados para uso em funções específicas, que é o caso da função createX, que nada mais é do que um uma rotina para tornar uma matriz X, na forma prevista pela função LogitMultinomial e Probit, também apresentadas no pacote. Passando por funções, como o exemplo da rWishart, que gera valores de uma Distribuição de Wishart, que é comumente utilizada como priori para o Desvio Padrão em modelos multivariados. Até chegar em funções com um maior nível de complexidade, como é o caso da rhierLinearMixture que implementa o amostrador de Gibbs para um modelo linear hierárquico com mistura de prioris.

Dentre os modelos possíveis de serem gerados através do pacote, temos:

- Regressão bayesiana (uni ou multivariada dep var);
- Regressão aparentemente independentes bayesiana (SUR);
- Logitmultinomial (MNL), ProbitMultinomial (MNP), Probit multivariada;
- Binomial (Poisson) Regressão negativo;
- Misturas de normais multivariadas (incluindo agrupamento);
- Modelos Lineares hierárquicos com uma prévia e covariáveis normal;
- Modelos Lineares hierárquicos com uma mistura de normais antes e covariáveis;
- Logitmultinomial hierárquico com uma mistura de normais antes e covariáveis;
- Hierárquicos binomial negativa Modelos de Regressão;
- Análise Bayesiana de dados conjuntas baseadas em escolha;
- Tratamento Bayesiana de modelos lineares de variáveis instrumentais;
- Analise de levantamento de dados ordinais multivariadas, com heterogeneidade uso de escala.

O pacote também apresenta uma série de bancos de dados para auxiliar na compreensão de que tipos de dados são necessários para trabalhar com as funções em questão, deste modo, os bancos presentes no pacote são:

- Bank - dados de pesquisa sobre as preferências de cartão de crédito para 946 respondentes.
- Cheese - Os dados do painel com um volume de vendas de queijo assim como uma medida da atividade de exibição e preço deste em estabelecimentos.
- CustomerSat - dados da pesquisa de satisfação dos clientes de um produto de publicidade.
- Detailing- dados em painel sobre médico prescritor e visitas de vendas para uma determinada droga.
- Margarine - Dados em painel sobre as compras de margarina de 516 famílias.
- Scotch - levantamento de dados sobre as preferências para as marcas de Uísque.
- Tuna - Volume de vendas de atum em conserva, bem como uma medida da atividade de exibição do produto.

Através destes bancos de dados, e de algumas das funções presentes no *Bayesm*, foram, portanto, implementados estudos de caso em questão.

5. Estudo de caso: aplicação dos dados de vendas de queijo

Conforme apresentado na Seção 2.2, serão utilizados os dados do exemplo para a obtenção de um modelo que explique o comportamento do volume de vendas de queijo baseados no percentual de queijo exposto no estabelecimento e seu respectivo preço de compra.

5.1. Modelo:

A variável resposta de interesse (Y: volume de venda de queijo) é quantitativa e observada um certo número de vezes para cada estabelecimento.

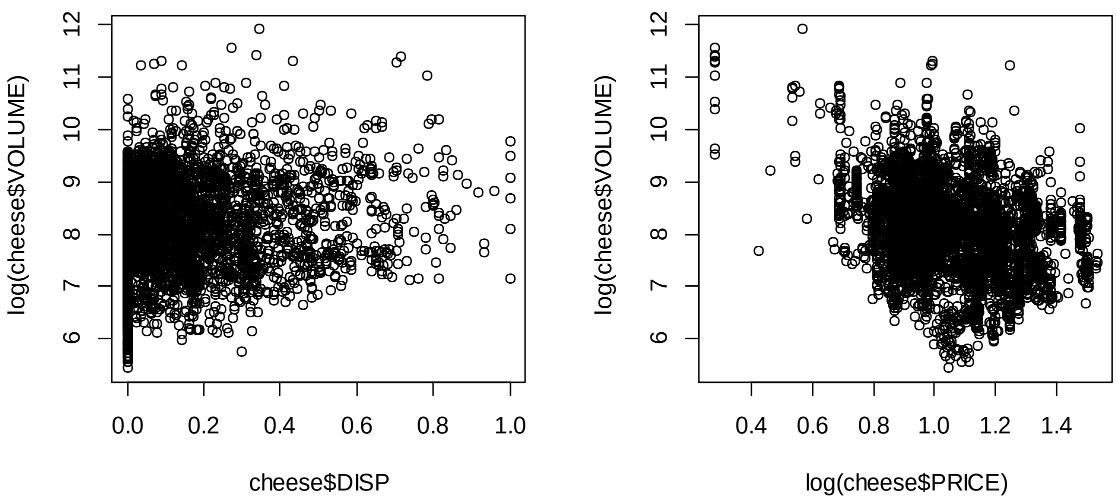


Figura 3: Gráfico de dispersão da variável resposta (Y: Volume) com as covariáveis percentual exposto e preço, respectivamente

Portanto, analisando a Figura 3, será ajustado um modelo de regressão Linear Hierárquica , para este modelo deve-se atender a suposição de normalidade da variável resposta, para isto, é observada sua distribuição, bem como entendido um pouco o comportamento das demais variáveis através de uma análise descritiva:

Volume da venda de queijo: apresenta uma grande amplitude de possíveis valores, cujo mínimo é de 231 unidades e o máximo é de 148.109, o valor médio é de 4.771 e o desvio padrão de 5.967,92, a distância interquartílica, entretanto é de apenas 3.530, e então pode-se constatar uma grande presença de valores atípicos, que podem ser observados através da Figura 4 que traz o histograma dos dados.

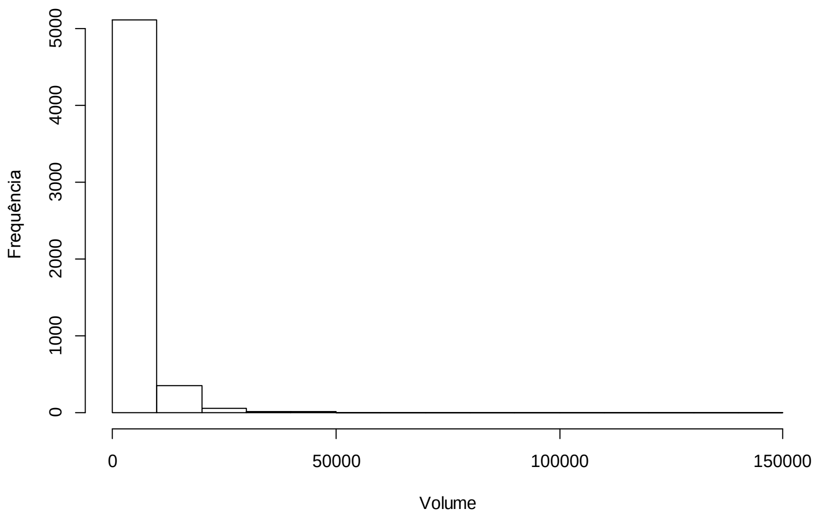


Figura 4: Histograma do Volume de venda de queijo

O histograma apenas comprovou o que as medidas descritivas já indicavam, os dados, visivelmente, não possuem distribuição Normal, desta forma, é proposta uma transformação logarítmica, que não altera a ordenação dos dados, mas diminui sua amplitude; a Figura 5 mostra então o histograma dos dados transformados, agora com uma distribuição que, visivelmente, se aproxima da distribuição Normal.

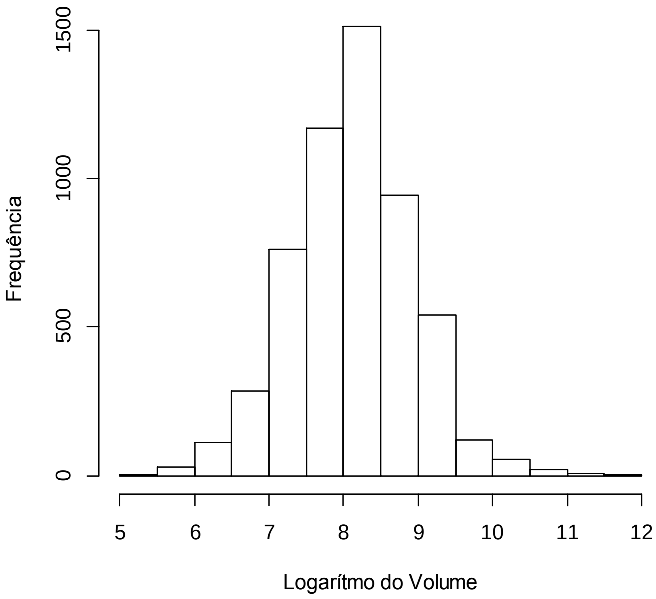


Figura 5: Histograma da transformação logarítmica do Volume de venda de queijo

Percentual de queijo exposto no estabelecimento: por ser dada em percentual, esta variável não possui problemas como a anterior, sabemos que seu mínimo é 0 e seu máximo é 1, sua média é de 0,11 e seu desvio padrão de 0,16.

Preço do queijo: em uma análise geral, esta variável não apresenta problemas, possui média e mediana muito semelhantes de 2,869 e 2.703, respectivamente, e desvio padrão de 0,5317, porém, como será proposto um modelo por estabelecimento. A fim de melhorar a interpretação prática do modelo e sabendo que isto não ocasiona nenhum prejuízo, é aplicada a transformação logarítmica também a esta variável.

O modelo proposto será:

1º Nível: modela a variabilidade dentro de cada um dos 88 estabelecimentos através de um modelo linear.

$$\ln(y_{ij}) = \beta_o + \beta_{1i} x_{1i} + \beta_{2i} ln(x_{2i}) + e_i,$$

em que, y_{ij} é a variável dependente, no caso, o volume de queijo vendido na j-ésima observação do estabelecimento *i*, x_{1j} o percentual de queijo exposto para venda na observação *j*, e x_{2j} a informação do preço do queijo na observação *j*, de modo que *i* vai de 1 a 88 (número de estabelecimentos) e *j* de 1 a *ni*, número de observações do estabelecimento *i*, $e_i \sim Normal\ (0, tau_i)$.

2º Nível: modela a variabilidade entre os 88 estabelecimentos estudados através de um modelo linear.

$\beta_{ik} = \delta_{ik} z_{ik}$

em que, β_i é cada um dos *i* coeficientes dos 88 estabelecimentos, em que *k* assume valores de 1 a 3, correspondendo aos coeficientes β_o, β_1e β_2 da equação anterior, δ_{ik} é o coeficiente de cada um dos *k* valores de β, em cada um dos *i* estabelecimentos.

O segundo nível do modelo poderia contar com covariáveis, informações que descrevem os estabelecimentos, como: faturamento, número de funcionários, tamanho do estabelecimento; estas informações poderiam auxiliar, por exemplo, no momento de explicar a variabilidade do preço de queijo (uma grande empresa consegue comprar maiores quantidades de queijo e com isso oferecer um preço mais atrativo para o consumidor, por exemplo). Deste modoos parâmetros β_k poderiam ser modelados utilizando, por exemplo, outro modelo linear, isto será exemplificado no próximo capítulo.

Neste estudo, por não existirem informações a respeito dos estabelecimentos os δ_{ik} são iguais aos β_{ik} e, deste modo, este nível do modelo é estabelecido apenas para formalização, neste caso, diferente do que será exposto no próximo capítulo.

3º Nível: são apresentados os valores a priori para os parâmetros do modelo, deste modo, serão utilizadas prioris conjugadas, não informativas, em que a média dos parâmetros β será dada pela distribuição Normal e a variância pela Inversa Wishart. A seguir, será mostrado como se indica, na função, estas proris.

5.2. Implementação do Modelo

O modelo linear hierárquico é facilmente implementado utilizando a função rhierLinearModel do pacote bayesm. A seguir é apresentado um passo a passo, onde, nos retângulos em destaque, são apresentados todos os comandos utilizados para a implementação do modelo.

Para utilizar esta função é necessário a definição de três argumentos: *Data*, *Prior*, *Mcmc*. O argumento Data compreende a uma lista composta por dois elementos: *regdata* e *Z*, sendo o primeiro a matriz de dados com informação de y, variável dependente, e X covariáveis; o segundo elemento compreende a uma matriz de observações. O argumento Prior compreende a uma lista composta por cinco elementos: *Deltabar*, *A*, *nu.e*, *ssq*, *nu* e *V*. E por último, o argumento Mcmc traz uma lista de dois elementos: *R* e *keep*. Como resultado da função tem-se os valores gerados pelo MCMC para cada um dos parâmetros do modelo: *Betadraw*, *Taudraw*, *Deltadraw*, *Vbetadraw*. Para fazer uso da função, precisamos então alimentá-la da maneira descrita anteriormente, deste modo:

1º Passo: definir o argumento Data

O argumento Data será dado por *regdata*, que será composto das informações da variável dependente e das covariáveis do modelo (x_i); por termos dados em painel, ou seja, mais de uma informação de *y* e *x* para cada unidade experimental (U.E.), precisamos organizar os dados de forma que tenhamos identificação de qual a U.E. cada observação pertence, então, para cada U.E., teremos uma lista de valores de *y* e outra lista com os valores de *x*, devendo esta conter uma coluna de valor "1", constante, para que possamos também estimar um valor para o intercepto do modelo. A segunda informação do argumento Data é *Z*, sendo esta, opcional, caso não tenhamos um segundo nível de informação, como acredita-se que possa haver variação de vendas entre os estabelecimentos, definimos Z como um vetor de valor “1” com 88 observações, ou seja, uma para cada estabelecimento.

```
retailer=levels(cheese$RETAILER)
nreg=length(retailer)
nvar=3
regdata=NULL
for (reg in 1:nreg) {
y=log(cheese$VOLUME[cheese$RETAILER==retailer[reg]])
iota=c(rep(1,length(y)))
X=cbind(iota,cheese$DISP[cheese$RETAILER==retailer[reg]],
log(cheese$PRICE[cheese$RETAILER==retailer[reg]]))
regdata[[reg]]=list(y=y,X=X)
}

Z=matrix(c(rep(1,nreg)),ncol=1)
nz=ncol(Z)

lscoef=matrix(double(nreg*nvar),ncol=nvar)
for (reg in 1:nreg) {
coef=lsfit(regdata[[reg]]$X,regdata[[reg]]$y,intercept=FALSE)$coef
if (var(regdata[[reg]]$X[,2])==0) { lscoef[reg,1]=coef[1]; lscoef[reg,3]=coef[2]}
else {lscoef[reg,]=coef }
}

Data=list(regdata=regdata,Z=Z)
```

2º Definir o argumento *Prior:*

O argumento prior estabelece valores a priori para parâmetros do modelo, sendo todas informações opcionais, como não se tem informações a priori, não foram indicados valores iniciais, deixando então os valores *default* do pacote, ou seja, utilizar-se-á prioris não informativas para: Deltabar, priori para média, no caso, igual a 0; A, priori para informação de precisão, default em 0,011, nu.e, parâmetro para a priori do erro da variância (indicado por 3), ssq, como parâmetro de escala para a priori do erro, igual a var(y_i), nu priori para a variância dos β, como padrão dado pelo número de variáveis mais três; e V, matriz de locação para variância dos β, por definição, dado como nu*I.

3º Passo: definir o argumento Mcmc:

O argumento Mcmc estabelece o número de iterações da cadeia, elemento *R* da função e o *thin*, elemento *keep* a ser utilizado pelo algoritmo do MCMC. Como já visto, estes valores devem ser definidos de modo a garantir as propriedades básicas do MCMC. Para definir estes valores vamos primeiramente estabelecer (keep=1, R=2000) e avaliar os principais parâmetros do modelo, no nosso caso, os valores gerados para: β_0, β_1 e β_2.

```
R=2000
Keep = 1
Mcmc=list(R=R, keep=Keep)
```

4º Passo: ajustar o modelo piloto:

Conforme orientações dadas nos passos anteriores, define-se então o primeiro modelo de acordo com os comandos abaixo, cabe destacar que, por se ter deixado todos os parâmetros a priori não informativos, não se precisa indicar o argumento *Prior* na especificação da função:

```
out=rhierLinearModel(Data=Data,Mcmc=Mcmc)
```

Com o primeiro modelo ajustado, agora precisa-se verificar a validade dos resultados gerados através da análise do *Burn-in*, *Thin* e Número de iterações. Através do comando abaixo gera-se a Figura 6 que traz os gráficos da distribuição dos dados e da sua autocorrelação:

```
plot(out$Deltadraw)
```

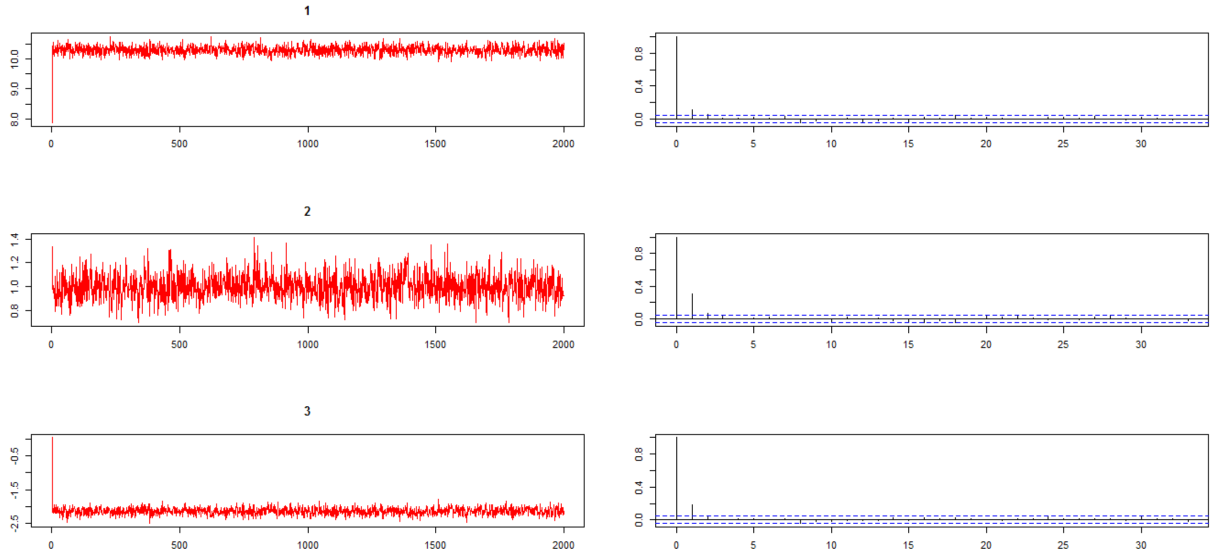


Figura 6: Análise do MCMC – modelo piloto

Burn-in:

Através da Figura 6, percebe-se que logo após a primeira iteração, os demais valores passaram a ser gerados em torno de uma mesma média e uma mesma variância. Logo, pode-se dizer que houve convergência das informações, e a partir deste valor elas estão aptas para interpretação, sendo este valor o *Burn-in*.

Esta análise visual pode ser realizada sem preocupação com a consequência desta escolha já que o pacote estabelece o *Burn-in* automaticamente.

Thin:

A Figura 6 apresenta também os correlogramas dos parâmetros β_0, β_1 e β_2, em uma cadeia gerada com keep=1.Através destes gráficos é possível uma correlação ainda alta nos primeiros lags.

Desta forma define-se um novo valor para o *thin*, temos então, keep=20.

Número de Iterações:

Após analisar *Burn-in* e *thin* pode-se ter algumas ideias sobre qual o número de iterações necessárias para a análise dos resultados. Considerando que se utilizará um valor a cada 20 gerados, serão realizadas 40.000 iterações.

5º Passo: ajustar o modelo final

Conforme salientado no passo anterior, em que se fez o ajuste dos parâmetros do argumento Mcmc do modelo, tem-se o seguinte modelo final:

```
Data=list(regdata=regdata,Z=Z)
R = 40000
Keep = 20
Mcmc=list(R=R,keep=Keep)
out=rhierLinearModel(Data=Data,Mcmc=Mcmc)
```

A Figura 7 então mostra, como resultado da implementação do MCMC, a cadeia de valores simulados para os β's do modelo, e conforme analisado anteriormente, percebe-se melhora na convergência dos valores e ausência na correlação dos dados. Desta forma, aptos para a interpretação.

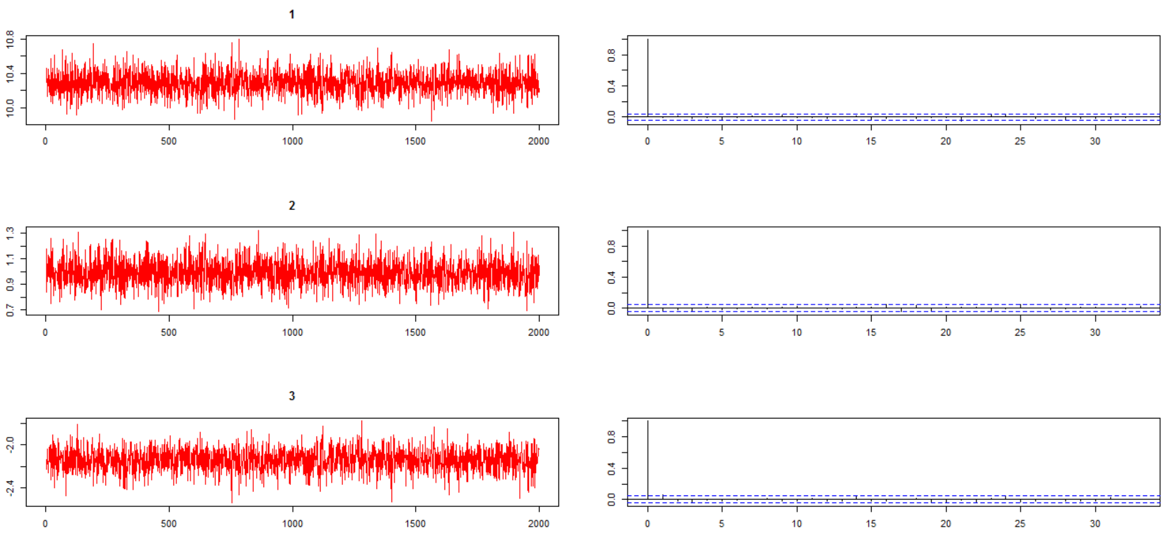


Figura 7: Análise do MCMC - modelo ajustado

Vale lembrar, que, quando utilizada a Inferência Bayesiana para a estimação dos parâmetros do modelo, se diz que cada um deles tem sua própria distribuição, e, desta forma, se pode calcular um intervalo de probabilidade para que aquele dado valor estimado.

Os valores simulados pelo MCMC são considerados uma amostra desta distribuição, pela Figura 8, pode-se observar a distribuição de cada um dos parâmetros estimados. Pode-se notar também que, em todos os casos a média, mediana e moda (representados em vermelho, laranja e amarelo), coincidem, podendo assim dizer que os valores pontuais são os que minimizam a perda quadrática e absoluta, propriedade para a média e mediana da posteriori, respectivamente, e são a estimativa de máxima verossimilhança generalizada, propriedade da moda da distribuição a posteriori.

```
1  plot(out$Deltadraw)
```

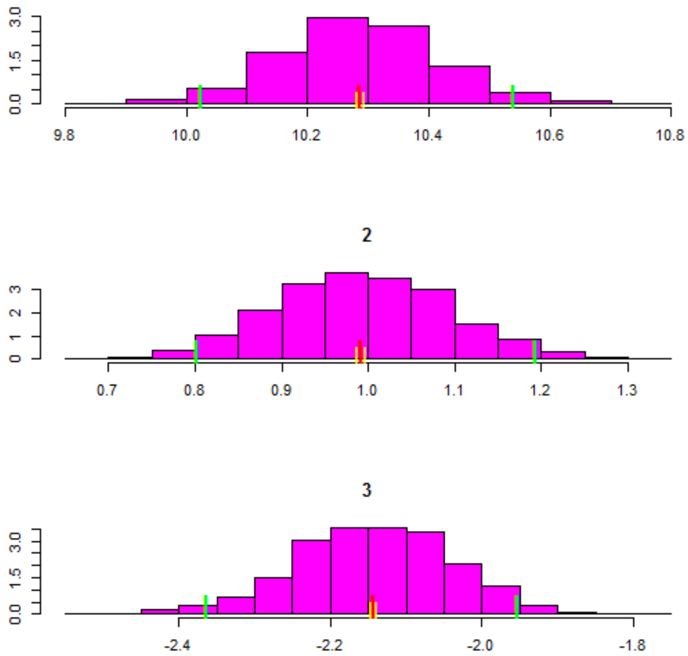


Figura 88: Distribuição a posteriori dos parâmetros β

6º Passo: interpretar os resultados

Com os resultados validados, pode-se agora realizar a interpretação dos mesmos. A Figura 8, apresenta, box-plots para alguns dos estabelecimentos, neste tipo de gráfico, a caixa representa a distância interquartílica, entre o 1º e 3º quartil, o traço central representa a mediana dos dados. Com isto, pode-se dizer que os valores gerados para os β's de cada estabelecimento são diferentes, explicando assim o uso de um segundo nível de informação.

Nota-se, por exemplo, que o estabelecimento 14 é o que apresenta maiores valores para β_0, já o estabelecimento 62, o que apresenta menor distância interquartílica. Em relação à segunda parte da Figura 9, onde se tem os resultados para β_1, observa-se uma distribuição completamente da anterior, por exemplo, o estabelecimento 15, que antes era um dos que possuía distribuição de β_0 com uma das maiores variabilidades, agora apresenta β_1 com uma das menores variabilidades. O mesmo pode ser observado para a última parte da Figura 9 em que as variabilidades novamente mudam.

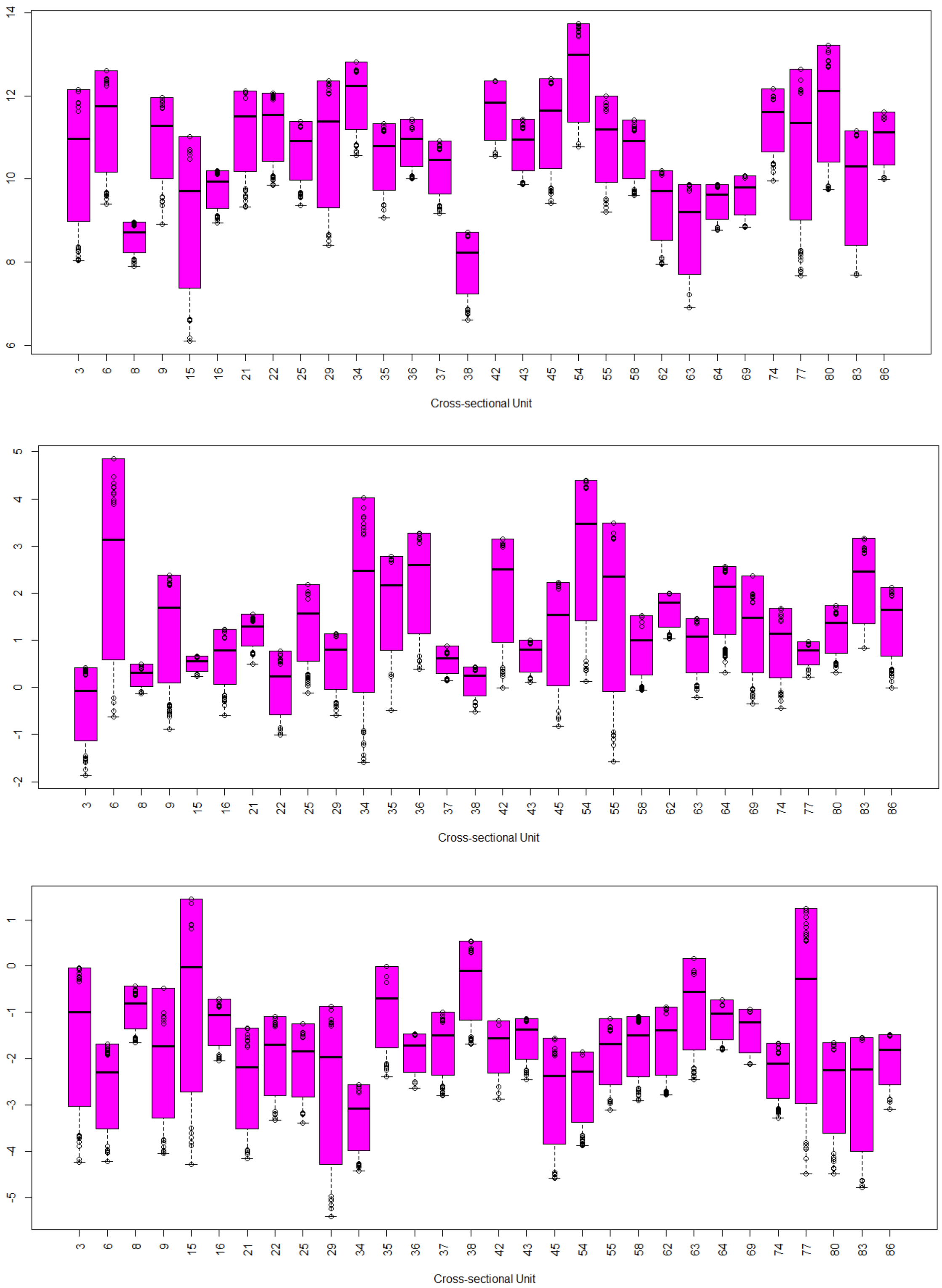


Figura 9: Box-plot dos β's ajustados para alguns dos estabelecimentos

Com os valores estimados apresentados na Figura 9 é possível estabelecer a reta de regressão estimada para cada estabelecimento. Com os comandos abaixo se obtêm as estatísticas descritivas de cada um β_{ik}, com *i* até 88 e *k* até 3. Para estabelecer a reta de regressão de cada um dos estabelecimentos então, é utilizado o valor mediano de cada β_{ik}.

```
median.beta0 = NULL
sd.beta0 = NULL
quantile5.beta0 = NULL
quantile95.beta0 = NULL
for (i in 1:88)
{
median.beta0 [i] = median (out$betadraw[i,1,])
sd.beta0 [i] = sd (out$betadraw[i,1,])
quantile2.5.beta0 [i] = quantile(out$betadraw[i,1,],0.025)
quantile97.5.beta0 [i] = quantile(out$betadraw[i,1,],0.975)
}
```

O comando acima traz as estatísticas descritivas de β_0 para cada um dos estabelecimentos, para a estimativa pontual será utilizado o valor mediano, que corresponde ao valor central dos dados, com desvio padrão é possível compreender a dispersão dos dados e com os quantis 2,5% e 97,5% estabelecer um intervalo de credibilidade de 95%.

```
median.beta1 = NULL
sd.beta1 = NULL
quantile5.beta1 = NULL
quantile95.beta1 = NULL
for (i in 1:88)
{
median.beta1 [i] = median (out$betadraw[i,2,])
sd.beta1 [i] = sd (out$betadraw[i,2,])
quantile2.5.beta1 [i] = quantile(out$betadraw[i,2,],0.025)
quantile97.5.beta1 [i] = quantile(out$betadraw[i,2,],0.975)
}
```

Novamente tem-se então o comando acima que traz as estatísticas descritivas, agora para β_1 de cada um dos estabelecimentos. O β_1 corresponde ao coeficiente angular para a covariável percentual de queijo exposto no estabelecimento. Através do comando abaixo, por fim, é possível obter as estatísticas descritivas para β_2 de cada estabelecimento, correspondendo assim ao coeficiente angular da covariável preço.

```
median.beta2 = NULL
sd.beta2 = NULL
quantile5.beta2 = NULL
quantile95.beta2 = NULL
for (i in 1:88)
{
median.beta2 [i] = median (out$betadraw[i,3,])
sd.beta2 [i] = sd (out$betadraw[i,3,])
quantile2.5.beta2 [i] = quantile(out$betadraw[i,3,],0.025)
quantile97.5.beta2 [i] = quantile(out$betadraw[i,3,],0.975)
}
```

Pode-se ainda estimar um modeloem que as estimativas de cada um dos estabelecimentos são combinadas em um modelo geral, neste caso dado pela reta obtida pelo comando abaixo:

$$\ln{(y_{ij})} = 10{,}29 + 0{,}99x_{1i} - 2{,}14\ln{(x_{2i})}.$$

```
summary(out$Betadraw)
```

Através do modelo geral, portanto, tem-se que, um aumento de 1% na quantidade de queijo exposta faz com que a venda de queijo seja mantida a mesma. Já o aumento de 1% unidade monetária no preço do queijo faz com exista um decréscimo de 2,14% na quantidade de queijo vendida.

A partir destes comandos, foram selecionados alguns dos estabelecimentos para análise gráfica, justamente com a análise do modelo geral.

$\ln{(y_{ij})} = 10{,}52 + 1.20x_{1i} - 3.69\ln{(x_{2i})}$, em laranja

$\ln(y_{ij}) = 11.96 + 1.47x_{1i} - 2.54\ln{(x_{2i})}$, em vermelho

$\ln(y_{ij}) = 9.29 + 1.73x_{1i} - 2.39\ln{(x_{2i})}$, em azul

$\ln(y_{ij}) = 9.96 + 0.62x_{1i} - 1.82\ln{(x_{2i})}$, em roxo

$\ln(y_{ij}) = \quad 9.44 + \ 0.16x_{1i} - 1.985\ln\ (x_{2i})$, em cinza

$\ln\ (y_{ij}) = \ 10{,}29 + \ 0{,}99x_{1i} \ - 2{,}14\ln\ (x_{2i})$, em preto

E então, fixando o valor do preço em 1, e variando o percentual de queijo exposto, de 1% a 100%, é possível observar as diferentes variações de *y*. Por ter um grande número de estabelecimentos, a Figura 9 mostra apenas os casos selecionados anteriormente. O ANEXO 1, traz os comandos para a obtenção da Figura 9, onde podem ser visualizadas as variações de inclinação das retas descritas abaixo:

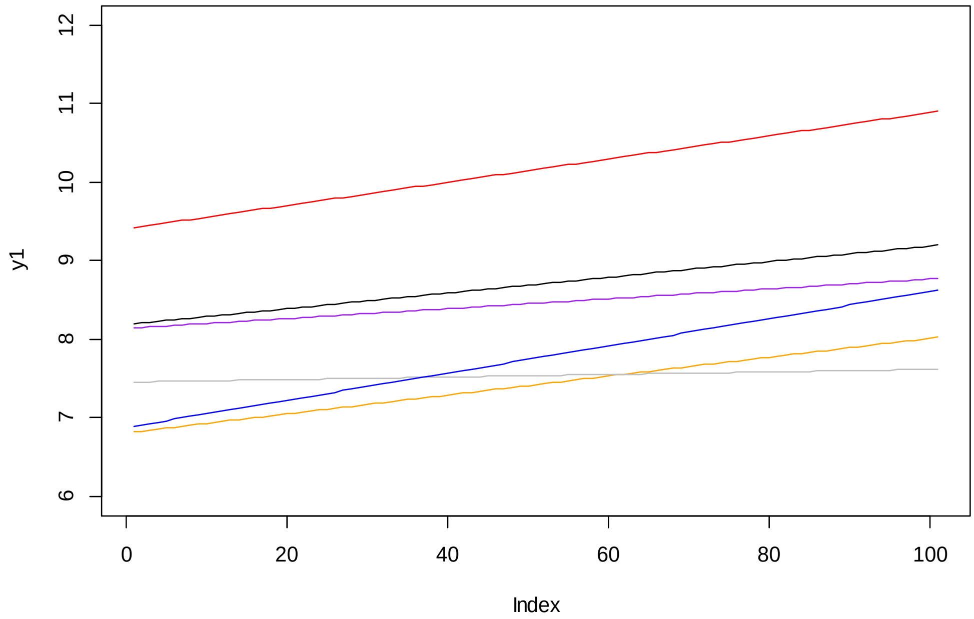


Figura 10: Reta de regressão de alguns dos estabelecimentos

Através da Fgura9, facilmente percebe-se que o aumento do percentual de queijo exposto para venda faz com que aumente a quantidade de queijo vendido, porém, este efeito no acréscimo nas vendas seja mais perceptível em alguns estabelecimentos.

6. Estudo de caso: aplicação dos dados de compra de diferentes marcas de margarina

Para estimar a probabilidade de compra de cada uma das marcas em diferentes situações de preço é proposto um modelo Logístico Multinomial Hierárquico Bayesiano, com o qual, pode-se verificar a influência do preço das margarinas na decisão de compra do consumidor.

Considerando que aspectos demográficos, como por exemplo,a renda familiar, podem influenciar na escolha de uma determinada margarina, é interessante elaborar um modelo em que as informações demográficas ajudem a explicar a escolha, deste modo, será empregado um segundo nível de informações em que será gerado um modelo para cada uma das famílias presentes no estudo.

Para melhor explicar o modelo Multinomial Hierárquico Logístico Bayesiano iremos, em um primeiro momento, considerar um caso simplificado, utilizando apenas duas das dez marcas de margarina.

6.1. Modelo Binomial Hierárquico Logístico Bayesiano – sem mistura

Neste caso, foram utilizadas as marcas 1 e 2, que foram respectivamente, a primeira e a segunda mais vendida entre as marcas analisadas. Neste caso o banco de dados é composto de 2.465 observações de compra, realizadas por 210 famílias que escolheram uma das duas marcas analisadas até o momento. Informações demográficas tais como o número de pessoas e renda da família serão também consideradas.

Preço da Marca 1: de maneira geral, pode-se dizer que o preço possui bom comportamento, mínimo de 0,19 e máximo de 0,67, com valores médio e mediano muito semelhantes, 0,5184 e 0,58 respectivamente, um desvio padrão de 0,15.

Preço da Marca 2: a marca 2, pode possuir alguns dados atípicos, dada diferença entre o terceiro quartil e o máximo, porém, ainda assim, possui média e mediana bem semelhantes, 0,5432 e 0,58 respectivamente, e um desvio padrão de 0,12.

Renda familiar: a renda, como habitual, possui grande amplitude (de 2,5 a 130), a distância entre média e mediana comprova a presença de dados outlier positivos (média de 27,56 e mediana de 22,5). Ainda deve-se considerar os valores do 1º e 3º quartil que são de 17,5 e 32,5, respectivamente.

Número de pessoas: o número de pessoas na família varia de 1 a 8, tendo como média 2,92 e mediana 3, a distância interquartílica de 2 sendo 2 o valor do 1º quartil e 4 o valor do 3º.

A fim de diminuir a variabilidade dos dados, facilitando assim a convergência das iterações da Cadeia de Markov será utilizada a transformação logarítmica nas informações de preço da margarina, renda familiar e tamanho da família.

O modelo proposto é definido em três níveis de hierarquia da seguinte forma:

1º nível: modela a variabilidade dentro das famílias usando a função de ligação Logito. A escolha da margarina e seus preços são considerados como as covariáveis para explicar esta variabilidade.

$$P(Y = m|X) = \frac{e^{g_m}}{\sum_{k=0}^{1} e^{g_k}}$$

onde, $g_m(x) = \ logito\ [P(Y = 1|x)] = \ln\ [\frac{P(y=1|x)}{P(y=0|x)}] = \ \beta_{1i}(x_1) + \ \beta_{2i}\ln\ (x_2)$

onde, x_{i1} é a variável dicotômica de qual das margarinas é preferida, sendo *m* = 0 para a primeira margarina e *m* = 1 para a segunda margarina e x_{i2} é o preço.

Foi considerada a marca da margarina 1 como nossa categoria de referência (0) e que por definição $g_0(x) = 0$, e por consequência, $g_1(x)$ será a função Logito para a marca 2.

2º nível: modela a variabilidade entre as famílias através de um modelo linear. A renda e o número de pessoas na família são considerados para explicar esta variabilidade.

$$\beta_{ij} = \ \delta_{1i}\ln\ (z_{1i}) + \ \delta_{2i}\ln\ (z_{2i})$$

ondez_{1i} é a renda da família *i* e z_{2i} é o tamanho da família i (i=1,...,n).

Note que o segundo nível da hierarquia possui covariáveis. Deste modo, as variações do βfaz com o efeito das covariáveis do primeiro nível também sejam alterados, deste modo, o modelo final estimado, que em geral é o do primeiro nível, agora sofrerá modificações, e seus coeficientes ficarão condicionados ao segundo nível do modelo.

3º nível: são apresentados os valores a priori para os parâmetros, deste modo, serão utilizadas prioris não informativas em que a média dos parâmetros será dada pela distribuição Normal e a variância pela Inversa Wishart. E para o parâmetro que indica a escolha da margarina, será utilizada priori conjugada Dirichlet. A seguir, será mostrado como se indica, na função, estas proris.

Implementação do modelo:

Para a implementação do modelo acima é utilizada a função *rhierMnlRwMixture*disponível no pacote *bayesm*do software R. Para utilizar a função é necessária a definição de três argumentos: Data, Prior, Mcmc.

1º Passo: definir o argumento Data

O argumento Data compreende a uma lista composta por três elementos: *p*, *lgtdata* e *Z* O elemento *p* representa o número de possíveis escolhas que, no exemplo é o valor 2 representando as duas marcas de margarina. O elemento *lgtdata*, são os dados das variáveis Y, X e Z. No exemplo, Y representa a marca de margarina escolhida, X e as informações de covariáveis, no caso, uma coluna de 0 ou 1, indicando qual das duas margarinas foi escolhida e a informação do logaritmo do preço de cada uma das marcas em cada uma das situações de venda; e *Z*, as informações das covariáveis demográficas, no caso as informações de renda e número de pessoas da família.

```
select= c(1,2)
chPr=as.matrix(margarine$choicePrice)
chPr=cbind(chPr[,1],chPr[,2],log(chPr[,2+select]))
demos=as.matrix(margarine$demos[,c(1,2,5)])

chPr=chPr[chPr[,2] <= 2,]

chPr[chPr[,2] == 7,2]=6
hhidl=levels(as.factor(chPr[,1]))
lgtdata=NULL
nlgt=length(hhidl)
p=length(select)
ind=1
```

```
for (i in 1:nlgt) {
nobs=sum(chPr[,1]==hhidl[i])
if(nobs >=5) {
data=chPr[chPr[,1]==hhidl[i],]
y=data[,2]
names(y)=NULL
X=createX(p=p,na=1,Xa=data[,3:4],nd=NULL,Xd=NULL,INT=TRUE,base=1)
lgtdata[[ind]]=list(y=y,X=X,hhid=hhidl[i]); ind=ind+1
}
}
nlgt=length(lgtdata)
Z=NULL
nlgt=length(lgtdata)
for(i in 1:nlgt){
Z=rbind(Z,demos[demos[,1]==lgtdata[[i]]$hhid,2:3])
}
Z=log(Z)
Z[,1]=Z[,1]-mean(Z[,1])
Z[,2]=Z[,2]-mean(Z[,2])

Data=list(p=p,lgtdata=lgtdata,Z=Z)
```

2º Passo: definir o argumento *Prior*:

O argumento Prior compreende a uma lista composta por oito elementos: *a*, *Deltabar*, *Ad*, mubar, *Amu*, *nu*, *V* e *ncomp* O segundo argumento, diz respeito das informações a priori; neste caso, serão utilizadas prioris não informativas, e então o único dado necessário será o de número de misturas de Normais utilizadas para a priori (*ncomp*), que neste caso, será definido como 1, já que não serão utilizadas prioris com mistura de Normais neste primeiro momento.

Os demais argumentos, que serão indicados pelo *default* da função são: *a*, vetor com os parâmetros da distribuição de Dirichlet a priori (default: rep(5,ncomp)); *Deltabar*, priori para média dos valores de Delta (default: 0); *Ad*, priori para vec(Delta) (default= 0,011); *mubar* priori para a média do erro (default: 0); *Amu* priori de precisão para a média do erro (default 0,011); *nu* priori para o número de componentes da Inversa Wishart para a matriz

Sigma (default: nvar + 3) e *V* priori para o parâmetro de locação da Inversa Wishart (default: nulo).

```
Prior=list(ncomp=1)
```

3º Passo: definir o argumento Mcmc:

O argumento Mcmc traz uma lista de quatro elementos: *s*, *w*, *R* e *keep*. Por fim, o argumento *Mcmc*, contará com as informações do parâmetro de escala para o algoritmo de Ruido Branco (*s*), que será mantido como default, s=2.93/sqrt(nvar); parâmetro de ponderação de probabilidades (*w*), novamente definido como default, 1; número de iterações do algoritmo, *R = 20.000*, e do intervalo entre as observações utilizadas para resultados, em um primeiro momento,*keep = 5*. Com isso avaliar-se-á um primeiro modelo, observando os valores gerados para δ_{11}, δ_{12}, δ_{21} e δ_{22}.

```
keep=5
R=20000
mcmc1=list(keep=keep,R=R)
```

4º Passo: ajustar modelo piloto:

Conforme as informações dadas nos passos de 1 a 3, define-se o primeiro modelo, deste modo, o primeiro modelo será especificado como:

```
out=rhierMnlRwMixture(Data=Data ,Prior=Prior, Mcmc=mcmc1)
```

Agora então, pode-se realizar a primeira avaliação do modelo, assim como no caso anterior, os resultados gerados apenas são válidos após análise das componentes: *Burn-in*, *Thin* e Número de iterações.

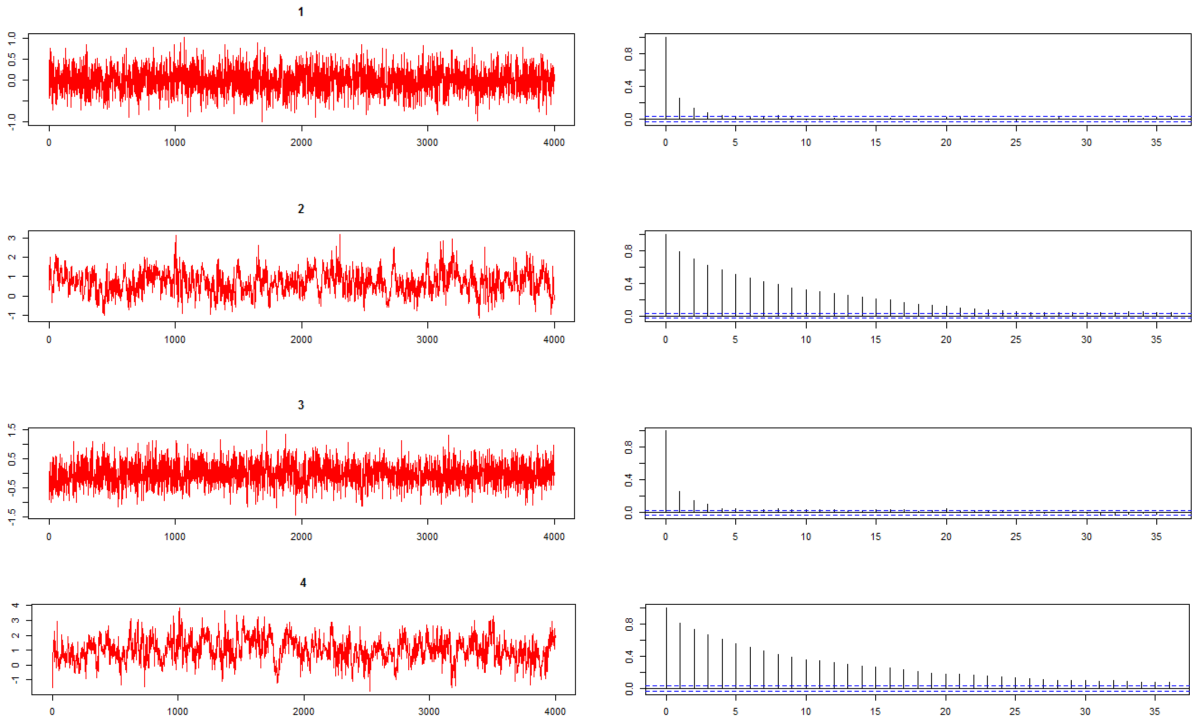


Figura 10: Análise do MCMC - modelo piloto

Burn-in:

Analisando a Figura 10, os gráficos da primeira coluna, pode-se avaliar o *Burn-in* necessário para a convergência dos dados, as estimativas 2 e 4, em especial, percebe-se que, mesmo "flutuando" em um mesmo valor central, ainda possui grande variabilidade.

Thin:

Através da Figura 10, percebe-se que para os argumentos referentes à renda familiar das equações a correlação persiste até o lag5, como é o caso dos argumentos 1 e 3; já para os outros dois argumentos, referentes ao tamanho da família, percebe-se que a correlação persiste até o lag 30.

Portanto, neste caso, a escolha inicial para o número de iterações e intervalo entre os dados a serem analisados não foi assertiva, então, será realizada uma nova tentativa em que as informações do argumento *Mcmc*, serão gerados mais valores, *R = 60.000*, e serão utilizadas informações mais espaçadas, no caso *keep = 30*.

Número de Iterações:

Como já analisado *Burn-in* e *thin*, assim como no estudo anterior, pode-se agora indicar o número mínimo de iterações necessárias para a análise, deste modo, considerando que serão descartados os primeiros 400 valores, e que se poderá apenas utilizar valores gerados com um intervalo de 30 iterações, precisar-se-á de, no mínimo, 60.000 iterações.

5º Passo: ajustar o modelo final:

Com os valores definidos anteriormente, pode-se agora definir na função os novos parâmetros:

```
Data=list(p=p,lgtdata=lgtdata,Z=Z)
Prior=list(ncomp=1)
keep=5
R=20000
mcmc1=list(keep=keep,R=R)
out=rhierMnlRwMixture(Data=Data ,Prior=Prior, Mcmc=mcmc1)
```

Conforme ilustrado anteriormente, na Figura 11, sabe-se que as informações geradas estão mais aptas para uso e interpretação então, pode-se responder aos objetivos do estudo; primeiro, pode-se verificar se realmente existe variação entre as famílias na importância do preço na hora da compra.

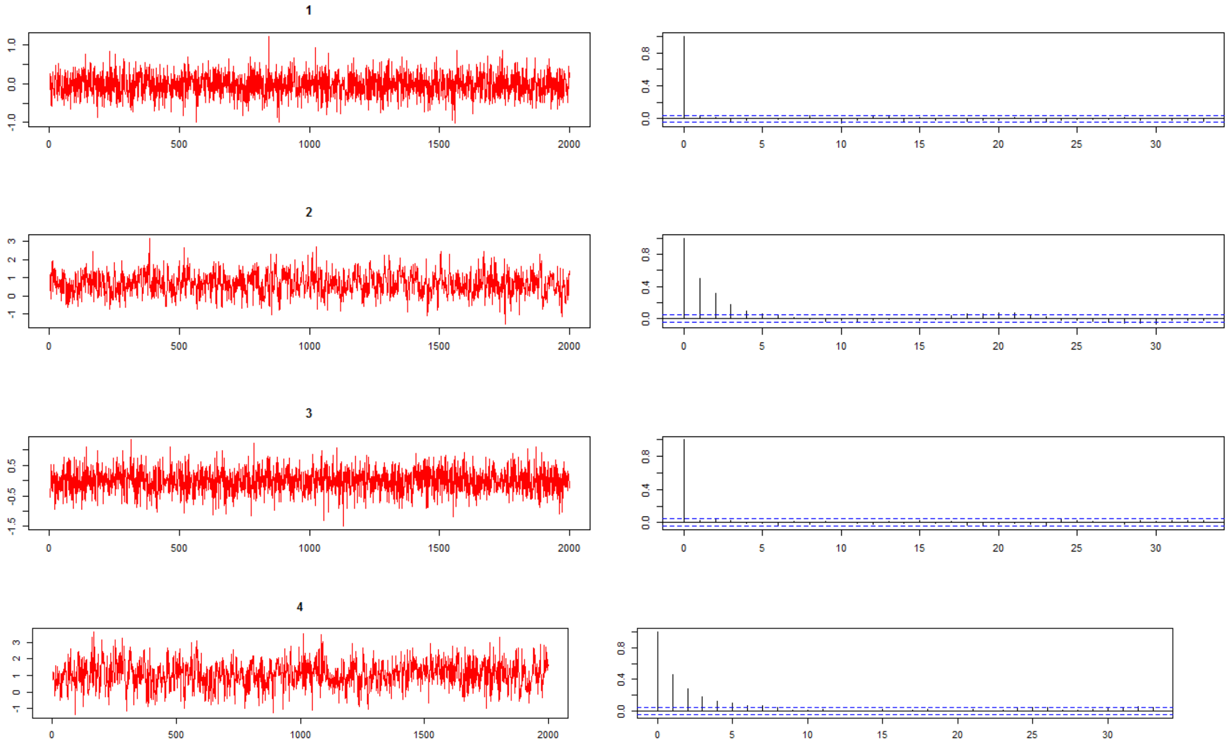


Figura 111: Análise do MCMC - modelo ajustado

Quando é utilizada a Inferência Bayesiana para a estimação dos parâmetros do modelo, diz-se que cada um deles tem sua própria distribuição, e, desta forma, pode-se calcular um intervalo de probabilidade para que aquele dado valor estimado.

Pela Figura abaixo 12, pode-se observar a distribuição de cada um dos parâmetros estimados através do MCMC, em todos os casos a média, mediana e moda (representados em vermelho, laranja e amarelo), coincidem, podendo assim dizer que os valores pontuais são os que minimizam a perda quadrática e absoluta, propriedade para a média e mediana da posteriori, respectivamente, e são a estimativa de máxima verossimilhança generalizada, propriedade da moda da distribuição a posteriori.

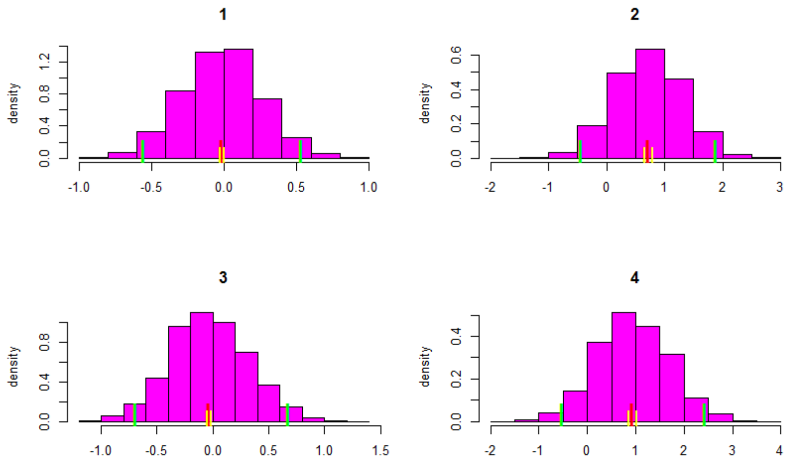


Figura 12: Histograma da distribuição a posteriori de cada um dos quatro parâmetros

Na Figura 13 então, através dos box-plot de algumas das famílias analisadas, pode-se perceber que existe variação entre os coeficientes gerados para cada uma delas, e então percebe-se a importância de estabelecer um segundo nível no problema.

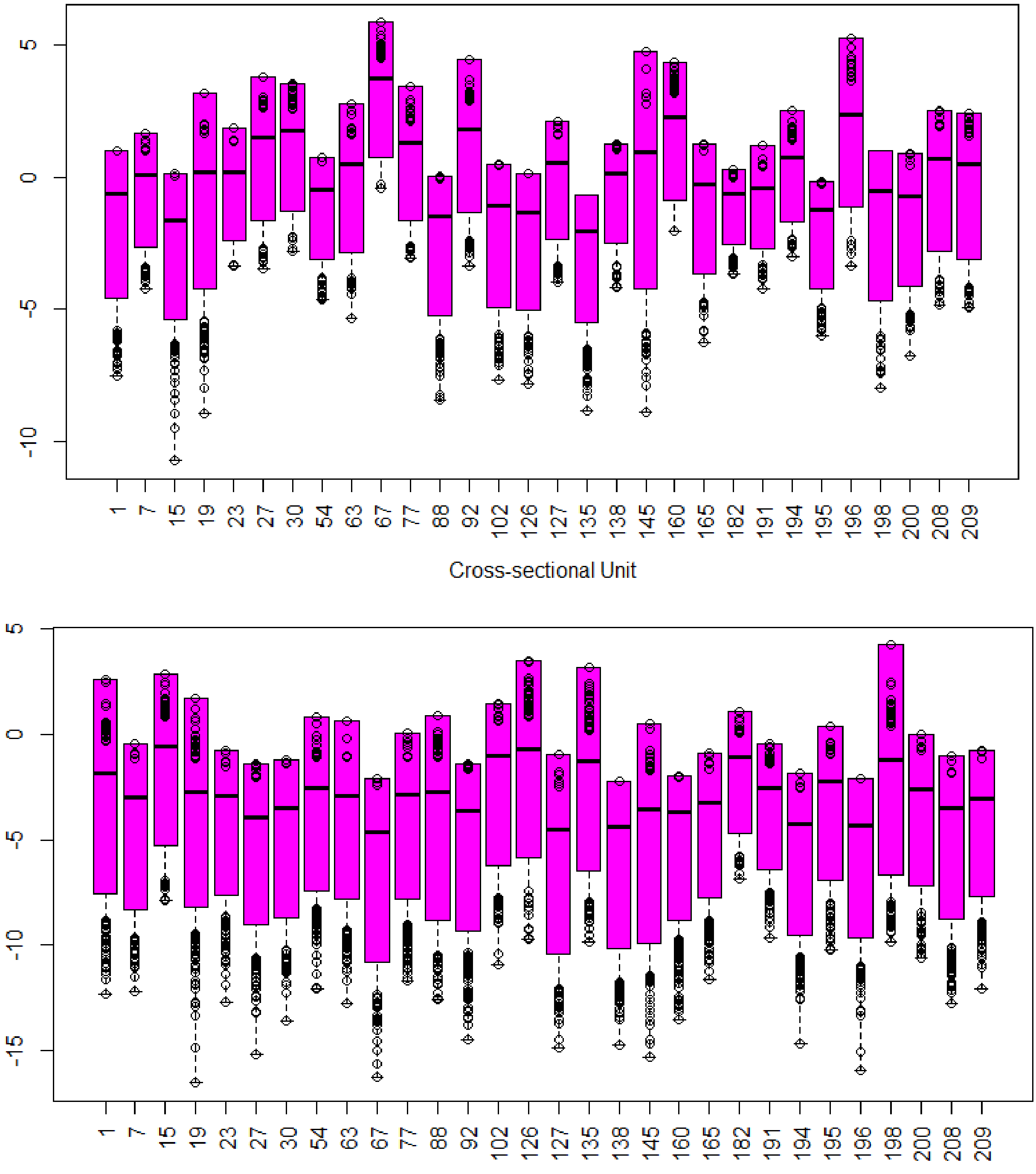


Figura 1123: Box-plot dos β's ajustados para algumas das famílias

Com os valores estimados acima, pode-se estabelecer um diferente modelo para cada uma das famílias estudadas.

```
median.beta1 = NULL
sd.beta1 = NULL
quantile5.beta1 = NULL
quantile95.beta1 = NULL
for (i in 1:88)
{
median.beta1 [i] = median (out$betadraw[i,2,])
sd.beta1 [i] = sd (out$betadraw[i,2,])
quantile2.5.beta1 [i] = quantile(out$betadraw[i,2,],0.025)
quantile97.5.beta1 [i] = quantile(out$betadraw[i,2,],0.975)
}
```

O quadro acima traz as estatísticas descritivas, agora para β_1 de cada uma das famílias estudadas. O β_1 corresponde ao coeficiente angular para a covariável renda familiar. Através do comando abaixo, por fim, é possível obter as estatísticas descritivas para β_2 de cada estabelecimento, correspondendo assim ao coeficiente angular da covariável tamanho da família.

```
median.beta2 = NULL
sd.beta2 = NULL
quantile5.beta2 = NULL
quantile95.beta2 = NULL
for (i in 1:88)
{
median.beta2 [i] = median (out$betadraw[i,3,])
sd.beta2 [i] = sd (out$betadraw[i,3,])
quantile2.5.beta2 [i] = quantile(out$betadraw[i,3,],0.025)
quantile97.5.beta2 [i] = quantile(out$betadraw[i,3,],0.975)
}
```

Atendendo ao objetivo geral do estudo, entender o efeito do preço na probabilidade de compra de determinada marca de margarina, tem-se, portanto, o seguinte resultado:

$$\beta_1 = -0.010 \ln(z_1) + 0.719 \ln(z_2)$$
$$\beta_2 = -0.041 \ln(z_1) + 0.927 \ln(z_2)$$

e então,

$$g_1(x) = (-0.01 \ln(z_{11}) + 0.72 \ln(z_{12}))(x_1) + (-0.04 \ln(z_{21}) + 0.93 \ln(z)) \ln(x_2)$$

De forma geral, tem-se que, um aumento de 1% na renda familiar provoca uma redução de 0,0001 na estimativa de β_1, dado que o número de pessoas na família é constante; tem-se também que um aumento de 1% no número de pessoas na família acarreta em um aumento de 0,00719 na estimativa de β_1 mantendo a renda familiar constante. A mesma interpretação é válida para a segunda equação, referente a estimativa de β_2. Deste modo, pode-se dizer que quando se tem um público de alta renda, β_1 torna-se negativo, indicando uma não preferência pela marca em questão.

Conforme explicado anteriormente, $g_1(x)$ não possui interpretação prática direta, porém, a partir de sua estimativa é possível estimar a probabilidade de compra de cada uma das marcas de margarina. Esta estimativa é feita utilizando a expressão abaixo.

$$P(Y = m|X) = \frac{e^{g_m}}{\sum_{k=0}^{1} e^{g_k}}$$

Para uma situação em que uma família composta de 3 pessoas, possui renda familiar de 32,5 unidades monetárias, tem preferência pela marca de margarina 2, considerando o preço das margarinas em 0,4 e 0,8 unidades monetárias, obtém-se os seguintes valores de probabilidade de compra:

Situação 1: $P(Y = 1|X)$ = 0,38
$P(Y = 2|X)$ = 0,62

Situação 2: $P(Y = 1|X)$ = 0,47
$P(Y = 2|X)$ = 0,53

Considerando agora, um tamanho de família constante, de três pessoas, e três diferentes valores de renda: 17,5, correspondendo ao 1º quartil, 22,5, indicando a mediana e 32,5, o 3º quartil. Analisar-se-á a influência do Preço na compra das Marcas.

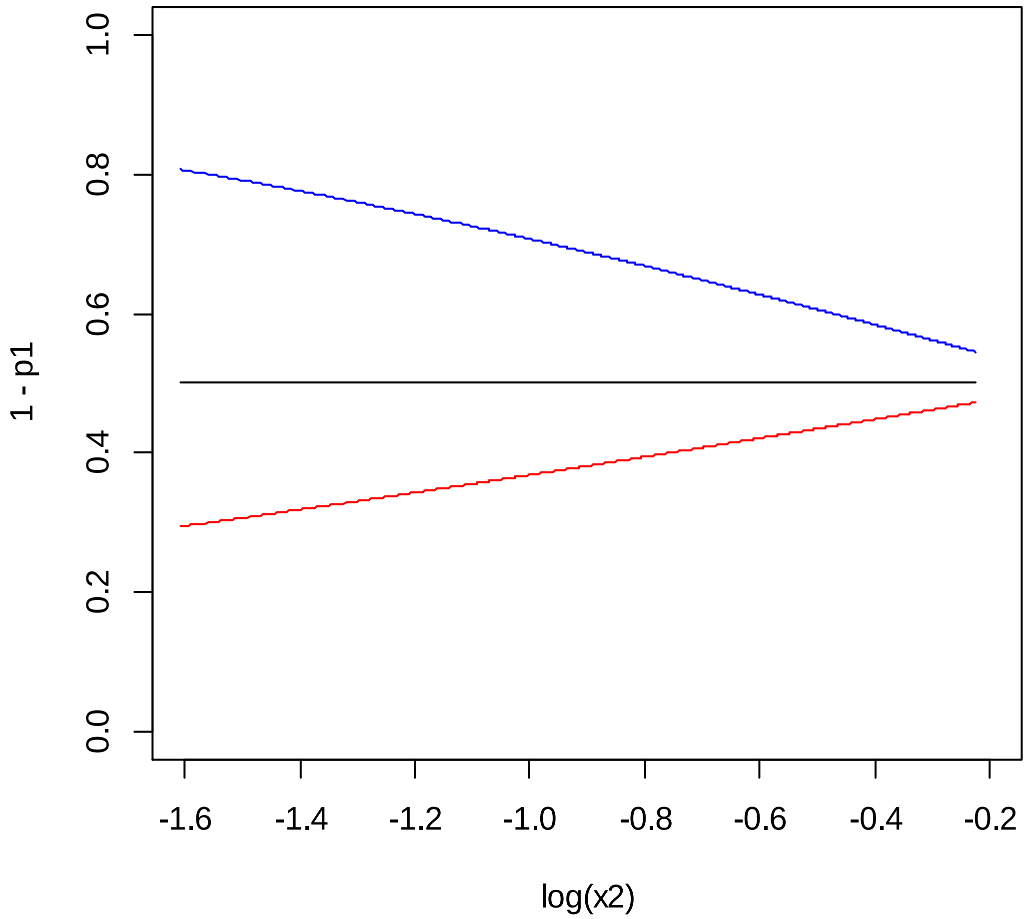


Figura 134: Variações de probabilidade de compra da Marca 1

Na Figura 14 tem-se então, que comparada a preferência da marca 2 (x_1), as diferentes variações de probabilidade para diferentes perfis de renda, sendo azul, baixa renda, preto, renda mediana e vermelho alta renda. Em um primeiro momento, observa-se que, para o público de alta renda, conforme o preço da margarina aumenta, aumenta também a probabilidade de compra da marca 1; já para o público de baixa renda, o efeito é o contrário, conforme o preço da margarina aumenta, diminui a sua probabilidade de compra da marca 1. Enquanto para o cliente de renda mediana a probabilidade de compra permanece constante.

6.2. Modelo Binomial Hierárquico Logístico Bayesiano – com mistura

A fim de obter melhores resultados para o modelo estimado, será proposto um modelo análogo ao anterior, tendo uma principal diferença: prioris de Mistura de Normais, elas continuarão sendo não informativas, porém agora, poderão ser captados valores mais distantes da média para as análises.

Vem se tornando usual, em modelos de marketing Prioris de distribuição para o primeiro nível do modelo hierárquico bayesiano, a Normal oferece a flexibilidade ideal e se ajusta de forma convencional para grandes modelos de regressão ou de análise multivariada bayesianos com grande número de observações.

Embora modelos de Priori Normal sejam “flexíveis” eles tendem a “puxar” dados mais distantes para o centro das observações, o que acaba sendo uma desvantagem para modelos que procuram exatamente descobrir estruturas nos dados. Em dados de marketing em geral, é natural ter-se modelos multimodais, neste caso, modelos com mistura podem auxiliar a incluir esta variação.

O número de componentes desta priori, entretanto, é definido de maneira intuitiva, (Rossi, 2005), em geral, cada mistura representa um segmento e, portanto, um diferente caso de estimação.

Em casos de modelos bayesianos hierárquicos e aplicações ao Marketing, modelos de mistura podem ser aplicados a vetores de parâmetros com grande dimensão e, consequentemente, prioris poderão facilmente exceder uma mistura de mais 200 ou 300 componentes (Rossi, 2005).

Deste modo, compreendendo que será aumentada a complexidade do problema, foi utilizado um maior número de iterações, já que foi aumentado também o *Thin*, intervalo entre dados utilizados.

Ajuste do modelo:

Conforme estabelecido então, modificar-se-á o número de iterações, o *thin* e, por objetivo principal, indicar-se-á o uso de mistura de duas Normais para as distribuições a priori do modelo. Deste modo, o modelo será definido na função da seguinte forma:

```
Data=list(p=p,lgtdata=lgtdata,Z=Z)
Prior=list(ncomp=2)
keep=50
R=100000
mcmc1=list(keep=keep,R=R)
out=rhierMnlRwMixture(Data=Data ,Prior=Prior, Mcmc=mcmc1)
```

O algoritmo utilizado pelo pacote, ao utilizar prioris de mistura não informativas, separa os dados em grupos – indicando o número de misturas – e passa a interpolá-los de um grupo para o outro até o momento em que nenhum dado troca mais de grupo, tento assim, duas diferentes distribuições e nelas os dados que se aproximam melhor de cada uma delas.

Utilizou-se a Figura 15 para verificar a validade dos resultados obtidos pelo algoritmo.

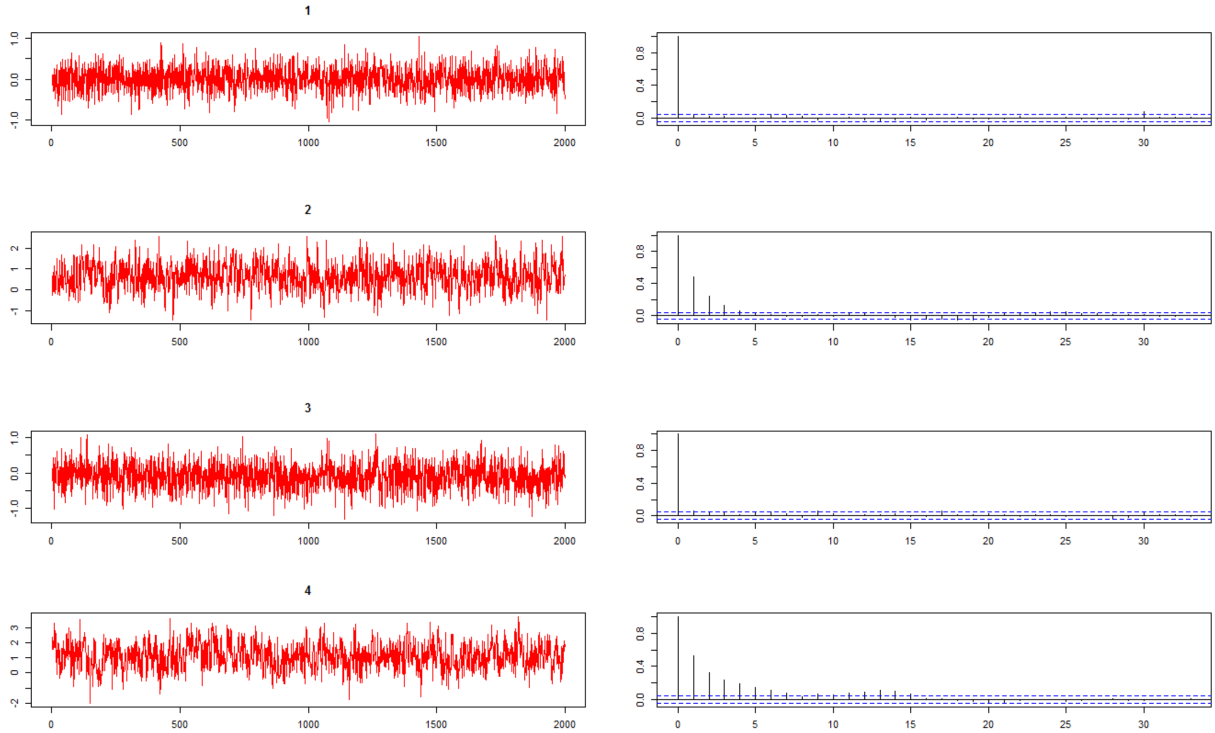


Figura 14: Análise do MCMC - modelo ajustado

Considerando a complexidade do modelo utilizado, embora ainda exista uma correlação de Lag5 no segundo e quarto parâmetro, estes resultados podem ser considerados válidos.

Seguindo nas interpretações então, a Figura 16, pode ser interpretada do mesmo modo que a Figura 13; o que deve ser destacado neste ponto diz respeito ao efeito da mistura nos valores gerados para β de cada família; que comprova que a Priori de Mistura consegue utilizar até valores mais extremos para realização do método, já que como se pode observar, tem-se uma amplitude maior para os valores de β gerados.

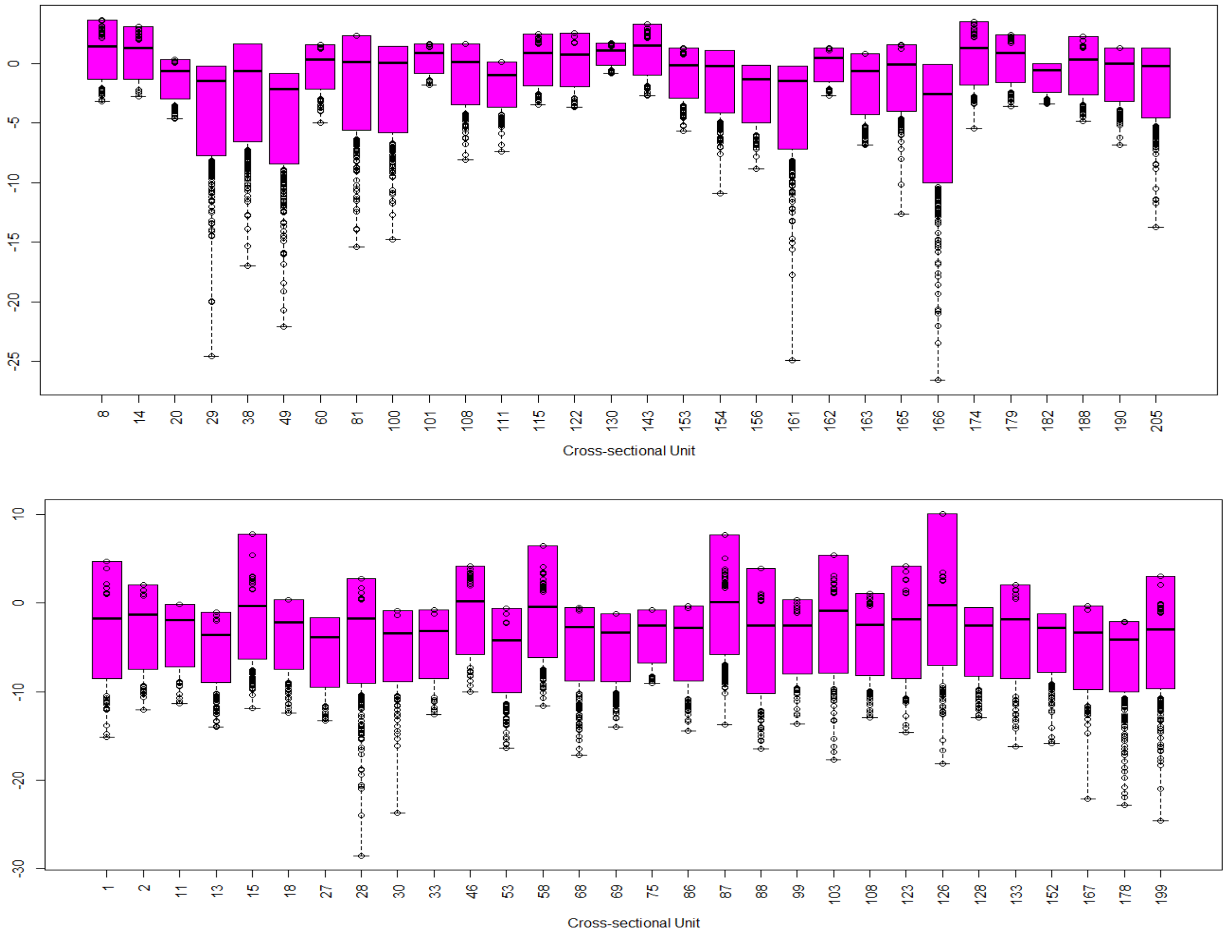


Figura 16: Box-plot dos β's ajustados para algumas das famílias

Isso faz com que se consiga captar ainda melhor as variações entre as famílias, e então, o segundo nível de informações, que já era útil pelo fato de explicar β em função da renda e do número de pessoas, agora passa a ser ainda mais importante, pois consegue absorver ainda mais as diferenças entre famílias.

Considerando então, estas variações, o modelo final será dado da seguinte forma:

$$\beta_1 = 0.002 \ln(z_1) + 0.65 \ln(z_2)$$

$$\beta_2 = -0.095 \ln(z_1) + 1.12 \ln(z_2)$$

e então,

$$g_1(x) = (0.002 \ln(z_1) + 0.65 \ln(z_2))(x_1) + (-0.095 \ln(z_1) + 1.12 \ln(z_2)) \ln(x_2)$$

Será analisada agora a variação da influência do preço na probabilidade de compra das margarinas para famílias de mesmo número de pessoas, 3, e três diferentes rendas: 17,5, correspondendo ao 1º quartil, 22,5, indicando a mediana e 32,5, o 3º quartil. O ANEXO 3 traz os comandos utilizados para a elaboração da Figura 17.

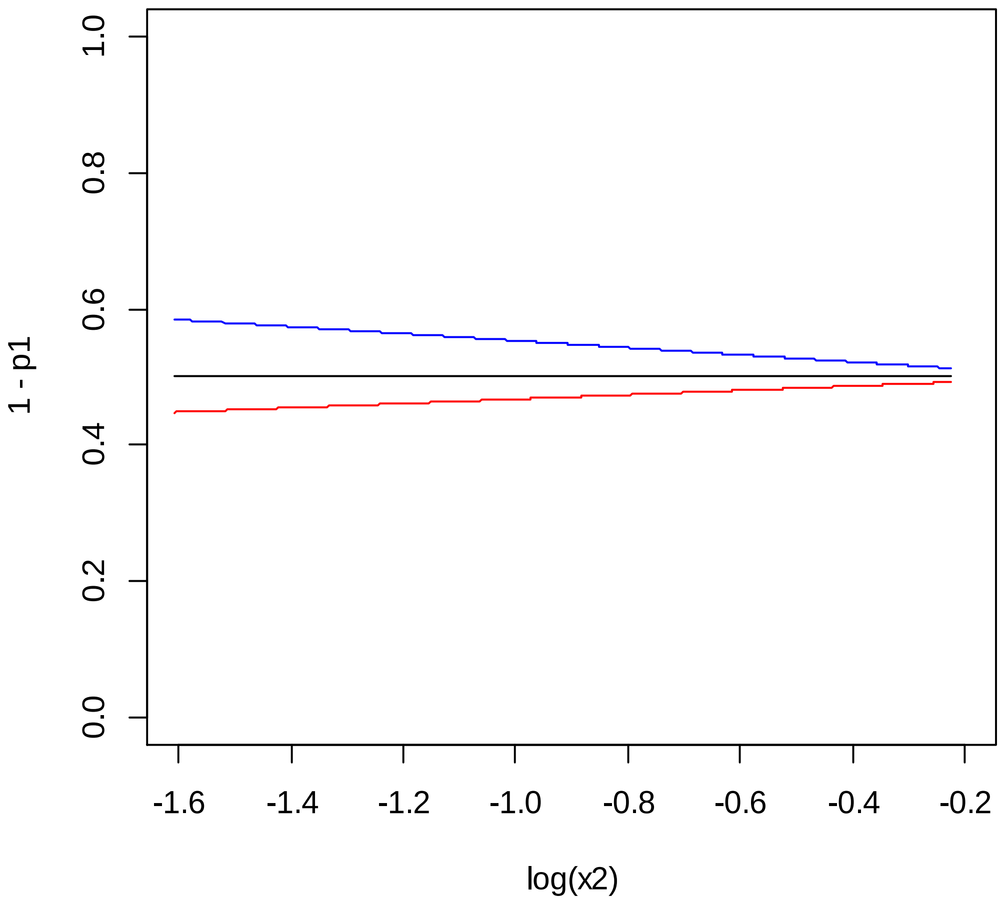


Figura 17: Variações de probabilidade de compra da Marca 1

Pela Figura 17 temos então, uma interpretação análoga a da Figura 14, em que a única diferença é o efeito da renda nas probabilidades, que agora é bem mais sutil, já que o coeficiente de inclinação passa a ser menor para ambas as retas em questão, o que já era imaginado devido os coeficientes δ_{11} e δ_{21} serem bem próximos de zero.

6.3. Modelo Multinomial Hierárquico Logístico Bayesiano – sem mistura

Compreendendo os casos anteriores, considerando apenas duas marcas de margarina, torna-se mais fácil de entender o caso específico, onde o objetivo principal é comparar seis das dez marcas observadas.

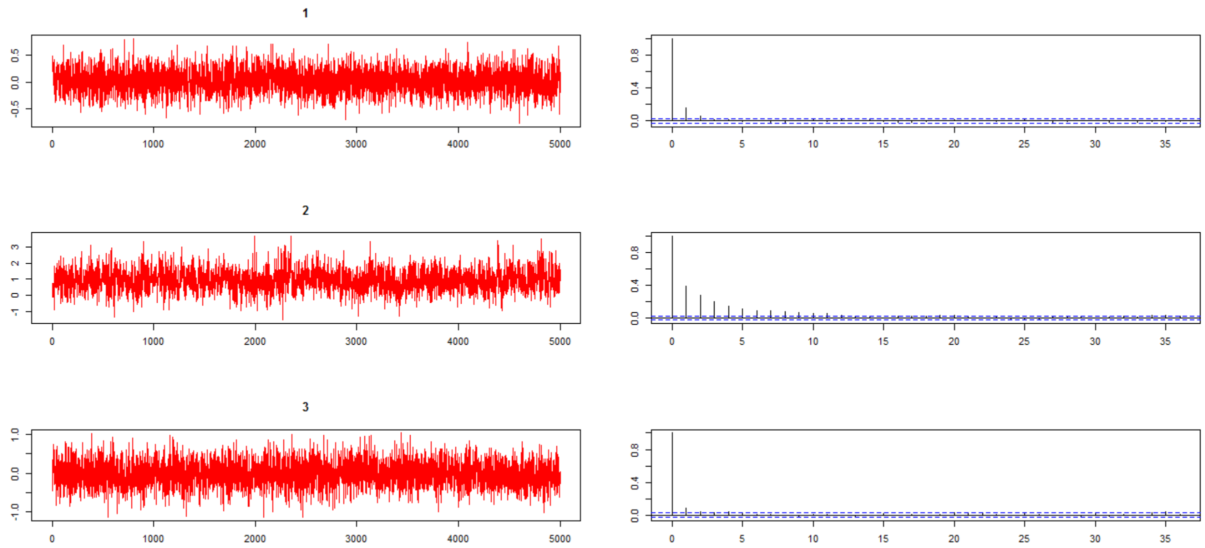
Deste modo, utilizou-se as informações demográficas das famílias e comparou-se 6 das marcas apresentadas, sendo elas: 1, 2, 3, 4, 5 e 7 e tendo então o caso politômico de um modelo hierárquico logístico. Deste modo, foram consideradas as observações de compra de 314 famílias; fixando a marca 1 como nossa categoria de referência e estimou-se o modelo do mesmo modo que o caso dicotômico.

Implementação do modelo:

Assim como no exemplo anterior, utilizou-se a função *rhierMnlRwMixture*para a obtenção dos resultados, e os argumentos foram fixados de forma muito semelhante ao anterior, apenas atentando ao fato de compararmos seis marcas (*p* = *6*). Continuou-se contando com prioris não informativas.

```
Data=list(p=p,lgtdata=lgtdata,Z=Z)
Prior=list(ncomp=2)
keep=30
R=150000
mcmc1=list(keep=keep,R=R)
out=rhierMnlRwMixture(Data=Data ,Prior=Prior, Mcmc=mcmc1)
```

Observando a Figura 18 percebe-se que foram atendidos os critérios de convergência do modelo, com 150.000 repetições e intervalo entre dados de 30.



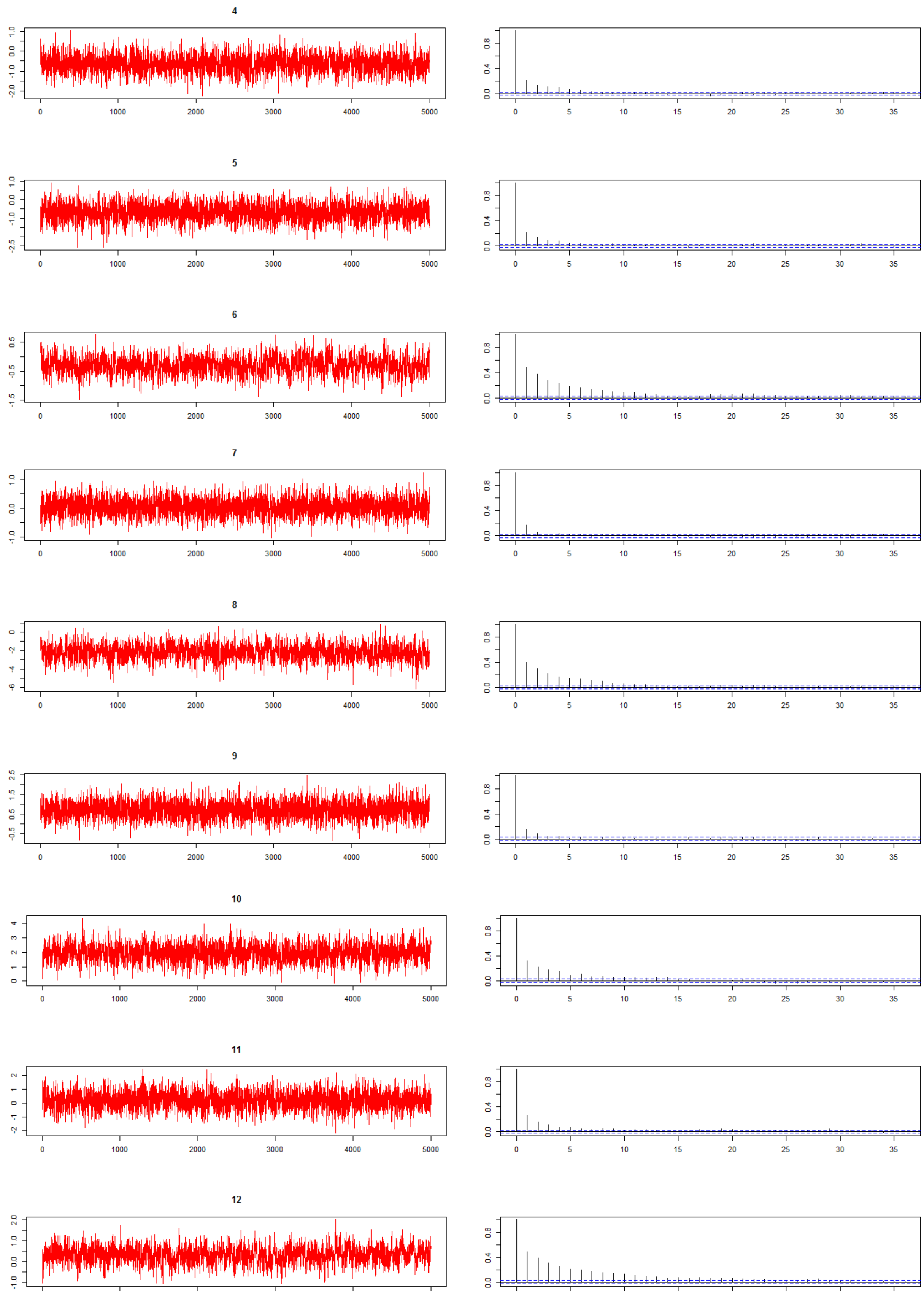
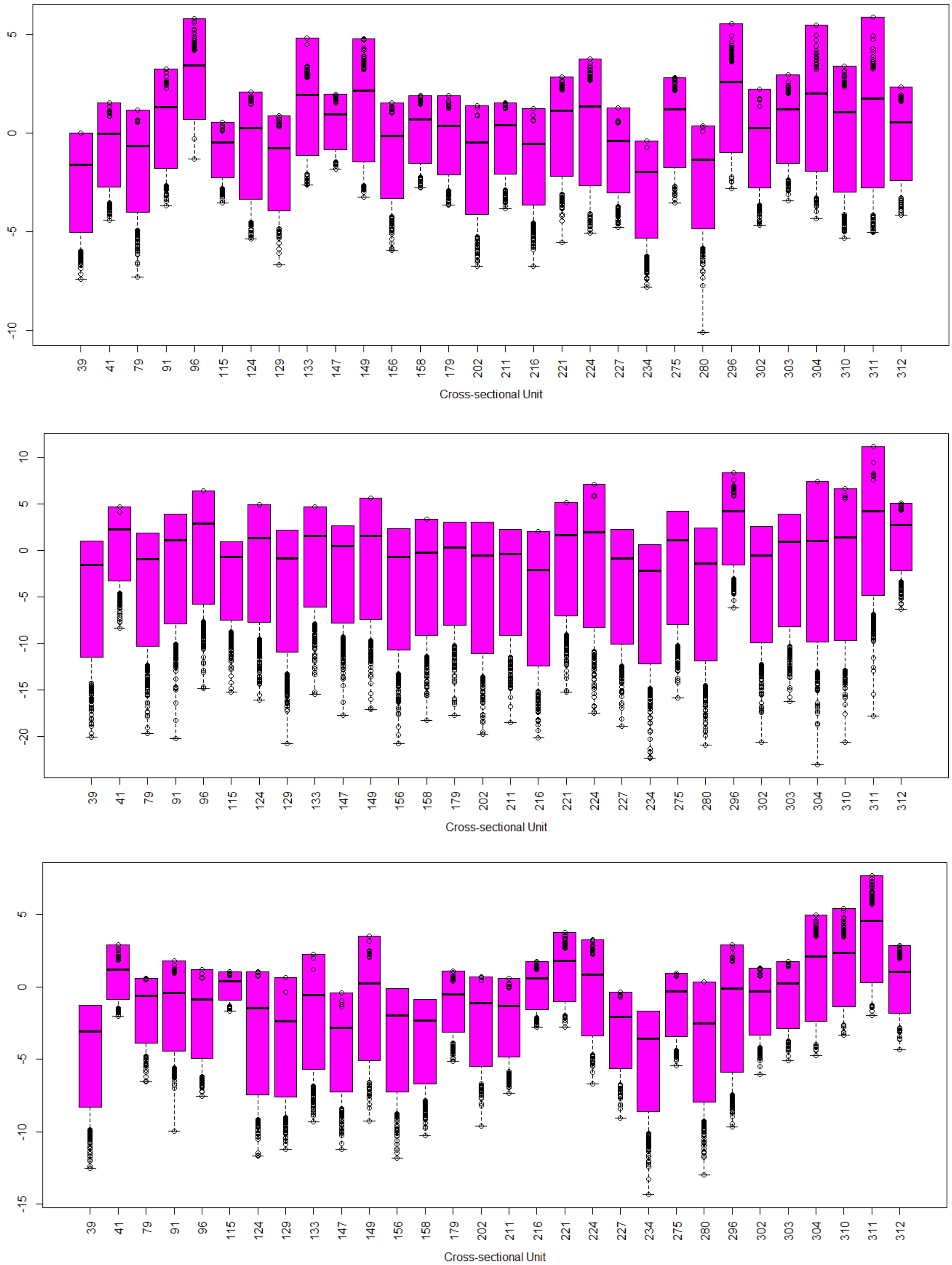


Figura 18: Análise do MCMC - modelo ajustado

É possível demonstrar a importância de considerar as diferenças entre os perfis demográficos das famílias no momento de escolha. Deste modo, a partir da Figura 19, apresenta as estimativas dos β_i para cada uma das famílias:



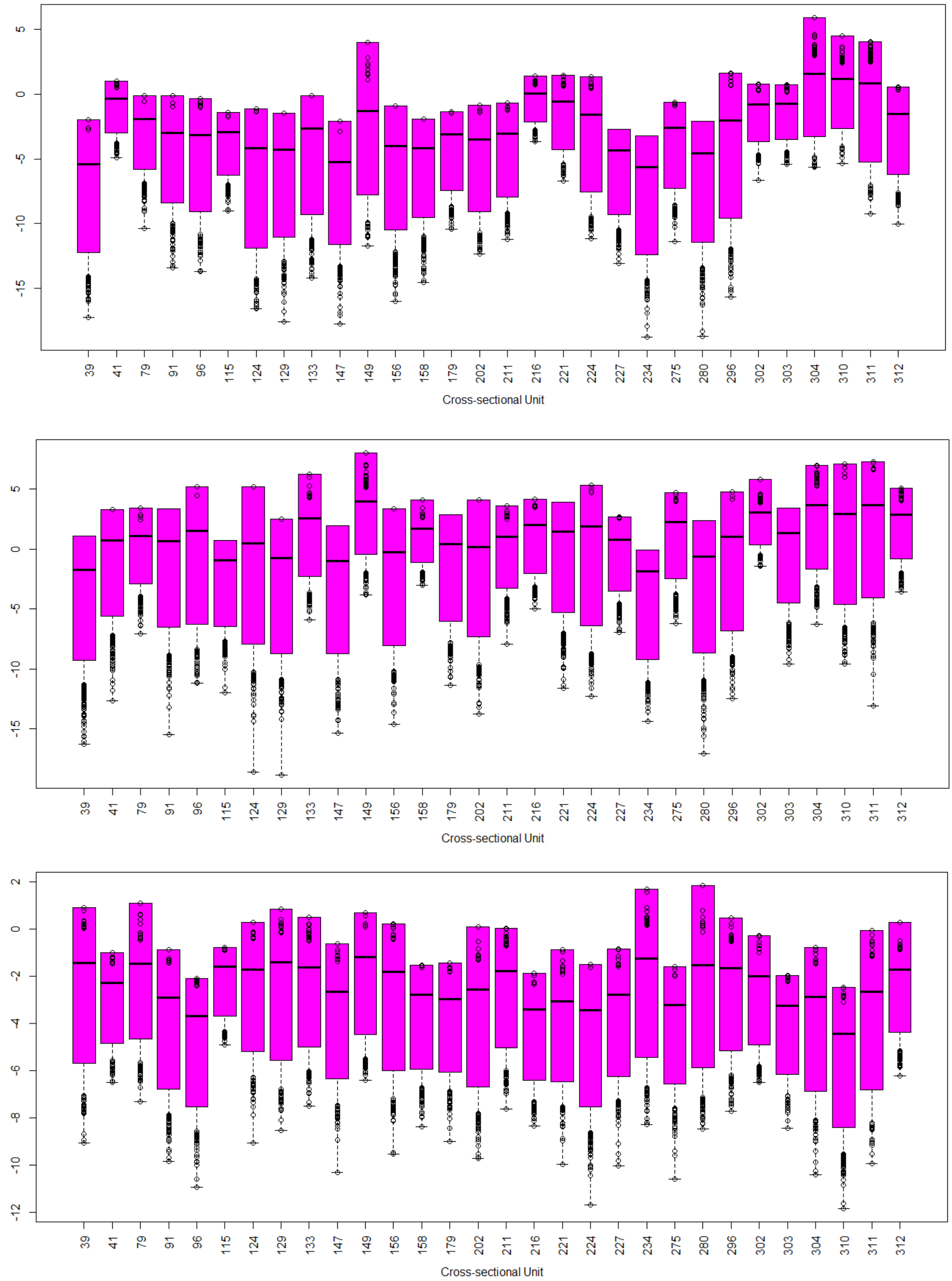


Figura 159: Box-plot dos β's ajustados para algumas das famílias

Seguindo nas interpretações, que são todas análogas ao caso dicotômico, o modelo final pode ser escrito da seguinte forma:

Agora, por ter um maior número de comparações, torna-se mais interessante realizar as predições para a probabilidade de compra de cada uma das marcas.

Deste modo, tem-se:

$$g_1(x) = (0.016\ln(z_{11}) + 0.92\ln(z_{12}))(x_1) + (0.202\ln(z_{21}) + 0.328\ln(z))\ln(x_2)$$
$$g_2(x) = (-0.02\ln(z_{11}) - 0.642\ln(z_{12}))(x_1) + (0.202\ln(z_{21}) + 0.328\ln(z))\ln(x_2)$$
$$g_3(x) = (-0.682\ln(z_{11}) - 0.296\ln(z_{12}))(x_1) + (0.202\ln(z_{21}) + 0.328\ln(z))\ln(x_2)$$
$$g_4(x) = (0.040\ln(z_{11}) - 2.180\ln(z_{12}))(x_1) + (0.202\ln(z_{21}) + 0.328\ln(z))\ln(x_2)$$
$$g_5(x) = (0.722\ln(z_{11}) + 1.961\ln(z_{12}))(x_1) + (0.202\ln(z_{21}) + 0.328\ln(z))\ln(x_2)$$

E então, a probabilidade de compra de cada marca, pode ser calculada a partir de:

$$P(Y = i|X) = \frac{e^{g_i}}{\sum_{k=0}^{1} e^{g_k}}$$

Para fins de exemplificar como é realizado este cálculo, foi fixado um valor para a Renda, um para o Número de pessoas na família e combinados com diferentes variações de preço, verificando assim as mudanças na probabilidade de compra de cada uma das 6 marcas:

Tabela 1: Tabela de probabilidade de compra de 6 diferentes de margarina

Preço	p0	p1	p2	p3	p4	p5
0.2	0.08321061	0.1541590	0.1457799	0.05491420	0.1600480	0.4018882
0.3	0.06026632	0.1580297	0.1494130	0.05617173	0.1640876	0.4120317
0.4	0.04483886	0.1606325	0.1518553	0.05701432	0.1668046	0.4188543
0.5	0.03335894	0.1625696	0.1536725	0.05763979	0.1688269	0.4239323
0.6	0.02429311	0.1640993	0.1551074	0.05813280	0.1704242	0.4279432
0.7	0.01684686	0.1653559	0.1562858	0.05853714	0.1717363	0.4312380
0.8	0.01055751	0.1664172	0.1572811	0.05887823	0.1728447	0.4340212

Cabe destacar brevemente, que, com o aumento do preço das margarinas, percebe-se uma redução na probabilidade de compra da marca 1, e um aumento na probabilidade da compra das demais. Os comandos para o cálculo da probabilidade são encontrados no ANEXO 4.

Assim como no caso dicotômico, o modelo poderia ter sido ajustado utilizando prioris mistura de Normais. Ele seria implementado de forma análoga a mostrada na Seção 6.2. porém, destacando que, por um aumento na complexidade do problema, seria necessário um maior *Thin*, *Bur in* e Número de iterações.

7. Conclusão

Dado o objetivo principal do trabalho, trazer os benefícios da modelagem Estatística Bayesiana para estudos de marketing, foram utilizadas duas principais abordagens, uso de modelo linear hierárquico e de modelo multinomial hierárquico.

O trabalho apresentou de forma objetiva quais são e como são apresentados os problemas de marketing e como a estatística pode ser importante para solucioná-los, além disso, foram apresentados os principais conceitos estatísticos bayesianos necessários para a compreensão das técnicas utilizadas na implementação dos modelos exemplificados.

Os conceitos de Inferência Bayesiana, com maior enfoque ao conceito e as informações a priori, sua importância, como utilizá-las e defini-las.Foram também abordados os conceitos de MCMC, método iterativo para simular valores amostrados da distribuição a posteriori dos parâmetros de interesse. Foi apresentado o pacote *bayesm* do software livre R, que através de diferentes funções soluciona problemas diretamente relacionados ao marketing e a microeconomia.

Por fim, para ilustrar e demonstrar a importância dos conceitos apresentados anteriormente, foram apresentados dois estudos de caso, o primeiro, referente ao volume de vendas de queijo, ajusta um modelo linear hierárquico bayesiano capaz de explicar o volume de vendas a partir do preço do produto e do percentual exposto para o consumidor. Ao utilizar um modelo hierárquico, são consideradas as diferenças entre os estabelecimentos estudados, tornando o modelo mais eficaz.

O segundo estudo de caso, onde o objetivo era estudar a probabilidade de compra de diferentes marcas de margarina foi dividido em três partes para facilitar o entendimento, primeiro foram consideradas apenas duas marcas de margarina e ajustado um modelo binomial hierárquico, depois considerado o uso de prioris de mistura neste caso binomial, e por último, o caso multinomial, considerando 6 marcas de margarina. Em todos os casos foi utilizado um modelo hierárquico, neste estudo, o segundo nível de informações conta com covariáveis, deste modo, além de ajustar um modelo diferente para cada família, pode-se também estudar o efeito das variáveis do segundo nível na informação do primeiro nível e consequentemente na resposta final, a probabilidade.

Como limitações das aplicações pode-se destacar a escolha de covariáveis, que não foi possibilitada neste trabalho. Sabe-se que com o auxílio de covariáveis é possível compreender ainda melhor as particularidades do segundo nível, no caso, consumidor e que métodos para a escolha destas covariáveis devem ser considerados.

Outro ponto a ser destacado é a escolha de um número adequado de misturas, segundo Rossi (2005) sabe-se que este número pode variar fortemente de acordo com o problema encontrado, no trabalho foi apenas ilustrado o uso de prioris de mistura e o foco não foi em encontrar um número ótimo de misturas, o que poderia fazer com que os resultados obtidos fossem ainda mais relevantes.

Por fim, sabe-se que ao utilizar MCMC é necessário tomar alguns cuidados para se ter resultados aptos para a interpretação, um número elevado de iterações associado a um *Thin* e *Burn in* adequados garantem a validade dos resultados obtidos, nos casos mais complexos, como os apresentados nas Seções 6.2 e 6.3torna-se difícil a convergência das informações, e deste modo, as interpretações realizadas nestas Seções podem estar levemente viesadas.

Referências Bibliográficas

ANDO, T. **Bayesian Model Selection and Statistical Modeling**1ª ed. Chapman & Hall, 2010.

BERGER, P. D.; NASR. N. Customer Lifetime Value: Marketingmodels and applications. **Journal Interactive Marketing**,v. 12, n. 1,p. 17–30, 1998.

CONGDON, P. **Bayesian Statistical Modeling**.John Wiley & Sons, Ltd, 2001.

GELMAN, A.; CARLIN, J. B.; STERN, H. S.; RUBIN, D. B. **Bayesian Data Analysis**. 4ª ed. Nova Iorque: Chapman & Hall, 1998.

GILKS, W.R.; RICHARDSON, S.; SPIEGELHALTER, D.J. **Markovchain Monte Carlo in practice**1ª ed. Chapman & Hall, 1996.

GUPTA, S.; LEHMANN, D. R. Customer Lifetime Value and Firm Valuation.**Journal of Relationship Marketing**, v. 5, n. 2/3, p. 87-110, 2006.

HOSMER D.W.; LEMESHOW S. **Aplied Logistic Regression**.2ª ed. John Wiley & Sons, Ltd, 2000.

JAIN, D.; SINGH, S. S. Customer Lifetime Value Research in Marketing: a review andfuture directions. **Journal of Interactive Marketing**, v. 16, n. 2, p. 34-46, 2002.

KUMAR, V.; SHAH, D. Expanding the Role of Marketing: From Customer Equity to Market Capitalization.**Journal of Marketing**, v. 73, n. 6, p. 119-136, 2009.

ROSSI P.E.; ALLENBY G.M.; MCCULLOCH R. **BayesianStatisticsand Marketing**. John Wiley& Sons, Ltd, 2005

ROSSI, P. (2012). bayesm: Bayesian Inference for Marketing/Micro-econometrics. R package version 2.2-5. http://CRAN.R-project.org/package=bayesm

SCHMITTLEIN, D. C.; MORRISON, D. G.; COLOMBO, R. Counting your customers: Who are they and what will they do next? **Management Science**, v. 33, p. 1-24, 1987.

SCHMITTLEIN, D. C., PETERSON, R A. Customer Base Analysis: An IndustrialPurchase Process Application. **Marketing Science**, v. 13, n. 1, p. 41-68, 1994.

VENKATESAN, R.; KUMAR, V.A Customer LifetimeValue Framework for Customer Selection and Optimal Resource Allocation Strategy.**Journal of Marketing**,v. 68, p. 106-125, 2004.

Anexos

ANEXO 1: Comandos para obtenção da Figura 9

```
x1 = seq(0,1,0.01)
n = length(x1)
y1 = NULL
for (i in 1:n){
y1[i] = mean.beta0[1] + (mean.beta1[1]*x1[i]) +(mean.beta2[1] * 1)}

x23 = seq(0,1,0.01)
n = length(x23)
y23 = NULL
for (i in 1:n){
y23[i] = mean.beta0[23] + (mean.beta1[23]*x1[i]) +(mean.beta2[23] * 1)}

x39 = seq(0,1,0.01)
n = length(x39)
y39 = NULL
for (i in 1:n){
y39[i] = mean.beta0[39] + (mean.beta1[39]*x1[i]) +(mean.beta2[39] * 1)}

x44 = seq(0,1,0.01)
n = length(x44)
y44 = NULL
for (i in 1:n){
y44[i] = mean.beta0[44] + (mean.beta1[44]*x1[i]) +(mean.beta2[44] * 1)}

x82 = seq(0,1,0.01)
n = length(x82)
y82 = NULL
for (i in 1:n){
y82[i] = mean.beta0[82] + (mean.beta1[82]*x1[i]) +(mean.beta2[82] * 1)}
```

```
xgeral = seq(0,1,0.01)
n = length(xgeral)
ygeral = NULL
for (i in 1:n){
ygeral[i] = 10.3 + xgeral[i] + (-2.1 * 1)}

plot(y1,type = 'l',col = "orange", ylim=c(6,12))
lines(y23,type = 'l',col = "purple")
lines(y39,type = 'l',col = "grey")
lines(y44,type = 'l',col = "red")
lines(y82,type = 'l',col = "blue")
lines(ygeral,type = 'l', col = "black")
```

ANEXO 2: Comandos para obtenção da Figura 14

```
F = -9.571e-17

R1 = -2.205e+00
R2 = -1.184e-16
R3 = 1.350e+00

x1 = 1
x2 = seq (0.2, 0.8, 0.001)

R1 = R1 - R2
R3 = R3 - R2
R2 = R2 - R2

beta11=-0.01*R1 + 0.719*0
beta12=-0.41*R1 + 0.927*0
beta21=-0.01*R2 + 0.719*0
beta22=-0.41*R2 + 0.927*0
beta31=-0.01*R3 + 0.719*0
beta32=-0.41*R3 + 0.927*0

g1=beta11*(x1)+beta12*log(x2)
g2=beta21*(x1)+beta22*log(x2)
g3=beta31*(x1)+beta32*log(x2)

p1=exp(g1)/(1+exp(g1))
p2=exp(g2)/(1+exp(g2))
p3=exp(g3)/(1+exp(g3))


plot(log(x2),1-p1,xlim = c(-1.6,-0.2), ylim = c(0,1),type="l",col="blue")
lines (log(x2),1-p2,type="l",col="black")
lines (log(x2),1-p3,type="l",col="red")
```

ANEXO 3: Comandos para obtenção da Figura 16

```
F = -9.571e-17

R1 = -2.205e+00
R2 = -1.184e-16
R3 = 1.350e+00

x1 = 1
x2 = seq (0.2, 0.8, 0.001)

R1 = R1 - R2
R3 = R3 - R2
R2 = R2 - R2

beta11=0.002*R1 + 0.65*0
beta12=-0.095*R1 + 1.12*0
beta21=0.002*R2 + 0.65*0
beta22=-0.095*R2 + 1.12*0
beta31=0.002*R3 + 0.65*0
beta32=-0.095*R3 + 1.12*0

g1=beta11*(x1)+beta12*log(x2)
g2=beta21*(x1)+beta22*log(x2)
g3=beta31*(x1)+beta32*log(x2)

p1=exp(g1)/(1+exp(g1))
p2=exp(g2)/(1+exp(g2))
p3=exp(g3)/(1+exp(g3))

plot(log(x2),1-p1,xlim = c(-1.6,-0.2), ylim = c(0,1),type="l",col="blue")
lines (log(x2),1-p2,type="l",col="black")
lines (log(x2),1-p3,type="l",col="red")
```

ANEXO 4: Comandos para obtenção da Tabela 1.

```
R = 1.35
F = 0
preco = seq (0.2,0.8,0.1)
g0 = 0
g1 = (0.016*R + 0.92*F)*1 + (0.202*R + 0.328*F)*log(preco)
g2 = (-0.02*R - 0.642*F)*1 + (0.202*R + 0.328*F)*log(preco)
g3 = (-0.682*R - 0.296*F)*1 + (0.202*R + 0.328*F)*log(preco)
g4 = (0.04*R - 2.18*F)*1 + (0.202*R + 0.328*F)*log(preco)
g5 = (0.722*R + 1.961*F)*1 + (0.202*R + 0.328*F)*log(preco)

p1 = exp(g1)/(exp(g0)+exp(g2)+exp(g3)+exp(g4)+exp(g5))
p2 = exp(g2)/(exp(g0)+exp(g1)+exp(g3)+exp(g4)+exp(g5))
p3 = exp(g3)/(exp(g0)+exp(g1)+exp(g2)+exp(g4)+exp(g5))
p4 = exp(g4)/(exp(g0)+exp(g1)+exp(g2)+exp(g3)+exp(g5))
p5 = exp(g5)/(exp(g0)+exp(g1)+exp(g2)+exp(g3)+exp(g4))
p0 = 1-(p1+p2+p3+p4+p5)

cbind(preco,p0,p1,p2,p3,p4,p5)
```