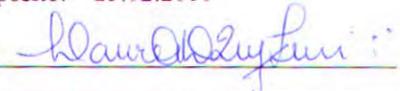


SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 25.02.2000

Assinatura: 

Análise Bayesiana de Modelos para Dados Binários Correlacionados

Vanderly Janeiro

Orientador: Prof. Dr. Jorge Alberto Achcar

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, para o exame de qualificação, como parte dos requisitos para obtenção do título de Mestre em Ciências – Área: Ciências de Computação e Matemática Computacional.

**USP – São Carlos
Fevereiro de 2000**

Agradecimentos

Sobretudo agradeço a Deus por me conceder a vida e estar aqui realizando mais este sonho.

Ao Prof. Dr. Jorge Alberto Achcar, por me confiar este trabalho, e pela orientação no decorrer de seu andamento, e especialmente por ter compartilhado comigo a sua grande experiência.

Aos professores que me auxiliaram e incentivaram para que eu pudesse chegar a mais esta etapa da minha vida.

Aos professores desta instituição: Profa. Dra. Maria Creusa B. Salles, Prof. Dr. Josemar Rodrigues, Prof. Dr. Marinho Gomes de Andrade Filho pela dedicação e empenho ao ensino.

Ao Prof. Dr. José Galvão Leite e ao Prof. Dr. Marinho pelas sugestões dadas em meu exame de qualificação, que vieram a contribuir muito com esta dissertação e com meu crescimento.

À todos os meus amigos que direta ou indiretamente contribuíram para que este trabalho fosse realizado, e em especial a Lucimary Afonso dos Santos que nos meus momentos de desespero manteve-se sempre do meu lado.

Com muito amor agradeço aos meus pais e meus irmãos por acreditarem em mim e sempre estarem ao meu lado.

Aos funcionários do ICMC pela dedicação às pessoas deste departamento.

E finalmente, à CAPES pelo apoio financeiro.

Abstract

In this dissertation, we develop a Bayesian analysis of regression models for correlated binary data in the presence of covariates, including the case with replicates. We consider probit and logistic regression models for correlated binary data assuming random effects with a mixture of normal distributions, since this model have great flexibility to the fitted for correlated binary data. We also present some considerations for the case with replicates. We assume informative prior distributions for the parameters of the model and we use Gibbs sampling and Metropolis-Hastings algorithms to get Monte Carlo estimates for the posterior quantities of interest. We also present some considerations for the selection of models using discrepancy measures between the fitted model and the data (Pearson residuals) and using the predictive densities (Bayes factor) estimated by MCMC (Markov Chain Monte Carlo). We present a numerical example to illustrate the proposed methodology.

Resumo

Nesta dissertação, desenvolvemos uma análise Bayesiana de modelos de regressão para dados binários correlacionados com covariáveis, podendo ocorrer réplicas. Assumimos os modelos de regressão logístico e probito para dados binários correlacionados considerando efeitos aleatórios com uma mistura de distribuições normais, pois este modelo tem uma grande flexibilidade para ser ajustado aos dados binários correlacionados em muitas aplicações. Também fazemos algumas considerações aos casos onde podem ocorrer repetições das observações ou réplicas. Assumimos distribuições a priori informativas para os parâmetros do modelo e consideramos os algoritmos Gibbs sampling e Metropolis-Hastings, para obter as estimativas de Monte Carlo para as quantidades a posteriori de interesse. Apresentamos também algumas considerações na seleção de modelos utilizando uma medida da discrepância entre o modelo ajustado e os dados (resíduo de Pearson) e utilizando as densidades preditivas (fator de Bayes) estimadas por MCMC (Monte Carlo em Cadeias de Markov). Apresentamos um exemplo numérico para ilustrar os métodos propostos.

Sumário

1	INTRODUÇÃO	1
1.1	Motivações, Modelos e Objetivos	3
1.1.1	Alguns Modelos Considerados na Literatura para Análise de Dados Binários Correlacionados	5
1.2	Organização da Dissertação	9
2	ANÁLISE DO MODELO DE PRENTICE	11
2.1	Função de Verossimilhança	12
2.2	Análise Clássica do Modelo	13
2.3	Análise Bayesiana do Modelo	16
2.4	Análise Bayesiana Hierárquica Assumindo Efeitos Aleatórios Normais . . .	19
2.5	Modelo de Efeitos Aleatórios com uma Mistura de Distribuições Normais .	21
3	ANÁLISE DO MODELO PROBITO	27
3.1	Análise Clássica do Modelo	28
3.2	Análise Bayesiana do Modelo	30
3.3	Análise Bayesiana Assumindo uma Distribuição Normal para os efeitos aleatórios α_i	32
3.4	Análise Bayesiana Assumindo uma Mistura de Normais para α_i	34
4	DADOS BINÁRIOS CORRELACIONADOS COM RÉPLICAS	39
4.1	Estimadores de Máxima Verossimilhança	40
4.2	Análise Bayesiana Assumindo $\alpha_i = \alpha$	41
4.3	Análise Bayesiana Assumindo $\alpha_i \sim N(0, \sigma_\alpha^2)$	42
4.4	Análise Bayesiana Assumindo uma Mistura de Normais para α_i	44

5	SELEÇÃO DE MODELOS	47
5.1	Resíduos de Pearson	47
5.2	Um Procedimento Bayesiano para Verificação da adequação de Modelos - Fator de Bayes	48
5.2.1	Aproximação de Monte Carlo - (MC)	49
5.2.2	Aproximações de Monte Carlo em Cadeias de Markov - (MCMC)	49
5.2.3	A Escolha do Modelo	51
6	APLICAÇÕES NUMÉRICAS	53
6.1	Ilustração Numérica para o Capítulo 2	53
6.1.1	EMV para os Parâmetros do Modelo com Ligação Logística	55
6.1.2	Aplicação do Modelo 1	56
6.1.3	Aplicação do Modelo 2	61
6.1.4	Aplicação do Modelo 3	63
6.1.5	Comparação dos Modelos 1, 2 e 3	66
6.2	Ilustração Numérica para o Capítulo 3	69
6.2.1	EMV para os Parâmetros do Modelo com Ligação Probita	70
6.2.2	Aplicação do Modelo 4	70
6.2.3	Aplicação do Modelo 5	73
6.2.4	Aplicação do Modelo 6	76
6.2.5	Comparação dos Modelos 4, 5 e 6	79
6.3	Ilustração Numérica para o Capítulo 4	81
6.3.1	EMV para os Parâmetros do Modelo com Ligação Logística Con- siderando Réplicas	84
6.3.2	Aplicação do Modelo 7	84
6.3.3	Aplicação do Modelo 8	86
6.3.4	Aplicação do Modelo 9	88
6.3.5	Comparação dos Modelos 7, 8 e 9	90
6.3.6	EMV para os Parâmetros do Modelo com Ligação Probita Con- siderando Réplicas	92
6.3.7	Aplicação do Modelo 10	92
6.3.8	Aplicação do Modelo 11	94

6.3.9	Aplicação do Modelo 12	97
6.3.10	Comparação dos Modelos 10, 11 e 12	99
7	CONCLUSÕES E CONSIDERAÇÕES FUTURAS	101
	Apêndices	102
A	ALGORITMO GIBBS SAMPLING	103
B	ALGORITMO DE METROPOLIS-HASTINGS	105
C	O CRITÉRIO DE CONVERGÊNCIA DE GELMAN E RUBIN (1992)	107
D	PROGRAMA COMPUTACIONAL PARA O MODELO 3	109
	Referências Bibliográficas	116

Capítulo 1

INTRODUÇÃO

Supor que associado a cada indivíduo ou unidade experimental temos uma observação onde a resposta assume um, de dois possíveis valores (sucesso ou fracasso). Isto pode ocorrer em muitas áreas de interesse. Por exemplo:

- um componente eletrônico é defeituoso, ou não é defeituoso;
- um experimento agrônomico tem por objetivo verificar se uma determinada semente germina ou não em diferentes condições experimentais (temperatura, umidade, adubação, etc.) controladas durante o experimento;
- um animal em teste morre de uma dose específica de veneno, ou vive;

Nesses exemplos os resultados de interesse nas unidades observadas assumem apenas dois valores possíveis.

Representamos a i -ésima observação ou resposta, por uma variável aleatória Y_i , que sem perda de generalidade pode ser codificada pelos valores 0 ou 1. Frequentemente tem sido conveniente definir $Y_i = 1$ como *sucesso* e $Y_i = 0$ como *fracasso* (distribuição de Bernoulli com probabilidade de sucesso θ_i); trata-se portanto de uma variável dicotômica, (binária), tendo probabilidades:

$$Pr[Y_i = 1] = \theta_i$$

$$Pr[Y_i = 0] = 1 - \theta_i,$$

com média e variância dadas por,

$$\mathbb{E}(Y_i) = \theta_i$$

$$Var(Y_i) = \theta_i(1 - \theta_i).$$

Em estudos como estes, estamos interessados em verificar o comportamento da probabilidade de sucesso em função de um vetor de variáveis ou covariáveis explicativas $\{X_1, X_2, \dots, X_k\}$. Num planejamento de experimentos, este vetor de covariáveis incluem usualmente várias variáveis indicadoras associadas com blocos e fatores de tratamento, junto com informações quantitativas sobre vários aspectos do material experimental. Uma análise estatística em geral investiga a relação entre a probabilidade $\theta = \theta(\underline{X})$ e as variáveis explicativas $\underline{X} = (X_1, X_2, \dots, X_k)$.

Para investigar a relação entre a probabilidade de resposta θ e o vetor de covariadas \underline{X} , é conveniente construir um modelo capaz de discriminar o efeito de θ pela transformação em (X_1, X_2, \dots, X_k) . Na prática, este modelo usualmente incorpora suposições tais como correlação zero ou independência, falta de interação e linearidade.

Modelos lineares fazem um importante papel tanto em trabalhos teóricos quanto aplicados. Supor inicialmente que a dependência de θ sobre (X_1, X_2, \dots, X_k) ocorre através da combinação linear

$$\eta = \alpha + \sum_{j=1}^k X_j \beta_j \quad (1.1)$$

de coeficientes desconhecidos $\beta = (\beta_1, \dots, \beta_k)$. De (1.1), temos $-\infty < \eta < \infty$. Assim, expressar θ como combinação linear (1.1) seria inconsistente com as leis de probabilidade, pois θ estaria fora do intervalo $[0; 1]$. Uma forma alternativa para evitar este problema é o uso de uma transformação $h(\theta)$ que mapeie o intervalo unitário sobre a reta real $(-\infty, \infty)$. Duas funções de ligação são comumente usadas na prática:

1) *Logito ou função logística*

$$h_1(\theta) = \log\left(\frac{\theta}{1-\theta}\right) = \alpha + \sum_{j=1}^k X_j \beta_j; \quad (1.2)$$

2) *Probit ou função Normal inversa*

$$h_2(\theta) = \Phi^{-1}(\theta) = \alpha + \sum_{j=1}^k X_j \beta_j \quad (1.3)$$

onde

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t \exp\left\{-\frac{u^2}{2}\right\} du$$

De (1.2) temos,

$$\frac{\theta}{1-\theta} = \exp\left(\alpha + \sum_{j=1}^k X_j \beta_j\right),$$

$$\theta = \frac{\exp\left(\alpha + \sum_{j=1}^k X_j \beta_j\right)}{1 + \exp\left(\alpha + \sum_{j=1}^k X_j \beta_j\right)}.$$

Em (1.3) temos que:

$$\theta = \Phi\left(\alpha + \sum_{j=1}^k X_j \beta_j\right).$$

De forma geral, supor um vetor de variáveis explicativas \underline{X} e uma variável latente W com função de distribuição acumulada $F(w; x)$.

Consideremos que $Y = 1$ se e somente se $W > 0$, isto é,

$$\theta = Pr[Y = 1; x] = 1 - F(0; x)$$

Se considerarmos agora que a distribuição de W é a função logística, temos:

$$F(0; x) = \left(1 + \exp\left(\alpha + \sum_{j=1}^k X_j \beta_j\right)\right)^{-1} = Pr[Y = 0; x]$$

Então,
$$\theta = Pr[Y = 1; x] = 1 - F(0; x) = \frac{\exp\left(\alpha + \sum_{j=1}^k X_j \beta_j\right)}{1 + \exp\left(\alpha + \sum_{j=1}^k X_j \beta_j\right)}$$

Considerando W com distribuição $W \sim N\left(\alpha + \sum_{j=1}^k X_j \beta_j, 1\right)$, temos,

$$p = \Phi\left(\alpha + \sum_{j=1}^k X_j \beta_j\right)$$

1.1 Motivações, Modelos e Objetivos

‡ Frequentemente quando duas ou mais medidas são tomadas ao mesmo tempo para alguma unidade experimental, indivíduo, ou quando medidas repetidas são tomadas para mesma unidade experimental temos dados binários longitudinais. Normalmente assumimos que tais observações binárias para n unidades experimentais são independentes. Considerando que Y_i assume resposta 0 ou 1, Y_i tem distribuição de Bernoulli com probabilidade de sucesso θ_i , então métodos de regressão padrões são inadequados por varias razões:

- (i) as covariáveis podem ser fixas para todos os tempos ou variar nos tempos;
- (ii) existe o fator tempo;
- (iii) observações repetidas para a mesma unidade experimental tendem a serem correlacionadas.

O uso dos modelos anteriores da forma com que foram construídos pode não ser coerente, pois em muitas aplicações temos dados binários correlacionados, já que para um mesmo indivíduo são feitas mais de uma medida. Muitos modelos tem sido propostos para ajustar modelos de regressão com variáveis binárias correlacionadas (ver por exemplo, Haseman & Kupper, 1979).

Alguns modelos podem ser generalizados para permitir regressão de variáveis que caracterize cada observação binária correlacionada ou bloco o que permite estudar a relação entre várias covariáveis e medidas de efeitos dicotômicos. Em geral obter estimadores de máxima verossimilhança é trabalhoso e requer resoluções de equações que não tem solução analítica. Procedimentos Bayesianos podem evitar a necessidade dessas soluções numéricas usando técnicas de Monte Carlo em Cadeias de Markov (MCMC) (ver por exemplo, Gelfand, A.E. & Smith, A.F.M., 1990), em particular o algoritmo *Gibbs Sampling* o algoritmo *Metropolis-Hastings*. Além disso no método Bayesiano é possível incorporar ao modelo um conhecimento prévio a respeito do parâmetro de interesse para a análise, isso porque na inferência Bayesiana o parâmetro é considerado uma variável aleatória, com uma densidade de probabilidade chamada de *distribuição a priori do parâmetro*.

O algoritmo *Gibbs Sampling* é um método de amostragem, para se obter amostras de uma distribuição de interesse, que explora as distribuições condicionais completas através de um algoritmo iterativo que define uma cadeia de Markov (ver Apêndice A). O algoritmo *Metropolis-Hastings* é um usado quando algum componente do modelo não possui distribuições condicionais padronizadas (ver Apêndice B).

Prentice (1988) apresenta uma revisão completa de várias estratégias de modelagem usando análise de regressão linear generalizada de dados binários correlacionados com associação de covariáveis a cada resposta binária. Seguindo Liang e Zeger (1986) e Zeger e Liang (1986), Prentice considera o uso de equação de estimação generalizada

(GEE) aproximada para obter estimadores consistentes e assintoticamente normais para os coeficientes de regressão.

Dey e Chen (1996) apresentam uma análise Bayesiana dos modelos propostos por Prentice (1988) e também por Ochi e Prentice (1984).

Nesta dissertação, exploramos diferentes estratégias de modelagem para a análise de dados binários correlacionados sob o enfoque Bayesiano, considerando métodos de Monte Carlo em Cadeias de Markov (MCMC) para simular as distribuições a posteriori de interesse. Para discriminar possíveis modelos usados para dados binários, usamos os dois métodos a seguir:

- Resíduos de Pearson;
- Fator de Bayes.

O fator de Bayes será aproximado pelas amostras geradas pelo método MCMC (ver por exemplo, Kass, R.E. & Raftery, A.E., 1995; Gelfand, et al., 1992; Dey, et al., 1995).

1.1.1 Alguns Modelos Considerados na Literatura para Análise de Dados Binários Correlacionados

Os modelos considerados para análise de dados binários correlacionados podem ser classificados em dois grupos. O primeiro grupo apresenta modelos logísticos binários generalizados para dados multivariados, considerando uma particular representação parametrizada para as correlações que usam a noção de efeitos aleatórios na estrutura de regressão logística em termos de *pairwise odds ratio*; o segundo grupo de modelos é obtido introduzindo uma função de ligação que usa uma função de distribuição cumulativa inversa. Modelos do tipo Probit Multivariado (MVP) e *t* de Ligação Multivariada (MVT) são modelos neste grupo. O modelo MVP foi introduzido por Ashford e Sowden (1970) e também considerado por Amemiya (1985). Recentemente Chib e Greenberg (1995) usaram o modelo MVP numa abordagem Bayesiana.

Inicialmente introduzimos algumas notações que serão usadas no decorrer deste trabalho. Suponha que observamos uma resposta binária Y_{ij} (0 ou 1) na i -ésima observação e j -ésima variável e temos $x_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp_j})$ correspondendo a linha do vetor de regressão de dimensão p_j com $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, J$. (Note que

x_{ij1} pode ser igual a 1, o qual corresponde ao intercepto). Seja $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{iJ})'$ e assumimos que $Y_{i1}, Y_{i2}, \dots, Y_{iJ}$ são dependentes e Y_1, Y_2, \dots, Y_n são independentes. Sejam $y_i = (y_{i1}, y_{i2}, \dots, y_{iJ})'$ e $y = y_1, y_2, \dots, y_n$ os dados observados. Também seja $\beta_j = (\beta_{1j}, \beta_{2j}, \dots, \beta_{pj})'$ um vetor coluna de coeficientes de regressão com dimensão p_j para $j = 1, 2, \dots, J$, correspondente à covariável x_{ij} , e $\beta = (\beta'_1, \beta'_2, \dots, \beta'_J)'$.

A) Modelos do Primeiro Grupo

(i) Modelos de Mistura e Estratificado

Com a forma de modelos de mistura e estratificação considerado em Prentice (1988), assumir o modelo de regressão logístico binário para Y_{ij} dado x_{ij} , por,

$$P(Y_{ij} = y_{ij} | \alpha_i, \beta_j, x_{ij}) = \frac{\exp\{(\alpha_i + x_{ij}\beta_j)y_{ij}\}}{1 + \exp(\alpha_i + x_{ij}\beta_j)} \quad (1.4)$$

onde α_i denota um efeito aleatório na i -ésima observação, o qual especifica a correlação entre $Y_{i1}, Y_{i2}, \dots, Y_{iJ}$. Assumir também,

$$\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2). \quad (1.5)$$

O modelo acima também é conhecido na literatura como um modelo log normal.

(ii) Modelos Condicionais

Em algumas aplicações de dados binários correlacionados, há uma ordem natural dos indivíduos dentre os blocos e os coeficientes de regressão nos modelos que podem caracterizar dependências de interesse. Como segundo modelo considerar um modelo de regressão logística binária de efeitos aleatórios condicionais simétricos (ver por exemplo, Prentice 1988), o qual é definido por,

$$\begin{aligned} P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \dots, Y_{iJ} = y_{iJ} | \alpha_i, \beta, x_{i1}, \dots, x_{iJ}) \\ = \frac{\exp\left\{\sum_{j=0}^{y_{i1}-1} \alpha_i(j) + \sum_{j=1}^J x_{ij}\beta_j y_{ij}\right\}}{\sum_{y_{i1}^*=0}^1 \sum_{y_{i2}^*=0}^1 \cdots \sum_{y_{iJ}^*=0}^1 \exp\left\{\sum_{j=0}^{y_{i1}^*-1} \alpha_i(j) + \sum_{j=1}^J x_{ij}\beta_j y_{ij}^*\right\}}, \end{aligned} \quad (1.6)$$

onde $y_i = \sum_{j=1}^J y_{ij}$ e $y_i^* = \sum_{j=1}^J y_{ij}^*$. Em (1.6), $\alpha_i(j)$ é escolhido para ser um

polinômio em j . Por exemplo, a escolha $\alpha_i(j) = \alpha_i$ leva ao modelo,

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \dots, Y_{iJ} = y_{iJ} | \alpha_i, \beta, x_{i1}, \dots, x_{iJ})$$

$$= \frac{\exp \left\{ \alpha_i y_i + \sum_{j=1}^J x_{ij} \beta_j y_{ij} \right\}}{\sum_{y_{i1}^*=0}^1 \sum_{y_{i2}^*=0}^1 \cdots \sum_{y_{iJ}^*=0}^1 \exp \left\{ \alpha_i y_i^* + \sum_{j=1}^J x_{ij} \beta_j y_{ij}^* \right\}} \quad (1.7)$$

onde $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$.

B) Alguns Modelos do Segundo Grupo

(i) Modelo Probito Multivariado

Prentice (1988) considera a distribuição conjunta para $Y_{i1}, Y_{i2}, \dots, Y_{iJ}$ com marginais idênticas, o qual é essencialmente uma versão de modelos binários correlacionados. Aqui, consideramos o modelo probito multivariado (MVP); (ver por exemplo, Chib & Greenberg, 1995).

Introduzimos um vetor J -dimensional $w_i = (w_{i1}, w_{i2}, \dots, w_{iJ})'$ tal que

$$Y_{ij} = \begin{cases} 1 & \text{se } w_{ij} > 0 \\ 0 & \text{se } w_{ij} \leq 0 \end{cases} \quad (1.8)$$

e assumimos que

$$w_i \stackrel{iid}{\sim} N(X_i \beta, \Sigma) \quad (1.9)$$

onde

$$X_i = \begin{pmatrix} x_{i1} & 0 & \cdots & 0 \\ 0 & x_{i2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{iJ} \end{pmatrix}.$$

Em (1.9), consideramos $\Sigma = (\rho_{ij})_{J \times J}$ uma matriz de correlação tal que $\rho_{ij} = 1$ para assegurar a identificabilidade dos parâmetros (ver Chib & Greenberg, 1995, para detalhes). Portanto, a distribuição de w_i determina a distribuição conjunta de Y_i por (1.8) e a matriz de correlação Σ captura as correlações entre Y_{ij} 's. Especificamente, temos,

$$\begin{aligned}
& P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \dots, Y_{iJ} = y_{iJ} | \beta, \Sigma, X_i) \\
&= \int_{A_1} \int_{A_2} \dots \int_{A_J} \left(\frac{1}{\sqrt{2\pi}} \right)^J \frac{1}{|\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (w_i - X_i\beta)' \Sigma^{-1} (w_i - X_i\beta) \right\} dw_i
\end{aligned} \tag{1.10}$$

onde A_j é um intervalo tal que

$$A_j = \begin{cases} (-\infty, 0] & \text{se } y_{ij} = 0 \\ (0, \infty) & \text{se } y_{ij} = 1 \end{cases} \tag{1.11}$$

(ii) Modelos t de Ligação Multivariado

Uma generalização dos modelos probito multivariados são modelos t multivariado (MVT). Semelhante aos modelos MVP, introduzimos um vetor aleatório J -dimensional $w_i = (w_{i1}, w_{i2}, \dots, w_{iJ})'$ tal que

$$Y_{ij} = \begin{cases} 1 & \text{se } w_{ij} > 0 \\ 0 & \text{se } w_{ij} \leq 0 \end{cases} \tag{1.12}$$

e assumimos que

$$w_i \stackrel{iid}{\sim} t_v(X_i\beta, \Sigma)$$

onde $t_v(X_i\beta, \Sigma)$ tem uma função de densidade de probabilidade

$$\pi(w_i | v, X_i\beta, \Sigma) = \frac{\Gamma\left(\frac{1}{2}(v+J)\right)}{(\pi v)^{\frac{1}{2}J} \Gamma\left(\frac{1}{2}v\right) |\Sigma|^{\frac{1}{2}}} \left(1 + v^{-1}(w_i - X_i\beta)' \Sigma^{-1} (w_i - X_i\beta)\right)^{-\frac{1}{2}(v+J)} \tag{1.13}$$

Note que um caso especial de (1.13) com $v = 1$ é denominado distribuição de Cauchy multivariada e outro caso especial de (1.13) com $v \rightarrow \infty$ é a distribuição normal multivariada (1.9). As distribuições conjuntas de Y_i são dadas por

$$P(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \dots, Y_{iJ} = y_{iJ} | \beta, \Sigma, v, X_i) = \int_{A_1} \int_{A_2} \dots \int_{A_J} \pi(w_i | v, X_i\beta, \Sigma) dw_i \tag{1.14}$$

onde $\pi(w_i | v, X_i\beta, \Sigma)$ e A_j são dados em (1.13) e (1.11) respectivamente. Para simplificação podemos considerar v fixado, o que facilita a análise Bayesiana do modelo.

1.2 Organização da Dissertação

Esta dissertação encontra-se dividida em 6 capítulos. Neste capítulo introdutório apresentamos alguns modelos para dados binários sem correlação, as motivações e os objetivos do trabalho, e ainda alguns modelos para dados binários correlacionados encontrados na literatura.

No capítulo 2, apresentamos uma análise do modelo (1.4) proposto por Prentice (1988), que é um modelo logístico generalizado. Para este modelo consideramos inicialmente efeitos fixos, ou seja, sem assumir correlação entre as observações, e apresentamos uma análise clássica e uma análise Bayesiana do modelo. A seguir, tratamos o modelo considerando efeitos aleatórios (1.5) onde α_i , $i = 1, \dots, n$ captura a correlação entre as respostas, sob uma análise Bayesiana hierárquica, e ainda considerando uma mistura de distribuições normais.

Como uma alternativa ao modelo logístico (1.4) apresentamos no capítulo 3, o modelo probito generalizado, e sobre este modelo fizemos as mesmas considerações feitas ao modelo logístico no capítulo 2, para desenvolver uma análise Bayesiana usando métodos MCMC.

Para análise de dados onde temos réplicas de unidades experimentais ou frequências nas observações, consideramos no capítulo 4 o modelo logístico e o modelo probito incorporando as réplicas das unidades experimentais, e desenvolvemos uma análise Bayesiana para este caso.

Com a finalidade de verificar qual ou quais modelos são mais adequados para fazer inferências sobre dados binários correlacionados, apresentamos no Capítulo 5 alguns procedimentos para discriminação dos modelos propostos.

Num capítulo 6, introduzimos exemplos de aplicação para ilustrar a metodologia proposta nesta dissertação.

Capítulo 2

ANÁLISE DO MODELO DE PRENTICE

Neste capítulo inicialmente introduzimos uma análise estatística do modelo logístico de Prentice (1988), sob o ponto de vista clássico, usando estimadores de máxima verossimilhança considerando α_i ($i = 1, 2, \dots, n$) iguais à α , portanto um modelo de efeitos fixos que considera nula a correlação entre as repetições. Também introduzimos uma análise Bayesiana onde são utilizados métodos de simulação de amostras baseados no método de Monte Carlo em Cadeias de Markov, considerando efeitos fixos, e efeitos aleatórios (1.5), com uma distribuição normal para α_i , $i = 1, 2, \dots, n$ e com uma mistura de distribuições normais.

Considerar,

J : número de tempos;

n : número de observações ;

Y_{ij} (1 ou 0): a resposta para i -ésima observação ($i = 1, 2, \dots, n$) no j -ésimo tempo ($j = 1, 2, \dots, J$), com distribuição de Bernoulli com parâmetro θ_{ij} .

$x_{ij} = (x_{ij1}, x_{ij2}, \dots, x_{ijp_j})$: um vetor linha de regressão de dimensão p_j ;

A tabela a seguir apresenta a disposição dos dados binários em uma forma geral.

Tabela 2.1: Dados Binários Longitudinais

i	T_1	...	T_j	...	T_J
1	$y_{11}; (x_{111}, x_{112}, \dots, x_{11p_j})$...	$y_{1j}; (x_{1j1}, x_{1j2}, \dots, x_{1jp_j})$...	$y_{1J}; (x_{1J1}, x_{1J2}, \dots, x_{1JP_j})$
2	$y_{21}; (x_{211}, x_{212}, \dots, x_{21p_j})$...	$y_{2j}; (x_{2j1}, x_{2j2}, \dots, x_{2jp_j})$...	$y_{2J}; (x_{2J1}, x_{2J2}, \dots, x_{2JP_j})$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
i	$y_{i1}; (x_{i11}, x_{i12}, \dots, x_{i1p_j})$...	$y_{ij}; (x_{ij1}, x_{ij2}, \dots, x_{ijp_j})$...	$y_{iJ}; (x_{iJ1}, x_{iJ2}, \dots, x_{iJP_j})$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	$y_{n1}; (x_{n11}, x_{n12}, \dots, x_{n1p_j})$...	$y_{nj}; (x_{nj1}, x_{nj2}, \dots, x_{njp_j})$...	$y_{nJ}; (x_{nJ1}, x_{nJ2}, \dots, x_{nJP_j})$

O modelo proposto por Prentice (1988) é dado por

$$\theta_{ij} = \frac{\exp\{\alpha_i + x_{ij}\beta_j\}}{1 + \exp\{\alpha_i + x_{ij}\beta_j\}} \quad (2.1)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, J.$$

Associado às covariáveis, temos $\beta_j = (\beta_{1j}, \beta_{2j}, \dots, \beta_{p_jj})'$ um vetor coluna de coeficientes de regressão com dimensão p_j .

Tratando-se de uma resposta binária, $y_{ij} = 1$ ou $y_{ij} = 0$, com distribuição de Bernoulli de parâmetro θ_{ij} , ou seja,

$$Pr[Y_{ij} = y_{ij} | \alpha_i, \beta_j, x_{ij}] = \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \quad (2.2)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, J.$$

Substituindo a função de ligação (2.1) em (2.2), temos:

$$Pr[Y_{ij} = y_{ij} | \alpha_i, \beta_j, x_{ij}] = \frac{\exp\{(\alpha_i + x_{ij}\beta_j)y_{ij}\}}{1 + \exp\{\alpha_i + x_{ij}\beta_j\}} \quad (2.3)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, J.$$

2.1 Função de Verossimilhança

Toda informação contida nos dados sobre os parâmetros do modelo é expressa pela função de verossimilhança (FV).

Assumindo que as observações são independentes, a função de verossimilhança é obtida como o produto dos termos dados pela expressão (2.2), ou seja, o produto da probabilidade de ocorrência de cada resposta observada, podendo ser escrita na forma,

$$L(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) = \prod_{i=1}^n \prod_{j=1}^J \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \quad (2.4)$$

onde $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$, e \mathcal{D} representa o conjunto de dados (y_{ij}, x_{ij}) $i = 1, 2, \dots, n$; $j = 1, 2, \dots, J$. Portanto, de acordo com a expressão (2.1) a função de verossimilhança é dada por,

$$L(\underline{\alpha}, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) = \prod_{i=1}^n \prod_{j=1}^J \frac{\exp\{(\alpha_i + x_{ij}\beta_j)y_{ij}\}}{1 + \exp\{\alpha_i + x_{ij}\beta_j\}}. \quad (2.5)$$

2.2 Análise Clássica do Modelo

Inicialmente, consideramos efeitos fixos, ou seja, $\alpha_i = \alpha$ ($i = 1, 2, \dots, n$). Para fazer as inferências sobre os parâmetros do modelo usamos o método da máxima verossimilhança, como é feito usualmente para parâmetros de modelos logísticos.

O objetivo é estimar os parâmetros α e β_j , $j = 1, 2, \dots, J$, que maximizem a função de verossimilhança. Para isso tomamos o logaritmo da função de verossimilhança (2.5), para simplificação matemática. A expressão do logaritmo da função de verossimilhança é dado por,

$$\begin{aligned} \ell(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) &= \log L(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) \\ &= \sum_{i=1}^n \sum_{j=1}^J \left\{ (\alpha + x_{ij}\beta_j)y_{ij} - \log[1 + \exp(\alpha + x_{ij}\beta_j)] \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^J \left\{ \alpha y_{ij} + x_{ij}\beta_j y_{ij} - \log[1 + \exp(\alpha + x_{ij}\beta_j)] \right\} \\ &= \sum_{i=1}^n \sum_{j=1}^J \alpha y_{ij} + \sum_{i=1}^n \sum_{j=1}^J x_{ij}\beta_j y_{ij} - \sum_{i=1}^n \sum_{j=1}^J \log[1 + \exp(\alpha + x_{ij}\beta_j)]. \end{aligned} \quad (2.6)$$

Derivando o logaritmo da função verossimilhança em relação à α e β_j ,

($j = 1, 2, \dots, J$), temos:

$$\begin{aligned} \frac{\partial \ell}{\partial \alpha} &= \sum_{i=1}^n \sum_{j=1}^J y_{ij} - \sum_{i=1}^n \sum_{j=1}^J \frac{\exp\{(\alpha + x_{ij}\beta_j)\}}{1 + \exp\{\alpha + x_{ij}\beta_j\}} = 0 \\ \frac{\partial \ell}{\partial \beta_{pj}} &= \sum_{i=1}^n x_{ij} y_{ij} - \sum_{i=1}^n x_{ijp_j} \frac{\exp\{(\alpha + x_{ij}\beta_j)\}}{1 + \exp\{\alpha + x_{ij}\beta_j\}} = 0 \end{aligned} \quad (2.7)$$

$$j = 1, 2, \dots, J.$$

Igualando as equações (2.7) à zero, temos:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^J y_{ij} - \sum_{i=1}^n \sum_{j=1}^J \frac{\exp\{(\hat{\alpha} + x_{ij}\hat{\beta}_j)\}}{1 + \exp\{\hat{\alpha} + x_{ij}\hat{\beta}_j\}} &= 0 \\ \sum_{i=1}^n x_{ij} y_{ij} - \sum_{i=1}^n x_{ijp_j} \frac{\exp\{(\hat{\alpha} + x_{ij}\hat{\beta}_j)\}}{1 + \exp\{\hat{\alpha} + x_{ij}\hat{\beta}_j\}} &= 0 \end{aligned} \quad (2.8)$$

$$j = 1, 2, \dots, J$$

As expressões (2.8) são as funções Escore. Os valores de α e β_j que são soluções das equações (2.8) serão os EMV denotados por $\hat{\alpha}$ e $\hat{\beta}_j$. Um método prático para obtenção das soluções é o método de Newton-Raphson, que é um algoritmo numérico, que coincide com o proposto por Fisher (Rao, 1973). As estimativas de máxima verossimilhança são obtidas através do seguinte esquema iterativo:

$$\begin{pmatrix} \hat{\alpha}_{(n+1)} \\ \hat{\beta}_{j(n+1)} \end{pmatrix} = \begin{pmatrix} \hat{\alpha}_{(n)} \\ \hat{\beta}_{j(n)} \end{pmatrix} + \left(\mathcal{I}(\hat{\alpha}_{(n)}, \hat{\beta}_{j(n)}) \right)^{-1} S(\hat{\alpha}_{(n)}, \hat{\beta}_{j(n)})$$

onde \mathcal{I} é a matriz de informação de Fisher e S o vetor Escore.

As segundas derivadas do logaritmo da função de verossimilhança são dadas por:

$$\frac{\partial^2 \ell}{\partial \alpha^2} = - \sum_{i=1}^n \sum_{j=1}^J \frac{\exp\{(\alpha + x_{ij}\beta_j)\}}{[1 + \exp(\alpha + x_{ij}\beta_j)]^2}$$

$$\frac{\partial^2 \ell}{\partial \beta_{pj}^2} = - \sum_{i=1}^n x_{ij}^2 \frac{\exp\{(\alpha + x_{ij}\beta_j)\}}{[1 + \exp(\alpha + x_{ij}\beta_j)]^2}$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta_{pj}} = - \sum_{i=1}^n x_{ij} \frac{\exp\{(\alpha + x_{ij}\beta_j)\}}{[1 + \exp(\alpha + x_{ij}\beta_j)]^2}$$

$$\frac{\partial^2 \ell}{\partial \beta_{pj} \partial \beta_{pk}} = 0 \quad \text{para } j \neq k \quad \text{onde } \begin{cases} j = 1, 2, \dots, J \\ k = 1, 2, \dots, J \end{cases}$$

$$\left\{ \begin{array}{l} \frac{\partial^2 \ell}{\partial \beta_{pj} \partial \beta_{pk}} = 0 \\ \frac{\partial^2 \ell}{\partial \beta_{pj} \partial \beta_{pk}} = - \sum_{i=1}^n x_{ij} x_{ik} \frac{\exp\{(\alpha + x_{ij}\beta_j)\}}{[1 + \exp(\alpha + x_{ij}\beta_j)]^2} \quad \text{c.c.} \end{array} \right. \quad \text{para } j \neq k$$

Os estimadores de máxima verossimilhança têm distribuição normal assintótica (ver por exemplo, Gourieroux & Monfort, 1995; Bickel & Doksum, 1977) , ou seja:

$$(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_J) \stackrel{a}{\sim} N[(\alpha, \beta_1, \beta_2, \dots, \beta_J), \mathcal{I}^{-1}]$$

onde \mathcal{I} é a matriz de informação de Fisher.

No passo final das iterações além de obtermos uma estimativa dos parâmetros temos uma estimativa da matriz de variância-covariância dos estimadores.

Este processo pode ser árduo se tivermos um número de parâmetros muito grande para serem estimados. Para usar este método de resolução iterativa temos que dar valores adequados muito aproximados dos valores dos parâmetros ou *chutes* iniciais, a fim de obter convergência para os pontos $\hat{\alpha}$ e $\hat{\beta}_j$ com poucas iterações . Uma forma de obtermos os *chutes* iniciais é fazer uma aproximação do logaritmo da verossimilhança (ver por exemplo,

Cox, 1970) como a seguir:

$$\ell(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) = \sum_{i=1}^n \sum_{j=1}^J \alpha y_{ij} + \sum_{i=1}^n \sum_{j=1}^J x_{ij} \beta_j y_{ij} - K,$$

onde

$$K = \sum_{i=1}^n \sum_{j=1}^J \log[1 + \exp(\alpha + x_{ij} \beta_j)].$$

Considerando os resultados,

- $\log(1 + x) \simeq x - \frac{x^2}{2} + \frac{x^3}{3} - \dots + \frac{(-1)^n x^{n+1}}{n+1} + \dots$
- $e^x \simeq 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \dots$

podemos escrever,

$$K \simeq e^\alpha \sum_{i=1}^n \sum_{j=1}^J [1 + x_{ij} \beta_j + \frac{1}{2} (x_{ij} \beta_j)^2]$$

Assim o logaritmo da função de verossimilhança aproximada é dado por,

$$\ell^*(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) = \sum_{i=1}^n \sum_{j=1}^J \alpha y_{ij} + \sum_{i=1}^n \sum_{j=1}^J x_{ij} \beta_j y_{ij} - e^\alpha \sum_{i=1}^n \sum_{j=1}^J [1 + x_{ij} \beta_j + \frac{1}{2} (x_{ij} \beta_j)^2].$$

Derivando $\ell^*(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D})$ nos parâmetros e igualando a zero, obtemos facilmente as estimativas aproximadas dos EMV's dos parâmetros que servirão de chutes iniciais para o método de Newton-Raphson.

2.3 Análise Bayesiana do Modelo

Modelo 1: Modelo logístico com efeitos fixos.

Nesta seção consideramos o uso de métodos Bayesianos para o modelo de Prentice (1988). Para isso, usamos métodos MCMC em particular o algoritmo Gibbs Sampling (ver, Gelfand, A.E. & Smith, A.F.M, 1990) e o algoritmo Metropolis-Hastings (ver, Chib, S. & Greenberg, E., 1995), os quais tem sido muito utilizados em análise Bayesiana.

Ambos os métodos baseiam-se nas distribuições condicionais a posteriori para geração dos parâmetros, sendo o algoritmo Gibbs Sampling usado quando a distribuição condicional possui uma forma padronizada conhecida e o algoritmo Metropolis-Hastings usado quando a distribuição condicional não tem uma forma conhecida.

Assumimos as seguintes densidades a priori informativas:

- (i) $\alpha \sim N(\mu_0, \sigma_0^2)$ com μ_0 e σ_0^2 conhecidos;
(ii) $\beta_{kj} \sim N(b_{kj}, c_{kj}^2)$ com b_{kj} e c_{kj}^2 conhecidos; (2.9)

$$j = 1, 2, \dots, J; \quad k = 1, 2, \dots, p_j.$$

Assumindo independência das distribuições a priori, a distribuição a priori conjunta é dada por,

$$\pi(\alpha, \beta_1, \beta_2, \dots, \beta_J) \propto \exp\left\{-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right\} \prod_{j=1}^J \prod_{k=1}^{p_j} \exp\left\{-\frac{(\beta_{kj} - b_{kj})^2}{2c_{kj}^2}\right\} \quad (2.10)$$

A densidade a posteriori conjunta é tal que,

$$\pi(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) \propto \pi(\alpha, \beta_1, \beta_2, \dots, \beta_J) L(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}).$$

Desta forma, temos,

$$\begin{aligned} \pi(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) \propto & \left[\exp\left\{-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right\} \prod_{j=1}^J \prod_{k=1}^{p_j} \exp\left\{-\frac{(\beta_{kj} - b_{kj})^2}{2c_{kj}^2}\right\} \right] \\ & \cdot \left[\prod_{i=1}^n \prod_{j=1}^J \frac{\exp\{(\alpha + x_{ij}\beta_j)y_{ij}\}}{1 + \exp\{\alpha + x_{ij}\beta_j\}} \right]. \end{aligned} \quad (2.11)$$

Para obtermos uma amostra da densidade a posteriori marginal dos parâmetros $(\alpha, \beta_1, \beta_2, \dots, \beta_J)$ com o uso do método de Monte Carlo em Cadeias de Markov, temos as seguintes distribuições condicionais a posteriori:

(i)

$$\pi(\alpha | \beta_1, \beta_2, \dots, \beta_J, \mathcal{D}) \propto \exp\left\{-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right\} \prod_{i=1}^n \prod_{j=1}^J \frac{\exp\{(\alpha + x_{ij}\beta_j)y_{ij}\}}{1 + \exp\{\alpha + x_{ij}\beta_j\}}, \quad \text{ou}$$

$$\pi(\alpha | \beta_1, \beta_2, \dots, \beta_J, \mathcal{D}) \propto N(\mu_0, \sigma_0^2) \cdot \Psi_1(\alpha, \beta_1, \beta_2, \dots, \beta_J)$$

onde

$$\Psi_1(\alpha, \beta_1, \beta_2, \dots, \beta_J) = \exp \left\{ \alpha y_{..} - \sum_{i=1}^n \sum_{j=1}^J \log \left[1 + \exp \left(\alpha + \sum_{l=1}^{p_j} x_{ijl} \beta_{lj} \right) \right] \right\}$$

$$y_{..} = \sum_{i=1}^n \sum_{j=1}^J y_{ij}$$

(ii)

$$\pi(\beta_{kl} | \alpha, \beta_1, \dots, \beta_{j-1}, \underset{\sim}{\beta}_{kl}, \beta_{j+1}, \dots, \beta_J, \mathcal{D}) \propto$$

$$\propto \exp \left\{ -\frac{(\beta_{kl} - b_{kl})^2}{2c_{kl}^2} \right\} \prod_{i=1}^n \frac{\exp\{(\alpha + x_{ij}\beta_j)y_{ij}\}}{1 + \exp\{\alpha + x_{ij}\beta_j\}}$$

$$\propto N(b_{kl}, c_{kl}^2) \cdot \Psi_2(\alpha, \beta_1, \beta_2, \dots, \beta_J, \mathcal{D})$$

onde

$$\Psi_2(\alpha, \beta_1, \beta_2, \dots, \beta_J, \mathcal{D}) = \exp \left\{ a_{kl} \beta_{kl} - \sum_{i=1}^n \log [1 + \exp(\alpha + x_{ij} \beta_j)] \right\};$$

$$a_{kl} = \sum_{i=1}^n x_{iik} y_{ii}; \quad k = 1, \dots, p_j; \quad l, j = 1, \dots, J$$

e $\underset{\sim}{\beta}_{kl}$ é um vetor de dimensão $p_l - 1$, ou seja:

$$\underset{\sim}{\beta}_{kl} = \begin{cases} (\beta_{2l}, \beta_{3l}, \dots, \beta_{p_l l}) & \text{se } k = 1 \\ (\beta_{1l}, \dots, \beta_{k-1,l}, \beta_{k+1,l}, \dots, \beta_{p_l l}) & \text{se } 2 \leq k \leq p_{l-1} \\ (\beta_{1l}, \dots, \beta_{p_{l-1},l}) & \text{se } k = p_l \end{cases} \quad (2.12)$$

e

$$\begin{cases} l = 1, 2, \dots, J \\ k = 1, 2, \dots, p_j \end{cases}$$

Como podemos observar, nenhuma das distribuições condicionais são padronizadas, portanto devemos usar o algoritmo de Metropolis-Hastings para simulação dos parâmetros,

e a partir dos valores simulados obter a média, variância e outras inferências a posteriori de interesse.

Para o uso algoritmo de Metropolis-Hastings consideramos a parte conhecida (padronizada) das distribuições condicionais a posteriori como núcleo de transição do algoritmo. Para verificar a convergência da amostra simulada de cada parâmetro para uma distribuição estabilizada utilizamos o algoritmo proposto por Gelman e Rubin (1992).

2.4 Análise Bayesiana Hierárquica Assumindo Efeitos Aleatórios Normais

Modelo 2: Modelo logístico assumindo efeitos aleatórios normais - $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$;

$$i = 1, \dots, n$$

Assumir agora que α_i , ($i = 1, \dots, n$) são efeitos aleatórios independentes com distribuição normal, isto é,

$$\alpha_i \sim N(0, \sigma_\alpha^2) \quad (2.13)$$

onde σ_α^2 é desconhecido com uma distribuição conhecida, portanto σ_α^2 é um hiperparâmetro e $N(0, \sigma_\alpha^2)$ uma *distribuição a priori hierárquica* para α_i , ($i = 1, \dots, n$), sendo assim necessário a atribuição de uma distribuição a priori para este hiperparâmetro.

Consideramos as seguintes distribuições a priori para σ_α^2 e β_{lj} , $l = 1, \dots, p_j$, $j = 1, \dots, J$.

- (i) $\sigma_\alpha^2 \sim \mathcal{IG}(a, b)$ com a e b conhecidos;
 - (ii) $\beta_{lj} \sim N(b_{lj}, c_{lj}^2)$ com b_{lj} e c_{lj}^2 conhecidos,
- $$l = 1, \dots, p_j, \quad j = 1, \dots, J \quad (2.14)$$

onde \mathcal{IG} é uma distribuição Gama Invertida,

$$\pi(\sigma_\alpha^2) = \frac{1}{\Gamma(a)} b^{-a} (\sigma_\alpha^2)^{-(a+1)} \exp\left\{-\frac{1}{b\sigma_\alpha^2}\right\}.$$

Isto é,

$$\pi(\sigma_\alpha^2) \propto (\sigma_\alpha^2)^{-(a+1)} \exp\left\{-\frac{1}{b\sigma_\alpha^2}\right\} \quad (2.15)$$

para $\sigma_\alpha^2 > 0$ com média $\frac{1}{b(a-1)}$ para $a > 1$ e variância $\frac{1}{b^2(a-1)^2(a-2)}$ para $a > 2$.

Assumir independência a priori para os parâmetros. A distribuição a posteriori conjunta (a partir de (2.5), (2.13) e (2.14)) para $\underline{\alpha} = (\alpha_1, \dots, \alpha_n)$, σ_α^2 e β_1, \dots, β_J é dada por,

$$\begin{aligned} \pi(\underline{\alpha}, \sigma_\alpha^2, \beta_1, \dots, \beta_J | \mathcal{D}) &\propto \\ &\propto (\sigma_\alpha^2)^{-\left(\frac{n}{2}+a+1\right)} \exp\left\{-\frac{1}{2\sigma_\alpha^2}\left[\frac{2}{b} + \sum_{i=1}^n \alpha_i^2\right]\right\} \\ &\cdot \left[\prod_{i=1}^{p_j} \prod_{j=1}^J \exp\left\{-\frac{1}{2c_{ij}^2}(\beta_{ij} - b_{ij})^2\right\}\right] \\ &\cdot \left[\prod_{i=1}^n \prod_{j=1}^J \frac{\exp\{(\alpha_i + x_{ij}\beta_j)y_{ij}\}}{1 + \exp\{\alpha_i + x_{ij}\beta_j\}}\right]. \end{aligned} \quad (2.16)$$

As distribuições condicionais necessárias para o algoritmo Gibbs sampling são dadas por,

i)

$$\pi(\alpha_i | \sigma_\alpha^2, \beta_1, \dots, \beta_J, \mathcal{D}) \propto \exp\left\{-\frac{\alpha_i^2}{2\sigma_\alpha^2}\right\} \frac{\exp\{\alpha_i y_i\}}{\prod_{j=1}^J \{1 + \exp(\alpha_i + x_{ij}\beta_j)\}}$$

para $i = 1, \dots, n$, isto é,

$$\pi(\alpha_i | \sigma_\alpha^2, \beta_1, \dots, \beta_J, \mathcal{D}) \propto N\{0, \sigma_\alpha^2\} \Psi_3(\alpha_i, \beta_1, \dots, \beta_J)$$

onde

$$\Psi_3(\alpha_i, \beta_1, \dots, \beta_J) = \exp\left\{\alpha_i \sum_{j=1}^J y_{ij} - \sum_{j=1}^J \log[1 + \exp(\alpha_i + x_{ij}\beta_j)]\right\}$$

ii)

$$\begin{aligned} \pi(\sigma_\alpha^2 | \underline{\alpha}, \beta_1, \dots, \beta_J, \mathcal{D}) &\propto (\sigma_\alpha^2)^{-\left(\frac{n}{2}+a+1\right)} \exp\left\{-\frac{1}{\sigma_\alpha^2}\left[\frac{1}{b} + \frac{1}{2} \sum_{i=1}^n \alpha_i^2\right]\right\} \\ &\propto (\sigma_\alpha^2)^{-\left(\frac{n}{2}+a+1\right)} \exp\left\{-\frac{1}{\sigma_\alpha^2 \left[\frac{1}{b} + \frac{1}{2} \sum_{i=1}^n \alpha_i^2\right]^{-1}}\right\} \end{aligned}$$

isto é,

$$\sigma_\alpha^2 | \underline{\alpha}, \beta_1, \dots, \beta_J, \mathcal{D} \sim IG\left[\frac{n}{2} + a; \left(\frac{1}{b} + \frac{1}{2} \sum_{i=1}^n \alpha_i^2\right)^{-1}\right]$$

iii)

$$\pi(\beta_{kl} | \underline{\alpha}, \beta_1, \dots, \beta_{j-1}, \beta_{\sim_{kl}}, \beta_{j+1}, \dots, \beta_J, \mathcal{D}) \propto N(b_{kl}, c_{kl}^2) \frac{\exp\{\beta_{kl} a_{kl}\}}{\prod_{i=1}^n \{1 + \exp(\alpha_i + x_{ij} \beta_j)\}}$$

isto é,

$$\pi(\beta_{kl} | \underline{\alpha}, \beta_1, \dots, \beta_{j-1}, \beta_{\sim_{kl}}, \beta_{j+1}, \dots, \beta_J, \mathcal{D}) \propto N(b_{kl}, c_{kl}^2) \Psi_4(\underline{\alpha}, \beta_1, \dots, \beta_J),$$

onde

$$\Psi_4(\underline{\alpha}, \beta_1, \dots, \beta_J) = \exp\left\{ \beta_{kl} a_{kl} - \sum_{j=1}^n \log[1 + \exp(\alpha_i + x_{ij} \beta_j)] \right\}$$

$$, a_{kl} = \sum_{i=1}^n x_{ik} y_{il}, \quad k = 1, \dots, p_j, \quad l, j = 1, \dots, J, \text{ e } \beta_{\sim_{kl}} \text{ definido em (2.12).}$$

Observar que α_i e β_{kl} devem ser gerados a partir do algoritmo Metropolis-Hastings.

2.5 Modelo de Efeitos Aleatórios com uma Mistura de Distribuições Normais

Modelo 3: Modelo logístico assumindo os efeitos aleatórios com uma mistura de distribuições Normais.

Quando assumimos $\alpha_i, i = 1, \dots, n$ como efeitos aleatórios independentes, podemos ter uma multimodalidade e portanto modelar estes parâmetros por uma única distribuição normal pode não ser apropriado. Sendo assim para maior flexibilidade, e consistência com a estratégia de modelagem hierárquica, construímos tais distribuições como misturas de distribuições normais.

Procurando construir um modelo mais realista assumimos para $\alpha_i (i = 1, \dots, n)$ uma mistura de duas distribuições normais, com densidade,

$$\begin{aligned} \pi(\alpha_i) &= \sum_{k=1}^2 p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2) \\ &= p_1 \phi_1(\alpha_i | \mu_1, \sigma_1^2) + (1 - p_1) \phi_2(\alpha_i | \mu_2, \sigma_2^2) \end{aligned} \tag{2.17}$$

onde ϕ_k denota uma densidade normal $N(\mu_k, \sigma_k^2)$, com a restrição $p_1 + p_2 = 1$.

Assumimos as seguintes distribuições a priori para μ_k , σ_k^2 , β_{lj} e p_1 :

- (i) $\mu_k \sim N(d_k, e_k^2)$ com d_k e e_k^2 conhecidos, $k = 1, 2$;
 - (ii) $\sigma_k^2 \sim \mathcal{IG}(a_k, b_k)$ com a_k e b_k conhecidos, $k = 1, 2$;
 - (iii) $\beta_{lj} \sim N(b_{lj}, c_{lj}^2)$ com b_{lj} e c_{lj}^2 conhecidos, $l = 1, \dots, p_j$; $j = 1, \dots, J$
 - (iv) $p_1 \sim \mathcal{B}(f, g)$ com f e g conhecidos
- (2.18)

onde $\mathcal{B}(f, g)$ denota uma distribuição beta com densidade

$$\pi(p_1) \propto p_1^{f-1} (1-p_1)^{g-1}, \quad 0 < p_1 < 1.$$

Assumir independência a priori entre os parâmetros. A distribuição a posteriori conjunta para $\underline{\alpha} = (\alpha_1, \dots, \alpha_n)$, $\underline{\mu} = (\mu_1, \mu_2)$, p_1 e $\underline{\sigma}^2 = (\sigma_1^2, \sigma_2^2)$ é dada por,

$$\begin{aligned} \pi(\underline{\alpha}, \beta_1, \dots, \beta_J, \underline{\mu}, \underline{\sigma}^2, p_1 | \mathcal{D}) &\propto \prod_{i=1}^n \left\{ \sum_{k=1}^2 p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2) \right\} \cdot \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(a_k+1)} \exp \left[\frac{-1}{b_k \sigma_k^2} \right] \right\} \\ &\cdot \left\{ \prod_{k=1}^2 \exp \left[-\frac{1}{2e_k^2} (\mu_k - d_k)^2 \right] \right\} \cdot p_1^{f-1} (1-p_1)^{g-1} \\ &\cdot \left\{ \prod_{l=1}^{p_j} \prod_{j=1}^J \exp \left[-\frac{1}{2c_{lj}^2} (\beta_{lj} - b_{lj})^2 \right] \right\} \\ &\cdot \frac{\exp \left\{ \sum_{i=1}^n \sum_{j=1}^J (\alpha_i + x_{ij} \beta_j) y_{ij} \right\}}{\prod_{i=1}^n \prod_{j=1}^J \left\{ 1 + \exp(\alpha_i + x_{ij} \beta_j) \right\}}. \end{aligned}$$

(2.19)

Como podemos observar, a mistura de distribuições normais envolve uma soma de distribuições dificultando o cálculo das distribuições à posteriori marginais dos parâmetros do modelo, para uma análise Bayesiana. Para simplificar as distribuições condicionais necessárias para o algoritmo Gibbs Sampling, consideramos a introdução de variáveis latentes, que irá transformar o modelo de misturas em um modelo de componentes independentes (ver por exemplo, Tanner & Wong, 1987), dados por $\underline{z}_i = (z_{i1}, z_{i2})$, $z_{i1} + z_{i2} = 1$,

$i = 1, \dots, n$ onde z_{i1} tem uma distribuição de Bernoulli $b(1, h_{i1})$, e

$$h_{i1} = \frac{p_1 \phi_1(\alpha_i | \mu_1, \sigma_1^2)}{p_1 \phi_1(\alpha_i | \mu_1, \sigma_1^2) + (1 - p_1) \phi_2(\alpha_i | \mu_2, \sigma_2^2)} \quad (2.20)$$

Isto é,

$$\begin{aligned} \pi(\tilde{z}_1, \dots, \tilde{z}_n) &\propto \prod_{i1}^n h_{i1}^{z_{i1}} (1 - h_{i1})^{z_{i2}} \\ &\propto \frac{\prod_{i=1}^n \prod_{k=1}^2 \left[p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2) \right]^{z_{ik}}}{\prod_{i=1}^n \left\{ \sum_{k=1}^2 p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2) \right\}} \end{aligned} \quad (2.21)$$

onde,

$$z_{i1} = \begin{cases} 1 & \text{com probabilidade } h_{i1}, \\ 0 & \text{com probabilidade } 1 - h_{i1}. \end{cases}$$

Combinando-se (2.19) com (2.21), temos a seguinte densidade à posteriori conjunta para os parâmetros (sendo $\tilde{z} = (\tilde{z}_1, \dots, \tilde{z}_n)$),

$$\begin{aligned} \pi(\underline{\alpha}, \beta_1, \dots, \beta_J, \mu, \sigma^2, p_1 | \tilde{z}, \mathcal{D}) &\propto \prod_{i=1}^n \prod_{k=1}^2 \left[p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2) \right]^{z_{i2}} \\ &\cdot \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(a_k+1)} \exp \left[\frac{-1}{b_k \sigma_k^2} \right] \right\} \cdot p_1^{f-1} (1 - p_1)^{g-1} \\ &\cdot \left\{ \prod_{k=1}^2 \exp \left[-\frac{1}{2e_k^2} (\mu_k - d_k)^2 \right] \right\} \\ &\cdot \left\{ \prod_{l=1}^{p_j} \prod_{j=1}^J \exp \left[-\frac{1}{2c_{lj}^2} (\beta_{lj} - b_{lj})^2 \right] \right\} \\ &\cdot \frac{\exp \left\{ \sum_{i=1}^n \sum_{j=1}^J (\alpha_i + x_{ij} \beta_j) y_{ij} \right\}}{\prod_{i=1}^n \prod_{j=1}^J \left\{ 1 + \exp(\alpha_i + x_{ij} \beta_j) \right\}} \end{aligned} \quad (2.22)$$

De (2.22), encontramos as distribuições condicionais necessárias para o algoritmo Gibbs sampling,

(i)

$$\pi(\alpha_i | \beta_1, \dots, \beta_J, \mu, \sigma^2, p_1, z, \mathcal{D}) \propto \frac{\phi_1^{z_{i1}}(\alpha_i | \mu_1, \sigma_1^2) \phi_2^{z_{i2}}(\alpha_i | \mu_2, \sigma_2^2) \exp(\alpha_i y_i)}{\prod_{j=1}^J \left\{ 1 + \exp(\alpha_i + x_{ij} \beta_j) \right\}}$$

onde $y_i = \sum_{j=1}^J y_{ij}$.

Isto é,

$$\pi(\alpha_i | \beta_1, \dots, \beta_J, \mu, \sigma^2, p_1, z, \mathcal{D}) \propto \exp \left\{ -\frac{(z_{i1}\sigma_2^2 + z_{i2}\sigma_1^2)}{2\sigma_1^2\sigma_2^2} \left[\alpha_i - \frac{z_{i1}\mu_1\sigma_2^2 + z_{i2}\mu_2\sigma_1^2}{z_{i1}\sigma_2^2 + z_{i2}\sigma_1^2} \right]^2 \right\} \cdot \frac{\exp\{\alpha_i y_i\}}{\prod_{j=1}^J \left\{ 1 + \exp(\alpha_i + x_{ij} \beta_j) \right\}}$$

ou,

$$\pi(\alpha_i | \beta_1, \dots, \beta_J, \mu, \sigma^2, p_1, z, \mathcal{D}) \propto N \left\{ \frac{z_{i1}\mu_1\sigma_2^2 + z_{i2}\mu_2\sigma_1^2}{z_{i1}\sigma_2^2 + z_{i2}\sigma_1^2}, \frac{\sigma_1^2\sigma_2^2}{z_{i1}\sigma_2^2 + z_{i2}\sigma_1^2} \right\} \cdot \Psi_5(\alpha_i, \beta_1, \dots, \beta_J),$$

onde,

$$\Psi_5(\alpha_i, \beta_1, \dots, \beta_J) = \exp \left\{ \alpha_i y_i - \sum_{j=1}^J \log \left[1 + \exp(\alpha_i + x_{ij} \beta_j) \right] \right\}$$

para $i = 1, \dots, n$.

(ii)

$$\pi(\sigma_k^2 | \alpha_i, \beta_1, \dots, \beta_J, \mu, \sigma_k^2, p_1, z, \mathcal{D}) \propto (\sigma_k^2)^{-(a_k+1)} \exp \left\{ -\frac{1}{b_k \sigma_k^2} \right\} (\sigma_k^2)^{-\frac{v_k}{2}} \cdot \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{i=1}^n z_{ik} (\alpha_i - \mu_k)^2 \right\},$$

onde $v_k = \sum_{i=1}^n z_{ik}$

Isto é,

$$\sigma_k^2 | \alpha_i, \beta_1, \dots, \beta_J, \mu, p_1, z, \mathcal{D} \sim \text{IG} \left[a_k + \frac{v_k}{2}; \left(\frac{1}{b_k} + \frac{\sum_{i=1}^n z_{ik} (\alpha_i - \mu_k)^2}{2} \right)^{-1} \right]$$

onde $k \neq \underline{k}$, $k, \underline{k} = 1, 2$;

(iii)

$$p_1 | \alpha, \beta_1, \dots, \beta_J, \mu, \sigma^2, z, \mathcal{D} \sim B(f + v_1, g + v_2);$$

(iv)

$$\pi(\beta_{kl} | \alpha, \beta_1, \dots, \beta_{j-1}, \beta_{\underline{kl}}, \beta_{j+1}, \dots, \beta_J, \mu, \sigma^2, p_1, z, \mathcal{D}) \propto N(b_{kl}, c_{kl}^2) \Psi_6(\alpha, \beta_1, \dots, \beta_J),$$

onde

$$\Psi_6(\alpha, \beta_1, \dots, \beta_J) = \exp \left\{ \beta_{kl} a_{kl} - \sum_{i=1}^n \log[1 + \exp(\alpha_i + x_{ij} \beta_j)] \right\},$$

$$a_{kl} = \sum_{i=1}^n x_{ik} y_{il}$$

β definido em (2.12)

$$\beta_{\underline{kl}} = 1, \dots, p_j; \quad l, j = 1, \dots, J;$$

(v)

$$\pi(\mu_k | \alpha, \beta_1, \dots, \beta_J, \mu_{\underline{k}}, \sigma^2, p_1, z, \mathcal{D}) \propto \exp \left\{ -\frac{1}{2e_k^2} (\mu_k - d_k)^2 \right\} \cdot \prod_{i=1}^n \exp \left\{ -\frac{z_{ik}}{2\sigma_k^2} (\alpha_i - \mu_k)^2 \right\}$$

Isto é

$$\pi(\mu_k | \alpha, \beta_1, \dots, \beta_J, \mu_{\underline{k}}, \sigma^2, p_1, \mathcal{D}, z) \sim N \left\{ \frac{d_k \sigma_k^2 + e_k^2 \sum_{i=1}^n \alpha_i Z_{ik}}{\sigma_k^2 + e_k^2 \sum_{i=1}^n Z_{ik}}, \frac{e_k^2 \sigma_k^2}{\sigma_k^2 + e_k^2 \sum_{i=1}^n Z_{ik}} \right\},$$

onde $k \neq \underline{k}$, $k, \underline{k} = 1, 2$.

Observações:

- Considerando $\Theta = (\alpha, \beta_1, \dots, \beta_J, \mu, \sigma^2, p_1)$, para gerar amostras da distribuição à posteriori conjunta (2.22) usamos o algoritmo Gibbs sampling. Começamos o método iterativo com valores iniciais $\Theta^{(0)} = (\Theta_1^{(0)}, \dots, \Theta_2^{(0)})$, e seguindo os seguintes passos:

- (i) Geramos uma amostra $\tilde{z}^{(1)} = (\tilde{z}_1^{(1)}, \dots, \tilde{z}_2^{(1)})$ da equação (2.21).
- (ii) Geramos uma amostra de Θ das distribuições condicionais

$$\begin{aligned} &\pi(\Theta_1 | \Theta_2^{(0)}, \dots, \Theta_p^{(0)}, \tilde{z}^{(1)}, \tilde{x}, \tilde{z}), \\ &\pi(\Theta_2 | \Theta_1^{(1)}, \Theta_3^{(0)}, \dots, \Theta_p^{(0)}, \tilde{z}^{(1)}, \tilde{x}, \tilde{z}), \\ &\vdots \\ &\pi(\Theta_p | \Theta_1^{(1)}, \dots, \Theta_{p-1}^{(1)}, \tilde{z}^{(1)}, \tilde{x}, \tilde{z}), \end{aligned}$$

Então continuamos as iterações repetindo os passos (i) e (ii).

- Observamos também que os parâmetros $\tilde{\alpha}$, $\tilde{\mu}$ e $\tilde{\beta}$ são gerados através do algoritmo Metropolis-Hastings.

Capítulo 3

ANÁLISE DO MODELO PROBITO

No estudo de modelos de regressão para o caso de onde a variável resposta Y_{ij} é binária, com distribuição de Bernoulli(θ_{ij}), ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, J$), a probabilidade de sucesso θ_{ij} pode variar em função de um vetor de covariáveis x_{ij} . Como uma alternativa ao modelo de regressão logística consideramos neste capítulo a função de ligação probito ou função Normal inversa.

Considerar Y_{ij} (1 ou 0) a resposta para i -ésima observação no j -ésimo tempo, com uma distribuição de Bernoulli com parâmetro θ_{ij} , ou seja,

$$\begin{aligned} Pr[Y_{ij} = y_{ij} | \alpha_i, \beta_j, x_{ij}] &= \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \\ i &= 1, \dots, n; \quad j = 1, \dots, J. \end{aligned} \tag{3.1}$$

onde θ_{ij} é dado por

$$\theta_{ij} = \Phi(\alpha_i + x_{ij}\beta_j) \quad (\text{Probit})$$

e $\Phi(\cdot)$ é a função de distribuição Normal acumulada, isto é,

$$\Phi(\alpha_i + x_{ij}\beta_j) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha_i + x_{ij}\beta_j} \exp\left\{-\frac{u^2}{2}\right\} du$$

e α_i denota um efeito aleatório na i -ésima observação, especificando a correlação entre $Y_{i1}, Y_{i2}, \dots, Y_{iJ}$. Assumimos inicialmente $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$.

Desta forma temos,

$$\begin{aligned} Pr[Y_{ij} = y_{ij} | \alpha_i, \beta_j, x_{ij}] &= \Phi(\alpha_i + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-y_{ij}} \\ i &= 1, \dots, n; \quad j = 1, \dots, J. \end{aligned}$$

Assumimos também que as observações são independentes, pois a correlação entre os tempos para cada observação é capturado pelo efeito aleatório α_i . Neste caso a função de verossimilhança é dada por,

$$L(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) = \prod_{i=1}^n \prod_{j=1}^J \Phi(\alpha_i + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-y_{ij}} \quad (3.2)$$

onde $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ e \mathcal{D} representa o conjunto de dados (x_{ij}, y_{ij}) , $i = 1, \dots, n$; $j = 1, \dots, J$.

3.1 Análise Clássica do Modelo

Inicialmente, faremos as inferências sobre os parâmetros do modelo a partir do método de verossimilhança. Considerando α_i ($i = 1, \dots, n$) igual à α , ou seja, um parâmetro de efeito fixo, estamos assumindo que não há correlação entre os tempos.

Como nosso objetivo é encontrar estimadores para os parâmetros que maximizem a função de verossimilhança utilizaremos o logaritmo da função de verossimilhança. Para simplificação da notação escrevemos Φ ao invés de $\Phi(\alpha + x_{ij}\beta_j)$. Assim, temos:

$$\begin{aligned} \ell(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) &= \log L(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) \\ &= \sum_{i=1}^n \sum_{j=1}^J \left\{ y_{ij} \log \Phi + (1 - y_{ij}) \log (1 - \Phi) \right\} \end{aligned}$$

As derivadas de primeira ordem do log da função de verossimilhança (ℓ) são dadas por,

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^n \sum_{j=1}^J \left\{ \frac{\Phi'_\alpha (y_{ij} - \Phi)}{\Phi(1 - \Phi)} \right\}$$

$$\frac{\partial \ell}{\partial \beta_{p,j}} = \sum_{i=1}^n \left\{ \frac{\Phi'_{\beta_{p,j}} (y_{ij} - \Phi)}{\Phi(1 - \Phi)} \right\}$$

$$j = 1, 2, \dots, J$$

Fazendo $\frac{\partial \ell}{\partial \alpha} = 0$ e $\frac{\partial \ell}{\partial \beta_{p,j}} = 0$, ($j = 1, 2, \dots, J$) temos as equações de Escore, cujas soluções são os possíveis estimadores de máxima verossimilhança e podem ser encontrados utilizando o método numérico de Newton-Raphson.

As derivadas de segunda ordem para matriz de informação de Fisher, necessária para para o método de Newton-Raphson, são dadas por,

$$\frac{\partial^2 \ell}{\partial \alpha^2} = \sum_{i=1}^n \sum_{j=1}^J \frac{[\Phi'_{\alpha}(y_{ij} - \Phi) - \Phi''_{\alpha}] \Phi(1 - \Phi) - \phi'^2_{\alpha}(1 - 2\Phi)(y_{ij} - \Phi)}{[\Phi(1 - \Phi)]^2},$$

$$\frac{\partial^2 \ell}{\partial \beta_{p,j}^2} = \sum_{i=1}^n \frac{[\Phi'_{\beta_{p,j}}(y_{ij} - \Phi) - \Phi''_{\beta_{p,j}}] \Phi(1 - \Phi) - \phi'^2_{\beta_{p,j}}(1 - 2\Phi)(y_{ij} - \Phi)}{[\Phi(1 - \Phi)]^2},$$

$$j = 1, 2, \dots, J$$

$$\frac{\partial^2 \ell}{\partial \alpha \partial \beta_{p,j}} = \sum_{i=1}^n \frac{[\Phi''_{\alpha \beta_{p,j}}(y_{ij} - \Phi) - \Phi'_{\beta_{p,j}} \Phi'_{\alpha}] \Phi(1 - \Phi) - \Phi'_{\beta_{p,j}} \Phi'_{\alpha}(1 - 2\Phi)(y_{ij} - \Phi)}{[\Phi(1 - \Phi)]^2},$$

$$j = 1, 2, \dots, J$$

$$\frac{\partial^2 \ell}{\partial \beta_{p,j} \partial \beta_{p,k}} = 0 \quad \text{para } j \neq k \quad \text{onde } j = 1, 2, \dots, J; \quad k = 1, 2, \dots, J$$

$$\frac{\partial^2 \ell}{\partial \beta_{p,j} \partial \beta_{p,k}} = 0 \quad \text{para } j \neq k \quad \text{onde } j = 1, 2, \dots, J; \quad k = 1, 2, \dots, J$$

$$\frac{\partial^2 \ell}{\partial \beta_{p,j} \partial \beta_{p,k}} = \sum_{i=1}^n \frac{[\Phi''_{\beta_{p,j} \beta_{p,k}}(y_{ij} - \Phi) - \Phi'_{\beta_{p,k}} \Phi'_{\beta_{p,j}}] \Phi(1 - \Phi) - \Phi'_{\beta_{p,k}} \Phi'_{\beta_{p,j}}(1 - 2\Phi)(y_{ij} - \Phi)}{[\Phi(1 - \Phi)]^2},$$

$$j = k \quad \text{onde } j = 1, 2, \dots, J; \quad k = 1, 2, \dots, J.$$

Definir,

$$\Phi'_{\alpha} = \frac{\partial \Phi(\alpha + x_{ij}\beta_j)}{\partial \alpha} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\alpha + x_{ij}\beta_j)^2}{2}\right\}$$

$$i = 1, \dots, n; \quad j = 1, 2, \dots, J$$

$$\Phi''_{\alpha} = \frac{\partial^2 \Phi(\alpha + x_{ij}\beta_j)}{\partial \alpha^2} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\alpha + x_{ij}\beta_j)^2}{2}\right\} [-(\alpha + x_{ij}\beta_j)]$$

$$i = 1, \dots, n; \quad j = 1, 2, \dots, J$$

$$\Phi'_{\beta_{p_j j}} = \frac{\partial \Phi(\alpha + x_{ij}\beta_j)}{\partial \beta_{p_j j}} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\alpha + x_{ij}\beta_j)^2}{2}\right\} x_{ij p_j}$$

$$i = 1, \dots, n; \quad j = 1, 2, \dots, J$$

(3.3)

$$\Phi''_{\beta_{p_j j}} = \frac{\partial^2 \Phi(\alpha + x_{ij}\beta_j)}{\partial \beta_{p_j j}^2} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\alpha + x_{ij}\beta_j)^2}{2}\right\} [-(\alpha + x_{ij}\beta_j)] x_{ij p_j}^2$$

$$i = 1, \dots, n; \quad j = 1, 2, \dots, J$$

$$\Phi''_{\alpha \beta_{p_j j}} = \frac{\partial^2 \Phi(\alpha + x_{ij}\beta_j)}{\partial \alpha \partial \beta_{p_j j}} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\alpha + x_{ij}\beta_j)^2}{2}\right\} [-(\alpha + x_{ij}\beta_j)] x_{ij p_j}$$

$$i = 1, \dots, n; \quad j = 1, 2, \dots, J$$

$$\Phi''_{\beta_{p_j j} \beta_{p_k k}} = \frac{\partial^2 \Phi(\alpha + x_{ij}\beta_j)}{\partial \beta_{p_j j} \partial \beta_{p_k k}} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(\alpha + x_{ij}\beta_j)^2}{2}\right\} [-(\alpha + x_{ij}\beta_j)] x_{ij p_j} x_{ij p_k}$$

$$i = 1, \dots, n; \quad j = k; \quad j, k = 1, 2, \dots, J$$

3.2 Análise Bayesiana do Modelo

Modelo 4: Modelo probito com efeito fixo.

Assumir os efeitos fixos, α_i ($i = 1, \dots, n$) iguais à α , com as seguintes distribuições

à priori:

- (i) $\alpha \sim N(\mu_0, \sigma_0^2)$ com μ_0 e σ_0^2 conhecidos;
- (ii) $\beta_{kj} \sim N(b_{kj}, c_{kj}^2)$ com b_{kj} e c_{kj}^2 conhecidos;
- $k = 1, 2, \dots, p_j; \quad j = 1, 2, \dots, J.$
- (3.4)

Assumindo independência das distribuições a priori, temos a distribuição à priori conjunta dada por,

$$\pi(\alpha, \beta_1, \beta_2, \dots, \beta_J) \propto \exp\left\{-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right\} \prod_{j=1}^J \prod_{k=1}^{p_j} \exp\left\{-\frac{(\beta_{kj} - \mu_{kj})^2}{2\sigma_{kj}^2}\right\}. \quad (3.5)$$

Combinando (3.2) com (3.5), temos a seguinte distribuição à posteriori conjunta:

$$\begin{aligned} \pi(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) \propto & \exp\left[-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right] \left\{ \prod_{j=1}^J \prod_{k=1}^{p_j} \exp\left[-\frac{(\beta_{kj} - \mu_{kj})^2}{2\sigma_{kj}^2}\right] \right\} \\ & \cdot \left\{ \prod_{i=1}^n \prod_{j=1}^J \Phi(\alpha + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha + x_{ij}\beta_j)]^{1-y_{ij}} \right\} \end{aligned}$$

As distribuições condicionais para o algoritmo de Metropolis-Hastings são dadas por:

(i)

$$\pi(\alpha | \beta_1, \beta_2, \dots, \beta_J, \mathcal{D}) \propto \exp\left[-\frac{(\alpha - \mu_0)^2}{2\sigma_0^2}\right] \prod_{i=1}^n \prod_{j=1}^J \Phi(\alpha + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha + x_{ij}\beta_j)]^{1-y_{ij}}$$

$$\pi(\alpha | \beta_1, \beta_2, \dots, \beta_J, \mathcal{D}) \propto N(\mu_0, \sigma_0^2) L(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D})$$

(ii)

$$\pi(\beta_{kl} | \alpha, \beta_1, \dots, \beta_{j-1}, \underset{\sim_{kl}}{\beta}, \beta_{j+1}, \dots, \beta_J, \mathcal{D}) \propto \exp \left[-\frac{(\beta_{kl} - \mu_{kl})^2}{2\sigma_{kl}^2} \right] \cdot \left\{ \prod_{i=1}^n \Phi(\alpha + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha + x_{ij}\beta_j)]^{1-y_{ij}} \right\}$$

$$\pi(\beta_{kl} | \alpha, \beta_1, \dots, \beta_{j-1}, \underset{\sim_{kl}}{\beta}, \beta_{j+1}, \dots, \beta_J, \mathcal{D}) \propto N(b_{kl}, c_{kl}^2) \cdot \Psi_1(\alpha, \beta_1, \beta_2, \dots, \beta_J)$$

onde

$$\Psi_1(\alpha, \beta_1, \beta_2, \dots, \beta_J) = \left\{ \prod_{i=1}^n \Phi(\alpha + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha + x_{ij}\beta_j)]^{1-y_{ij}} \right\}$$

$$k = 1, \dots, p_j; \quad l, j = 1, \dots, J; \quad \text{e } \underset{\sim_{kl}}{\beta} \text{ definido em (2.12)}$$

3.3 Análise Bayesiana Assumindo uma Distribuição Normal para os efeitos aleatórios α_i

Modelo 5: Modelo probito assumindo efeitos aleatórios normais - $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$;
 $i = 1, \dots, n$

Como no Capítulo 2, assumir que α_i , $i = 1, \dots, n$ sejam efeitos aleatórios independentes com distribuição normal, $N(0, \sigma_\alpha^2)$ com as seguintes distribuições a priori,

- (i) $\sigma_\alpha^2 \sim \mathcal{IG}(a, b)$ com a e b conhecidos;
(ii) $\beta_{lj} \sim N(b_{lj}, c_{lj}^2)$ com b_{lj} e c_{lj}^2 conhecidos; $l = 1, \dots, p_j$; $j = 1, \dots, J$, (3.6)

onde $\mathcal{IG}(a, b)$ denota a distribuição gama invertida com média $\frac{1}{b(a-1)}$ para $a > 1$ e variância $\frac{1}{b^2(a-1)^2(a-2)}$ para $b > 1$.

Assumindo independência das distribuições a priori e considerando a função de verossimilhança (3.2), a distribuição a posteriori conjunta para $\underline{\alpha} = (\alpha_1, \dots, \alpha_n)$, σ_α^2 e

β_1, \dots, β_J é dada por,

$$\begin{aligned} \pi(\alpha, \sigma_\alpha^2, \beta_1, \dots, \beta_J | \mathcal{D}) &\propto \\ &\propto (\sigma_\alpha^2)^{-\left(\frac{n}{2}+a+1\right)} \exp\left\{-\frac{1}{2\sigma_\alpha^2}\left[\frac{2}{b} + \sum_{i=1}^n \alpha_i^2\right]\right\} \\ &\cdot \left[\prod_{k=1}^{p_j} \prod_{j=1}^J \exp\left\{-\frac{1}{2c_{kj}^2}(\beta_{kj} - b_{kj})^2\right\}\right] \\ &\cdot \left[\prod_{i=1}^n \prod_{j=1}^J \Phi(\alpha_i + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-y_{ij}}\right]. \end{aligned} \quad (3.7)$$

Portanto as distribuições condicionais para o algoritmo Metropolis-Hastings, são dadas por,

i)

$$\pi(\alpha_i | \sigma_\alpha^2, \beta_1, \dots, \beta_J, \mathcal{D}) \propto \exp\left\{-\frac{\alpha_i^2}{2\sigma_\alpha^2}\right\} \prod_{j=1}^J \Phi(\alpha_i + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-y_{ij}}$$

para $i = 1, \dots, n$, isto é,

$$\pi(\alpha_i | \sigma_\alpha^2, \beta_1, \dots, \beta_J, \mathcal{D}) \propto N\{0, \sigma_\alpha^2\} \Psi_2(\alpha_i, \beta_1, \dots, \beta_J)$$

onde

$$\Psi_2(\alpha_i, \beta_1, \dots, \beta_J) = \prod_{j=1}^J \Phi(\alpha_i + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-y_{ij}}$$

ii)

$$\begin{aligned} \pi(\sigma_\alpha^2 | \alpha_i, \beta_1, \dots, \beta_J, \mathcal{D}) &\propto (\sigma_\alpha^2)^{-\left(\frac{n}{2}+a+1\right)} \exp\left\{-\frac{1}{\sigma_\alpha^2}\left[\frac{1}{b} + \frac{1}{2} \sum_{i=1}^n \alpha_i^2\right]\right\} \\ &\propto (\sigma_\alpha^2)^{-\left(\frac{n}{2}+a+1\right)} \exp\left\{-\frac{1}{\sigma_\alpha^2\left[\frac{1}{b} + \frac{1}{2} \sum_{i=1}^n \alpha_i^2\right]^{-1}}\right\} \end{aligned}$$

isto é,

$$\sigma_\alpha^2 | \alpha_i, \beta_1, \dots, \beta_J, \mathcal{D} \sim IG\left[\frac{n}{2} + a; \left(\frac{1}{b} + \frac{1}{2} \sum_{i=1}^n \alpha_i^2\right)^{-1}\right]$$

iii)

$$\pi(\beta_{kl} | \alpha_i, \beta_1, \dots, \beta_{j-1}, \underset{\sim_{kl}}{\beta}, \beta_{j+1}, \dots, \beta_J, \mathcal{D}) \propto N(b_{kl}, c_{kl}^2) \cdot \Psi_3(\alpha_i, \beta_1, \beta_2, \dots, \beta_J)$$

onde

$$\Psi_3(\alpha_i, \beta_1, \beta_2, \dots, \beta_J) = \prod_{i=1}^n \Phi(\alpha_i + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-y_{ij}}$$

$$k = 1, \dots, p_j; \quad l, j = 1, \dots, J; \quad \text{e } \underset{\sim_{kl}}{\beta} \text{ definido em (2.12).}$$

Observamos que a amostra para σ_α^2 é gerada diretamente pelo algoritmo Gibbs sampling.

3.4 Análise Bayesiana Assumindo uma Mistura de Normais para α_i

Modelo 6: Modelo probito assumindo os efeitos aleatórios com uma mistura de distribuições Normais.

Consideramos nesta seção, uma mistura de duas distribuições normais, para os efeitos aleatórios de α_i ($i = 1, \dots, n$), isto é,

$$\pi(\alpha_i) = \sum_{k=1}^2 p_k \Phi_k(\alpha_i | \mu_k, \sigma_k^2) \quad (3.8)$$

onde $\sum_{k=1}^2 p_k = 1$ e Φ_k denota a densidade normal $N(\mu_k, \sigma_k^2)$.

Assumir independência à priori entre os parâmetros e considerar as seguintes densidades à priori para μ_k, σ_k^2, p_1 e β_{lj} , $k = 1, 2$; $l = 1, \dots, p_j$; $j = 1, \dots, J$:

- (i) $\mu_k \sim N(d_k, e_k^2)$ com d_k e e_k conhecidos, $k = 1, 2$;
- (ii) $\sigma_k^2 \sim \mathcal{IG}(a_k, b_k)$ com a_k e b_k conhecidos, $k = 1, 2$;
- (iii) $\beta_{lj} \sim N(b_{lj}, c_{lj}^2)$ com b_{lj} e c_{lj}^2 conhecidos, $l = 1, \dots, p_j$; $j = 1, \dots, J$
- (iv) $p_1 \sim \mathcal{B}(f, g)$ com f e g conhecidos

(3.9)

onde $N(d_k, e_k)$ denota a distribuição normal com média d_k e variância e_k ; $\mathcal{IG}(a_k, b_k)$ uma distribuição gama invertida com média $\frac{1}{b_k(a_k-1)}$ para $a_k > 1$ e variância $\frac{1}{b_k^2[(a_k-1)^2(a_k-2)]}$ para $a_k > 2$ e $B(f, g)$ denota uma distribuição beta com média $\frac{f}{(f+g)}$ e variância $\frac{fg}{(f+g)^2(f+g+1)}$.

A densidade à posteriori conjunta para $\Theta = (\underline{\alpha}, p_1, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \beta_1, \dots, \beta_J)$ é dada por,

$$\pi(\Theta|\mathcal{D}) \propto \Psi(\Theta) \left\{ \prod_{i=1}^n \sum_{k=1}^2 p_k \Phi_k(\alpha_i | \mu_k, \sigma_k^2) \right\} \quad (3.10)$$

onde

$$\Psi(\Theta) = \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(a_k+1)} \exp\left(-\frac{b_k}{\sigma_k^2}\right) \right\} \cdot \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2e_k^2}(\mu_k - d_k)^2\right] \right\} \cdot$$

$$p_1^{f-1} (1-p_1)^{g-1} \cdot \left\{ \prod_{l=1}^{p_j} \prod_{j=1}^J \exp\left[-\frac{1}{2c_{lj}^2}(\beta_{lj} - b_{lj})^2\right] \right\} \cdot$$

$$\prod_{i=1}^n \prod_{j=1}^J \Phi(\alpha_i + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-y_{ij}}.$$

Considerando a introdução da variável artificial $\underline{z}_i = (z_{i1}, z_{i2})$, $i = 1, \dots, n$, $\underline{z} = (z_1, \dots, z_n)$ (dada em (2.21)), a densidade à posteriori conjunta para Θ é dada por:

$$\begin{aligned} \pi(\Theta|\mathcal{D}, \underline{z}) &\propto \prod_{i=1}^n \prod_{k=1}^2 \left[p_k \Phi_k(\alpha_i | \mu_k, \sigma_k^2) \right]^{z_{ik}} \cdot \\ &\cdot \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(a_k+1)} \exp\left[\frac{-1}{b_k \sigma_k^2}\right] \right\} \cdot p_1^{f-1} (1-p_1)^{g-1} \cdot \end{aligned} \quad (3.11)$$

$$\cdot \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2e_k^2}(\mu_k - d_k)^2\right] \right\} \cdot \left\{ \prod_{l=1}^{p_j} \prod_{j=1}^J \exp\left[-\frac{1}{2c_{lj}^2}(\beta_{lj} - b_{lj})^2\right] \right\} \cdot$$

$$\cdot \prod_{i=1}^n \prod_{j=1}^J \Phi(\alpha_i + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-y_{ij}}$$

As distribuições condicionais para gerar amostras dos parâmetros através do algoritmo Metropolis-Hastings são dadas por:

(i)

$$\pi(\alpha_i | \beta_1, \dots, \beta_J, \underline{\mu}, \underline{\sigma}^2, p_1, \underline{z}, \mathcal{D}) \propto N \left\{ \frac{Z_{i1}\mu_1\sigma_2^2 + Z_{i2}\mu_2\sigma_1^2}{Z_{i1}\sigma_2^2 + Z_{i2}\sigma_1^2}, \frac{\sigma_1^2\sigma_2^2}{Z_{i1}\sigma_2^2 + Z_{i2}\sigma_1^2} \right\} \Psi_4(\alpha_i, \beta_1, \dots, \beta_J),$$

onde

$$\Psi_4(\alpha_i, \beta_1, \dots, \beta_J) = \prod_{j=1}^J \Phi(\alpha_i + x_{ij}\beta_j)^{v_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-v_{ij}}, \quad (3.12)$$

$$\underline{\mu} = (\mu_1, \mu_2); \quad \underline{\sigma} = (\sigma_1^2, \sigma_2^2); \quad i = 1, 2, \dots, n;$$

(ii)

$$\pi(\sigma_k^2 | \alpha_i, \beta_1, \dots, \beta_J, \underline{\mu}, \sigma_{\underline{k}}^2, p_1, \underline{z}, \mathcal{D}) \propto (\sigma_k^2)^{-(a_k+1)} \exp \left\{ -\frac{1}{b_k \sigma_k^2} \right\} (\sigma_k^2)^{-\frac{v_k}{2}} \\ \cdot \exp \left\{ -\frac{1}{2\sigma_k^2} \sum_{i=1}^n z_{ik} (\alpha_i - \mu_k)^2 \right\},$$

$$\text{onde } v_k = \sum_{i=1}^n z_{ik}$$

Isto é,

$$\sigma_k^2 | \alpha_i, \beta_1, \dots, \beta_J, \underline{\mu}, p_1, \underline{z}, \mathcal{D} \sim \text{IG} \left[a_k + \frac{v_k}{2}; \left(\frac{1}{b_k} + \frac{\sum_{i=1}^n z_{ik} (\alpha_i - \mu_k)^2}{2} \right)^{-1} \right]$$

$$\text{onde } k \neq \underline{k}, \quad k, \underline{k} = 1, 2;$$

(iii)

$$p_1 | \underline{\alpha}, \beta_1, \dots, \beta_J, \underline{\mu}, \underline{\sigma}^2, \underline{z}, \mathcal{D} \sim B(f + v_1, g + v_2);$$

(iv)

$$\pi(\beta_{kl} | \underline{\alpha}, \beta_1, \dots, \beta_{j-1}, \underline{\beta}_{\underline{kl}}, \beta_{j+1}, \dots, \beta_J, \underline{\mu}, \underline{\sigma}^2, p_1, \underline{z}, \mathcal{D}) \propto N(b_{kl}, c_{kl}^2) \Psi_5(\underline{\alpha}, \beta_1, \dots, \beta_J),$$

onde

$$\Psi_5(\underline{\alpha}, \beta_1, \dots, \beta_J) = \prod_{i=1}^n \Phi(\alpha_i + x_{ij}\beta_j)^{v_{ij}} [1 - \Phi(\alpha_i + x_{ij}\beta_j)]^{1-v_{ij}},$$

 $\underline{\beta}$ definido em (2.12)

$$\underline{\beta}_{\underline{kl}} \\ k = 1, \dots, p_j; \quad l, j = 1, \dots, J;$$

(v)

$$\pi(\mu_k | \underline{\alpha}, \beta_1, \dots, \beta_J, \mu_{\underline{k}}, \underline{\sigma}^2, p_1, \mathcal{D}, \underline{z}) \sim N \left\{ \frac{d_k \sigma_k^2 + e_k^2 \sum_{i=1}^n \alpha_i Z_{ik}}{\sigma_k^2 + e_k^2 \sum_{i=1}^n Z_{ik}}, \frac{e_k^2 \sigma_k^2}{\sigma_k^2 + e_k^2 \sum_{i=1}^n Z_{ik}} \right\},$$

onde $k \neq \underline{k}$, $k, \underline{k} = 1, 2$.

Para gerar as amostras usamos o mesmo procedimento descrito no cap. 2 seção 5.

Capítulo 4

DADOS BINÁRIOS CORRELACIONADOS COM RÉPLICAS

Muitas vezes em um banco de dados de muitas unidades experimentais podemos verificar que alguns resultados repetidos ocorrem com uma dada frequência. Neste capítulo trataremos desse caso utilizando a mesma modelagem anterior incluindo a frequência dessas ocorrências nos modelos.

Consideramos neste capítulo, um dado número de observações N , dentre as quais podemos ter repetições dos dados analisados (réplicas). As N observações podem ser escritas como n grupos considerando suas respectivas frequências, i.e., $N = \sum_{i=1}^n f_i$.

Assim Y_{ij} (1 ou 0), a resposta para a i -ésima observação no j -ésimo tempo, segue distribuição de Bernoulli com parâmetro θ_{ij} , ou seja,

$$\begin{aligned} Pr[Y_{ij} = y_{ij} | \alpha_i, \beta_j, x_{ij}] &= \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \\ i &= 1, \dots, N; \quad j = 1, \dots, J. \end{aligned} \tag{4.1}$$

Assim temos duas possibilidades para modelar θ_{ij} :

- a) $\theta_{ij} = \frac{\exp(\alpha_i + x_{ij}\beta_j)}{1 + \exp(\alpha_i + x_{ij}\beta_j)}$ (Logístico)
- b) $\theta_{ij} = \Phi(\alpha_i + x_{ij}\beta_j)$ (Probit)

Assumindo que as observações são independentes, a função de verossimilhança é

dada por,

$$L(\underline{\alpha}, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) = \prod_{i=1}^N \prod_{j=1}^J \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \quad (4.2)$$

onde $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_N)$ e \mathcal{D} representa os dados (y_{ij}, x_{ij}) .

Como podemos escrever as N observações em n grupos de frequências f_i , $i = 1, \dots, n$, a função de verossimilhança pode ser escrita na forma,

$$L(\underline{\alpha}, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) = \prod_{i=1}^n \prod_{j=1}^J \left[\theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \right]^{f_i} \quad (4.3)$$

onde $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$

4.1 Estimadores de Máxima Verossimilhança

Considerando α_i ($i = 1, 2, \dots, n$) iguais a α , ou seja, efeitos fixos, e o modelo logístico, o logaritmo da função de verossimilhança é dada por,

$$\begin{aligned} \ell(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) &= \log L(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) \\ &= \sum_{i=1}^n \sum_{j=1}^J \alpha y_{ij} f_i + \sum_{i=1}^n \sum_{j=1}^J x_{ij} \beta_j y_{ij} f_i - \sum_{i=1}^n \sum_{j=1}^J f_i \log[1 + \exp(\alpha + x_{ij} \beta_j)]. \end{aligned} \quad (4.4)$$

Derivando o logaritmo da função verossimilhança em relação à α e β e igualando à 0 (zero), temos as seguintes equações *scores*:

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^J y_{ij} f_i - \sum_{i=1}^n \sum_{j=1}^J f_i \frac{\exp\{(\hat{\alpha} + x_{ij} \hat{\beta}_j)\}}{1 + \exp\{\hat{\alpha} + x_{ij} \hat{\beta}_j\}} &= 0 \\ \sum_{i=1}^n x_{ij} y_{ij} f_i - \sum_{i=1}^n x_{ij} f_i \frac{\exp\{(\hat{\alpha} + x_{ij} \hat{\beta}_j)\}}{1 + \exp\{\hat{\alpha} + x_{ij} \hat{\beta}_j\}} &= 0 \end{aligned} \quad (4.5)$$

$$j = 1, 2, \dots, J$$

Para o caso do modelo probito ($\theta_{ij} = \Phi(\alpha + x_{ij}\beta_j) = \Phi$) o logaritmo da função de verossimilhança é dado por,

$$\begin{aligned} \ell(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) &= \log L(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) \\ &= \sum_{i=1}^n \sum_{j=1}^J f_i \left\{ y_{ij} \log \Phi + (1 - y_{ij}) \log (1 - \Phi) \right\} \end{aligned}$$

As equações *scores* são dadas por,

$$\sum_{i=1}^n \sum_{j=1}^J f_i \left\{ \frac{\Phi'_\alpha (y_{ij} - \Phi)}{\Phi(1 - \Phi)} \right\} = 0 \quad (4.6)$$

$$\sum_{i=1}^n f_i \left\{ \frac{\Phi'_{\beta_{pj}} (y_{ij} - \Phi)}{\Phi(1 - \Phi)} \right\} = 0$$

$$j = 1, 2, \dots, J$$

com Φ'_α e $\Phi'_{\beta_{pj}}$ apresentadas em (3.3).

4.2 Análise Bayesiana Assumindo $\alpha_i = \alpha$

Para uma análise Bayesiana considerando $\alpha_i = \alpha$, assumir as seguintes densidades a priori informativas:

- (i) $\alpha \sim N(\mu_0, \sigma_0^2)$ com μ_0 e σ_0^2 conhecidos;
 - (ii) $\beta_{kj} \sim N(\mu_{kj}, \sigma_{kj}^2)$ com μ_{kj} e σ_{kj}^2 conhecidos; (4.7)
- $$j = 1, 2, \dots, J \quad k = 1, 2, \dots, p_j.$$

Com independência das distribuições a priori obtemos a seguinte densidade a posteriori conjunta para os parâmetros:

$$\begin{aligned} \pi(\alpha, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) &\propto \left[\exp \left\{ -\frac{(\alpha - \mu_0)^2}{2\sigma_0^2} \right\} \prod_{j=1}^J \prod_{k=1}^{p_j} \exp \left\{ -\frac{(\beta_{kj} - \mu_{kj})^2}{2\sigma_{kj}^2} \right\} \right] \\ &\cdot \left\{ \prod_{i=1}^n \prod_{j=1}^J \left[\theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \right]^{f_i} \right\} \end{aligned} \quad (4.8)$$

Com a ligação logística, as distribuições condicionais para o algoritmo de Gibbs sampling são dadas por,

(i)

$$\pi(\alpha|\beta_1, \beta_2, \dots, \beta_J, \mathcal{D}) \propto N(\mu_0, \sigma_0^2) \cdot \Psi_1(\alpha, \beta_1, \beta_2, \dots, \beta_J)$$

onde

$$\Psi_1(\alpha, \beta_1, \beta_2, \dots, \beta_J) = \exp\left\{\alpha \sum_{i=1}^n \sum_{j=1}^J f_i y_{ij} - \sum_{i=1}^n \sum_{j=1}^J f_i \log[1 + \exp(\alpha + x_{ij}\beta_j)]\right\}$$

(ii)

$$\pi(\beta_{kl}|\alpha, \beta_1, \dots, \beta_{j-1}, \underset{\sim kl}{\beta}, \beta_{j+1}, \dots, \beta_J, \mathcal{D}) \propto$$

$$\propto N(\mu_{kl}, \sigma_{kl}^2) \cdot \Psi_2(\alpha, \beta_1, \beta_2, \dots, \beta_J, \mathcal{D})$$

onde

$$\Psi_2(\alpha, \beta_1, \beta_2, \dots, \beta_J, \mathcal{D}) = \exp\left\{a_{kl}\beta_{kl} - \sum_{i=1}^n f_i \log[1 + \exp(\alpha + x_{ij}\beta_j)]\right\};$$

$$a_{kl} = \sum_{i=1}^n x_{ik} y_{il} f_i; \quad k = 1, \dots, p_j; \quad l, j = 1, \dots, J$$

e $\beta_{\sim kl}$ definido em (2.12).

Com a ligação proibito, temos,

$$\Psi_1(\alpha, \beta_1, \beta_2, \dots, \beta_J) = \prod_{i=1}^n \prod_{j=1}^J \left[\Phi(\alpha + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha + x_{ij}\beta_j)]^{1-y_{ij}} \right]^{f_i},$$

e

$$\Psi_2(\alpha, \beta_1, \beta_2, \dots, \beta_J) = \prod_{i=1}^n \left[\Phi(\alpha + x_{ij}\beta_j)^{y_{ij}} [1 - \Phi(\alpha + x_{ij}\beta_j)]^{1-y_{ij}} \right]^{f_i},$$

$$j = 1, 2, \dots, J.$$

4.3 Análise Bayesiana Assumindo $\alpha_i \sim N(0, \sigma_\alpha^2)$

Para esta seção assumimos efeitos aleatórios para α_i ($i = 1, \dots, n$), com distribuição normal $N(0, \sigma_\alpha^2)$ e consideramos as seguintes distribuições a priori,

(i) $\sigma_\alpha^2 \sim \mathcal{IG}(a, b)$ com a e b conhecidos;(ii) $\beta_{lj} \sim N(b_{lj}, c_{lj}^2)$ com b_{lj} e c_{lj}^2 conhecidos; (4.9)

$$l = 1, \dots, p_j; \quad j = 1, \dots, J.$$

Consideramos independência das distribuições a priori, e com (4.3) as distribuições condicionais para o algoritmo Gibbs sampling, considerando a ligação logística, são dadas por,

i)

$$\pi(\alpha_i | \sigma_\alpha^2, \beta_1, \dots, \beta_J, \mathcal{D}) \propto N\{0, \sigma_\alpha^2\} \Psi_3(\alpha_i, \beta_1, \dots, \beta_J)$$

onde

$$\Psi_3(\alpha_i, \beta_1, \dots, \beta_J) = \exp \left\{ \alpha_i f_i \sum_{j=1}^J y_{ij} - \sum_{j=1}^J f_i \log[1 + \exp(\alpha_i + x_{ij} \beta_j)] \right\}$$

ii)

$$\sigma_\alpha^2 | \alpha, \beta_1, \dots, \beta_J, \mathcal{D} \sim \text{IG} \left[\frac{n}{2} + a; \left(\frac{1}{b} + \frac{1}{2} \sum_{i=1}^n \alpha_i^2 \right)^{-1} \right]$$

iii)

$$\pi(\beta_{kl} | \alpha, \beta_1, \dots, \beta_{j-1}, \beta_{\sim kl}, \beta_{j+1}, \dots, \beta_J, \mathcal{D}) \propto N(b_{kl}, c_{kl}^2) \Psi_4(\alpha, \beta_1, \dots, \beta_J),$$

onde

$$\Psi_4(\alpha, \beta_1, \beta_2, \dots, \beta_J, \mathcal{D}) = \exp \left\{ a_{kl} \beta_{kl} - \sum_{i=1}^n f_i \log[1 + \exp(\alpha + x_{ij} \beta_j)] \right\};$$

$$a_{kl} = \sum_{i=1}^n x_{ilk} y_{il} f_i; \quad k = 1, \dots, p_j; \quad l, j = 1, \dots, J$$

e $\beta_{\sim kl}$ definido em (2.12).

Considerando a ligação proibito, as distribuições condicionais para o algoritmo Gibbs sampling serão as mesmas da ligação logística com,

$$\Psi_3(\alpha_i, \beta_1, \beta_2, \dots, \beta_J) = \prod_{j=1}^J \left[\Phi(\alpha + x_{ij} \beta_j)^{y_{ij}} [1 - \Phi(\alpha + x_{ij} \beta_j)]^{1-y_{ij}} \right]^{f_i},$$

$$i = 1, 2, \dots, n,$$

e

$$\Psi_4(\alpha, \beta_1, \beta_2, \dots, \beta_J) = \prod_{i=1}^n \left[\Phi(\alpha + x_{ij} \beta_j)^{y_{ij}} [1 - \Phi(\alpha + x_{ij} \beta_j)]^{1-y_{ij}} \right]^{f_i},$$

$$j = 1, 2, \dots, J.$$

4.4 Análise Bayesiana Assumindo uma Mistura de Normais para α_i

Novamente assumimos um modelo de regressão de efeitos aleatórios para dados binários correlacionados onde α_i ($i = 1, 2, \dots, n$) é uma mistura de duas distribuições normais, dada por,

$$\begin{aligned}\pi(\alpha_i) &= \sum_{k=1}^2 p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2) \\ &= p_1 \Phi_1(\alpha_i | \mu_1, \sigma_1^2) + (1 - p_1) \Phi_2(\alpha_i | \mu_2, \sigma_2^2)\end{aligned}\tag{4.10}$$

com a restrição $p_1 + p_2 = 1$; ϕ_k denota uma densidade normal.

Consideramos as seguintes densidades a priori para os parâmetros, $\mu_k, \sigma_k^2, p_1, \beta_{lj}$, $k = 1, 2; l = 1, \dots, p_j; j = 1, \dots, J$:

- (i) $\mu_k \sim N(d_k, e_k^2)$ com d_k e e_k conhecidos, $k = 1, 2$;
- (ii) $\sigma_k^2 \sim \mathcal{IG}(a_k, b_k)$ com a_k e b_k ; conhecidos, $k = 1, 2$;
- (iii) $\beta_{lj} \sim N(b_{lj}, c_{lj}^2)$ com b_{lj} e c_{lj} ; conhecidos,
 $l = 1, \dots, p_j; j = 1, \dots, J$,
- (iv) $p_1 \sim B(f, g)$ com f e g ; conhecidos;

(4.11)

A distribuição a posteriori conjunta para $\Theta = (\alpha_1, \dots, \alpha_n, \mu_1, \mu_2, p_1, \sigma_1^2, \sigma_2^2, \beta_1, \dots, \beta_J)$ é dada por

$$\pi(\Theta | \mathcal{D}) \propto \Psi(\Theta) \left\{ \prod_{i=1}^n \sum_{k=1}^2 p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2) \right\}\tag{4.12}$$

onde

$$\begin{aligned} \Psi(\Theta) &= \left\{ \prod_{k=1}^2 (\sigma_k^2)^{-(a_k+1)} \exp\left(-\frac{1}{b_k \sigma_k^2}\right) \right\} \cdot \left\{ \prod_{k=1}^2 \exp\left[-\frac{1}{2e_k^2} (\mu_k - d_k)^2\right] \right\} \cdot \\ & p_1^{f-1} (1-p_1)^{g-1} \cdot \left\{ \prod_{l=1}^{p_j} \prod_{j=1}^J \exp\left[-\frac{1}{2c_{lj}^2} (\beta_{lj} - b_{lj})^2\right] \right\} \cdot \\ & \prod_{i=1}^n \prod_{j=1}^J \left[p_{ij}^{y_{ij}} (1-p_{ij})^{1-y_{ij}} \right]^{f_i} \end{aligned}$$

Como anteriormente para simplificar as distribuições condicionais necessárias para o algoritmo de Gibbs sampling, introduzimos uma variável latente, $\tilde{z}_i = (z_{i1}, z_{i2})$, ($z_{i1} + z_{i2} = 1$), $i = 1, \dots, n$, com distribuição conjunta

$$\pi(\tilde{z}_1, \dots, \tilde{z}_n) \propto \frac{\prod_{i=1}^n \prod_{k=1}^2 [p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2)]^{z_{ik}}}{\prod_{i=1}^n \{\sum_{k=1}^2 p_k \Phi_k(\alpha_i | \mu_k, \sigma_k^2)\}} \quad (4.13)$$

Combinando a equação (4.12) com (4.13), obtemos,

$$\pi(\Theta, \tilde{z}_1, \dots, \tilde{z}_n | \mathcal{D}) \propto \Psi(\Theta) \left\{ \prod_{i=1}^n \prod_{k=1}^2 [p_k \phi_k(\alpha_i | \mu_k, \sigma_k^2)]^{z_{ik}} \right\} \quad (4.14)$$

Considerando a ligação logística $\theta_{ij} = \frac{\exp(\alpha_i + x_{ij}\beta_j)}{1 + \exp(\alpha_i + x_{ij}\beta_j)}$ as distribuições condicionais para o algoritmo Gibbs sampling, são dadas por,

(i)

$$\pi(\alpha_i | \beta_1, \dots, \beta_J, \mu, \tilde{\sigma}^2, p_1, \tilde{z}, \mathcal{D}) \propto N \left\{ \frac{z_{i1}\mu_1\sigma_2^2 + z_{i2}\mu_2\sigma_1^2}{z_{i1}\sigma_2^2 + z_{i2}\sigma_1^2}, \frac{\sigma_1^2\sigma_2^2}{z_{i1}\sigma_2^2 + z_{i2}\sigma_1^2} \right\}$$

$$\cdot \Psi_5(\alpha_i, \beta_1, \dots, \beta_J),$$

onde,

$$\Psi_5(\alpha_i, \beta_1, \dots, \beta_J) = \exp \left\{ \alpha_i y_i f_i - \sum_{j=1}^J f_i \log \left[1 + \exp(\alpha_i + x_{ij}\beta_j) \right] \right\}$$

para $i = 1, \dots, n$.

(ii)

$$\sigma_k^2 | \alpha_i, \beta_1, \dots, \beta_J, \mu, \sigma_{\underline{k}}^2, p_1, z, \mathcal{D} \sim \mathcal{IG} \left[a_k + \frac{v_k}{2}; \left(\frac{1}{b_k} + \frac{\sum_{i=1}^n z_{ik} (\alpha_i - \mu_k)^2}{2} \right)^{-1} \right]$$

onde $v_k = \sum_{i=1}^n z_{ik}$; $k \neq \underline{k}$, $k, \underline{k} = 1, 2$;

(iii)

$$p_1 | \alpha, \beta_1, \dots, \beta_J, \mu, \sigma^2, z, \mathcal{D} \sim \mathcal{B}(f + v_1, g + v_2);$$

(iv)

$$\pi(\beta_{kl} | \alpha, \beta_1, \dots, \beta_{j-1}, \beta_{\underline{kl}}, \beta_{j+1}, \dots, \beta_J, \mu, \sigma^2, p_1, z, \mathcal{D}) \propto N(b_{kl}, c_{kl}^2) \Psi_6(\alpha, \beta_1, \dots, \beta_J),$$

onde

$$\Psi_6(\alpha, \beta_1, \dots, \beta_J) = \exp \left\{ a_{kl} \beta_{kl} - \sum_{i=1}^n f_i \log[1 + \exp(\alpha + x_{ij} \beta_j)] \right\};$$

$$a_{kl} = \sum_{i=1}^n x_{ik} y_{il} f_i; \quad k = 1, \dots, p_j; \quad l, j = 1, \dots, J$$

e $\beta_{\underline{kl}}$ definido em (2.12).

(v)

$$\pi(\mu_k | \alpha, \beta_1, \dots, \beta_J, \mu_{\underline{k}}, \sigma^2, p_1, \mathcal{D}, z) \sim N \left\{ \frac{d_k \sigma_k^2 + e_k^2 \sum_{i=1}^n \alpha_i Z_{ik}}{\sigma_k^2 + e_k^2 \sum_{i=1}^n Z_{ik}}, \frac{e_k^2 \sigma_k^2}{\sigma_k^2 + e_k^2 \sum_{i=1}^n Z_{ik}} \right\},$$

onde $k \neq \underline{k}$, $k, \underline{k} = 1, 2$.

Capítulo 5

SELEÇÃO DE MODELOS

Nos capítulos anteriores apresentamos alguns modelos para modelagem de dados binários correlacionados, e usualmente temos que decidir qual se adequa melhor aos dados.

Para selecionar um modelo, descrevemos dois métodos para discriminação dos mesmos.

O primeiro método baseia-se no erro de previsão do modelo, onde observamos a distância entre os dados observados e dados esperados através do modelo, onde de uma forma empírica, quanto menor for a distância, melhor será o modelo. Para isso fazemos uso dos resíduos de Pearson (ver por exemplo, Dey & Chen, 1996).

No segundo método a avaliação do modelo passa pela verossimilhança, e como ela provém da distribuição dos dados baseada no modelo é chamada de verossimilhança preditiva ou marginal. Este método considera a razão das verossimilhanças preditivas, que é chamado de fator de Bayes, fornecendo uma medida relativa de avaliação do modelo.

5.1 Resíduos de Pearson

Na estatística clássica é empregado o teste *goodness of fit*, para verificar a plausibilidade do modelo ajustado aos dados. O teste clássico *goodness of fit* quantifica a extremidade de uma medida de discrepância particular calculando uma probabilidade da área sob as extremidades da curva do modelo sob específica hipótese de nulidade.

Como alternativa à estatística clássica *goodness of fit*, Dey e Chen (1996), desenvolvem um método Bayesiano de checagem de modelos para dados binários correlacionados.

Para incorporar a natural correlação das respostas binárias, eles introduzem um resíduo de Pearson multivariado para medir a discrepância entre os dados e o modelo, que é definido como

$$(d_{i1}(\theta), d_{i2}(\theta), \dots, d_{iJ}(\theta))' = (\Sigma_i^*(\theta))^{-1/2}(y_i - p_i(\theta)) \quad (5.1)$$

onde $p_i(\theta) = (p_{i1}(\theta), p_{i2}(\theta), \dots, p_{iJ}(\theta))'$ e $\Sigma_i^*(\theta)$ é a matriz de covariância de Y_i .

No caso de interesse nas observações individuais, por exemplo, identificar se determinada observação é um possível *outlier*, o resíduo de Pearson a nível de observação é definido por:

$$d_{ij} = \frac{(y_{ij} - p_{ij})}{\sqrt{p_{ij}(1 - p_{ij})}} \quad i = 1, 2, \dots, n; j = 1, 2, \dots, J. \quad (5.2)$$

Além disso para verificar o desempenho global do modelo de interesse, define-se a medida total de discrepância do resíduo de Pearson como (total de Pearson),

$$D = \sum_{i=1}^n \sum_{j=1}^J \frac{(y_{ij} - p_{ij})^2}{p_{ij}(1 - p_{ij})}. \quad (5.3)$$

Assim, de forma empírica, quanto menor o valor de D mais apropriado é o modelo.

5.2 Um Procedimento Bayesiano para Verificação da adequação de Modelos - Fator de Bayes

O fator de Bayes é uma ferramenta Bayesiana para inferência estatística, e pode ser visto como uma analogia da estatística da razão de verossimilhança.

O fator de Bayes é a razão das verossimilhanças marginais dos dados y sob dois modelos concorrentes M_1 e M_0 , permitindo assim que se faça uma comparação dos dois modelos. O fator de Bayes B_{10} para o modelo M_1 contra o modelo M_0 é a razão dada por,

$$B_{10} = \frac{p(y|M_1)}{p(y|M_0)}.$$

Considerando o k -ésimo modelo (M_k), a verossimilhança marginal é obtida através da seguinte integral:

$$p(y|M_k) = \int L(y|\underline{\theta}_k, M_k)\pi(\underline{\theta}_k|M_k)d\underline{\theta}_k.$$

onde

$L(y|\underline{\theta}_{\sim k}, M_k)$ é a função de verossimilhança sob o modelo M_k ;

$\pi(\underline{\theta}_{\sim k} | M_k)$ é a distribuição a priori para os parâmetros do modelo M_k ;

$\underline{\theta}_{\sim k}$ é o vetor de parâmetros do modelo M_k .

Indiferente da densidade à priori usada para o modelo em questão, o problema reside no cálculo das integrais que geralmente são intratáveis matematicamente. Assim a solução é fazermos uma aproximação do fator de Bayes através de aproximações das integrais do numerador e do denominador do fator de Bayes. A seguir temos algumas das possíveis aproximações, e como estamos interessados em estimar a integral independente do modelo omitiremos as referências para o modelo M_k , assim como os k 's subscritos dos $\underline{\theta}$.

5.2.1 Aproximação de Monte Carlo - (MC)

Como uma primeira forma de aproximação para integral do tipo

$$p(y) = \int L(y|\underline{\theta})\pi(\underline{\theta})d\underline{\theta}, \quad (5.4)$$

é a integração de Monte Carlo, que fornece o estimador

$$\hat{p}_{MC}(y) = \frac{1}{m} \sum_{t=1}^m L(y|\underline{\theta}^t),$$

onde $L(y|\underline{\theta}^t)$ é a verossimilhança de y , $\{\underline{\theta}^t : t = 1, \dots, m\}$ são as amostras da densidade $\pi(y)$ e $\pi(\underline{\theta})$ é a distribuição a priori. A integração de MC garante a convergência de seu estimador mas esta convergência pode ser lenta. Este resultado se deve a possíveis discrepâncias entre $L(y|\underline{\theta}^t)$ e $\pi(\underline{\theta})$.

5.2.2 Aproximações de Monte Carlo em Cadeias de Markov - (MCMC)

A potencial ineficiência da integração de MC conduz ao desenvolvimento de uma amostragem por importância, na qual a integral (5.4) é escrita por:

$$p(y) = \int L(y|\underline{\theta}) \left[\frac{\pi(\underline{\theta})}{k\pi^*(\underline{\theta})} \right] k\pi^*(\underline{\theta})d\underline{\theta},$$

onde $L(y|\underline{\theta})$ e $\pi(\underline{\theta})$ novamente, representam a verossimilhança e a distribuição a priori, respectivamente, mas agora a amostra $\underline{\theta}^t$ é tirada da distribuição $k\pi^*(\underline{\theta})$, chamada de densidade de importância sendo k a constante normalizadora. Portanto

$$p(y) = \mathbb{E} \left[\frac{L(y|\underline{\theta})\pi(\underline{\theta})}{k\pi^*(\underline{\theta})} \right]$$

assim,

$$\hat{p}_{IS}(y) = \frac{1}{m} \sum_{t=1}^m \frac{\pi(\underline{\theta}^t)}{k\pi^*(\underline{\theta}^t)} L(y|\underline{\theta}^t)$$

Como k é a constante normalizadora da densidade de importância $k\pi^*(\underline{\theta})$, k pode ser escrita como:

$$k = \int k\pi(\underline{\theta})d\underline{\theta} = \int \frac{\pi(\underline{\theta})}{\pi^*(\underline{\theta})} k\pi(\underline{\theta})d\underline{\theta} \quad (5.5)$$

e um estimador pelo método MCMC amostrando os $\underline{\theta}$ da densidade $k\pi^*$ é dado por

$$\hat{k} = \frac{1}{m} \sum_{t=1}^m \frac{\pi(\underline{\theta}^t)}{\pi^*(\underline{\theta}^t)} \quad (5.6)$$

logo,

$$\hat{p}_{IS}(y) = \frac{\sum_{t=1}^m \frac{\pi(\underline{\theta}^t)}{\pi^*(\underline{\theta}^t)} L(y|\underline{\theta}^t)}{\sum_{t=1}^m \frac{\pi(\underline{\theta}^t)}{\pi^*(\underline{\theta}^t)}} \quad (5.7)$$

Este estimador para a integral é um peso médio, onde o peso $\pi(\underline{\theta})/\pi^*(\underline{\theta})$ compensa as os vícios das tiragens incorretas da distribuição, mas a eficiência pode ser ganha amostrando de uma distribuição com uma densidade fechada para o integrando.

Newton e Raftery (1994) propuseram vários estimadores para a verossimilhança marginal baseados em diferentes escolhas de $\pi^*(\underline{\theta})$.

Algumas formas alternativas:

- O uso da priori $\pi(\underline{\theta})$ como função de amostragem de importância ($\pi^*(\underline{\theta}) = \pi(\underline{\theta})$), leva $\hat{p}_{IS}(y)$ ao estimador de Monte Carlo, pois $\sum_{t=1}^m \frac{\pi(\underline{\theta}^t)}{\pi^*(\underline{\theta}^t)} = m$, assim

$$\hat{p}_1(y) = \frac{1}{m} \sum_{t=1}^m L(y|\underline{\theta}^t), \quad (5.8)$$

- O uso da posteriori $\pi(\underline{\theta}|y)$ como função de amostragem de importância ($\pi^*(\underline{\theta}) = \pi(\underline{\theta}|y)$) leva ao estimador de média harmônica,

$$\hat{p}_2(y) = \left\{ \frac{1}{m} \sum_{t=1}^m L(y|\underline{\theta}^t)^{-1} \right\}^{-1}. \quad (5.9)$$

Este estimador, como $\hat{p}_{MC}(y)$, é consistente em simulação e é potencialmente mais eficiente porque é principalmente baseado em valores de $\underline{\theta}^t$ com grande verossimilhança. Entretanto, não satisfaz o teorema do limite central Gaussiano e pode ser instável. Isto porque amostras ocasionais de $\underline{\theta}^t$ pode ser altamente inconsistentes com $L(y|\underline{\theta})$, e alguns resultados somados $(L(y|\underline{\theta}^t))^{-1}$ pode dominar o estimador.

- Uma função de importance sampling que é uma mistura das densidades priori e posteriori: $\pi^*(\underline{\theta}) = \delta\pi(\underline{\theta}) + (1-\delta)\pi(\underline{\theta}|y)$, onde $0 < \delta < 1$. Este $\pi^*(\underline{\theta})$ dá lugar para o estimador iterativo

$$\begin{aligned} \hat{p}_3(y) &= \frac{\sum_{t=1}^m \frac{\pi(\underline{\theta}^t)L(y|\underline{\theta}^t)}{\delta\pi(\underline{\theta}^t)+(1-\delta)\pi(\underline{\theta}^t|y)}}{\sum_{t=1}^m \frac{\pi(\underline{\theta}^t)}{\delta\pi(\underline{\theta}^t)+(1-\delta)\pi(\underline{\theta}^t|y)}} = \frac{\sum_{t=1}^m \frac{\pi(\underline{\theta}^t|y)}{\delta\pi(\underline{\theta}^t)+(1-\delta)\pi(\underline{\theta}^t|y)}}{\sum_{t=1}^m \frac{\pi(\underline{\theta}^t)}{\delta\pi(\underline{\theta}^t)+(1-\delta)\pi(\underline{\theta}^t|y)}} \\ \hat{p}_3(y) &= \frac{\sum_{t=1}^m \frac{L(y|\underline{\theta}^t)}{\delta\hat{p}_3(y)+(1-\delta)L(y|\underline{\theta}^t)}}{\sum_{t=1}^m \frac{1}{\delta\hat{p}_3(y)+(1-\delta)L(y|\underline{\theta}^t)}} \end{aligned}$$

- Newton e Raftery (1994) também consideram uma variação em \hat{p}_2 no qual m amostras de $\underline{\theta}^t$ são tiradas da posteriori e $(\delta/(1-\delta))m$ adicionais amostras imaginárias são tiradas da priori, para qual eles assumiram que a verossimilhança resultante $L(y|\underline{\theta})$ é igual a sua expectativa $\pi(y)$. Substituindo o estimador próprio para $\pi(y)$ resulta outro estimador iterativo:

$$\hat{p}_3(y) = \frac{\frac{\delta m}{1-\delta} + \sum_{t=1}^m \frac{L(y|\underline{\theta}^t)}{\delta\hat{p}_3(y)+(1-\delta)L(y|\underline{\theta}^t)}}{\frac{\delta m}{(1-\delta)\hat{p}_3(y)} + \sum_{t=1}^m \frac{1}{\delta\hat{p}_3(y)+(1-\delta)L(y|\underline{\theta}^t)}}. \quad (5.10)$$

5.2.3 A Escolha do Modelo

Considerando dois modelos, M_0 e M_1 , temos que o fator de Bayes é dado por:

$$B_{10} = \frac{p(y|M_1)}{p(y|M_0)}. \quad (5.11)$$

Para avaliar a evidência de que o modelo M_1 representa melhor os dados que M_0 , Raftery (1996) apresenta uma calibração do fator de Bayes, considerando a estatística do teste da razão de verossimilhança, que é dada na tabela abaixo.

Tabela 5.1: Calibração do fator de Bayes

B_{10}	$2 \log B_{10}$	evidência de M_1
< 1	< 0	negativo (apóia M_0)
$1 \mapsto 3$	$0 \mapsto 2$	vale ser mencionado
$3 \mapsto 12$	$2 \mapsto 5$	positivo
$12 \mapsto 150$	$5 \mapsto 10$	forte
> 150	> 10	muito forte

Da qual tiramos, que dados dois modelos M_i e M_j , $i \neq j$, se $p(y|M_i) > p(y|M_j)$ decidimos pelo modelo M_i .

Para o caso de k modelos, M_1, M_2, \dots, M_k decidiremos pelo modelo de maior verossimilhança marginal.

Capítulo 6

APLICAÇÕES NUMÉRICAS

Nos capítulos anteriores descrevemos algumas modelagens para serem aplicadas para dados binários correlacionados. Neste capítulo apresentaremos um exemplo numérico para cada um dos modelos apresentados. Aqui o nosso objetivo é a comparação das diferentes modelagens, utilizando MCMC. Para execução dos programas computacionais utilizamos o software MATLAB 4.2.

6.1 Ilustração Numérica para o Capítulo 2

Para ilustrarmos as modelagens do capítulo 2 consideraremos os dados analisados por Huggins (1991), os quais são apresentados abaixo (tabela 1.6), onde as colunas representam sexo (m ou f), a idade (y: jovem, sa: semi-adulto, a: adulto), o peso em gramas, e as observações de captura de 36 animais apanhados em 6 ocasiões.

Tabela 6.1: Dados de Huggins

Captura de *peromyscus maniculatus* coletados por
V. Reid em East Stuart Gulch, Colorado

m	y	12	1	1	1	1	1	1	f	y	10	1	0	0	1	0	0
f	y	15	1	0	0	1	1	1	f	a	23	0	1	0	0	1	0
m	y	15	1	1	0	0	1	1	f	y	7	0	1	1	0	0	1
m	y	15	1	1	0	1	1	1	m	y	8	0	1	0	0	0	1
m	y	13	1	1	1	1	1	1	m	a	19	0	1	0	1	0	1
m	y	13	0	1	1	0	1	0	f	a	22	0	0	1	1	1	1
f	y	5	0	1	0	1	0	1	f	y	10	0	0	1	0	1	1
m	a	20	0	1	0	0	0	1	f	y	14	0	0	1	1	1	1
m	y	12	0	1	0	0	1	1	f	a	19	0	0	1	0	0	0
f	y	6	0	0	1	0	0	0	f	a	20	0	0	0	1	0	0
m	a	21	1	1	0	1	1	1	m	sa	16	0	0	0	1	1	1
m	y	11	1	1	1	1	1	0	f	y	11	0	0	0	1	1	0
m	sa	15	1	1	1	0	0	1	m	y	14	0	0	0	0	1	0
m	y	14	1	1	1	1	1	1	f	y	11	0	0	0	0	1	0
m	y	13	1	1	0	1	1	1	m	a	24	0	0	0	0	1	0
f	a	22	1	1	1	0	1	1	m	y	9	0	0	0	0	0	1
m	y	14	1	1	1	1	1	1	m	sa	16	0	0	0	0	0	1
m	y	11	1	0	1	1	1	0	f	a	19	0	0	0	0	0	1

Temos então:

$n = 36$: observações - animais capturados;

$J = 6$: tempos - ocasiões de captura;

y_{ij} : a resposta para a i -ésima observação ($i = 1, 2, \dots, 36$) no j -ésimo tempo ($j = 1, 2, \dots, 6$), onde:

$$y_{ij} = \begin{cases} 1 & \text{capturado,} \\ 0 & \text{não capturado.} \end{cases}$$

$x_{ij} = (x_{ij1}, x_{ij2}, x_{ij3})$: o vetor de regressão (covariadas).

Portanto o modelo terá três covariadas, sendo:

$$x_{ij1} = \begin{cases} 1 & \text{se é do sexo masculino,} \\ 0 & \text{se é do sexo feminino.} \end{cases}$$

$$x_{ij2} = \begin{cases} 1 & \text{se é jovem,} \\ 2 & \text{se é semi-adulto,} \\ 3 & \text{se é adulto.} \end{cases}$$

$$x_{ij3} = (\text{peso em gramas}) - 14.$$

Logo o modelo de Prentice(1988) é dado por,

$$\theta_{ij} = \frac{\exp\{\alpha_i + x_{ij1}\beta_{1j} + x_{ij2}\beta_{2j} + x_{ij3}\beta_{3j}\}}{1 + \exp\{\alpha_i + x_{ij1}\beta_{1j} + x_{ij2}\beta_{2j} + x_{ij3}\beta_{3j}\}} \quad (6.1)$$

$$i = 1, 2, \dots, 36$$

$$j = 1, 2, \dots, 6$$

onde:

α_i : mede a correlação entre as observações ;

β_{1j} : é o efeito da covariável x_{ij1} ;

β_{2j} : é o efeito da covariável x_{ij2} ;

β_{3j} : é o efeito da covariável x_{ij3} .

De (6.1) temos que a função de verossimilhança (2.5) é dada por,

$$L(\underset{\sim}{\alpha}, \underset{\sim}{\beta}_1, \underset{\sim}{\beta}_2, \underset{\sim}{\beta}_3 | \mathcal{D}) = \prod_{i=1}^{36} \prod_{j=1}^6 \frac{\exp\{(\alpha_i + x_{ij1}\beta_{1j} + x_{ij2}\beta_{2j} + x_{ij3}\beta_{3j})y_{ij}\}}{1 + \exp\{\alpha_i + x_{ij1}\beta_{1j} + x_{ij2}\beta_{2j} + x_{ij3}\beta_{3j}\}} \quad (6.2)$$

onde $\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{36})$ e \mathcal{D} representa o conjunto de dados (x_{ij}, y_{ij}) , $i = 1, \dots, 36$; $j = 1, \dots, 6$.

6.1.1 EMV para os Parâmetros do Modelo com Ligação Logística

Usando o método da máxima verossimilhança descrito na seção 2.1. Temos na tabela 6.2 os estimadores de máxima verossimilhança para os parâmetros do modelo de Prentice (1988), obtidos via método iterativo de Newton-Raphson, e seus respectivos intervalos de confiança (95%) obtidos por aproximação assintótica. Lembrando que para este caso consideramos efeitos fixos, ou seja, α_i ($i = 1, \dots, 36$) iguais a α .

Tabela 6.2: Estimadores de Máxima Verossimilhança

Parâmetro	EMV	Int. Conf. 95%
α	1.6929	(1.4001; 1.9858)
β_{11}	1.3060	(0.3655; 2.2464)
β_{12}	0.9334	(-0.0468; 1.9136)
β_{13}	-0.6831	(-1.6009; 0.2348)
β_{14}	0.0606	(-0.8408; 0.9619)
β_{15}	1.3575	(0.3091; 2.4058)
β_{16}	0.4252	(-0.5944; 1.4447)
β_{21}	-1.9722	(-2.4153; -1.5292)
β_{22}	-1.2288	(-1.6031; -0.8544)
β_{23}	-1.0135	(-1.3974; -0.6297)
β_{24}	-1.0931	(-1.4556; -0.7306)
β_{25}	-1.3414	(-1.7279; -0.9548)
β_{26}	-0.7094	(-1.0800; -0.3387)
β_{31}	0.3256	(0.1469; 0.5043)
β_{32}	0.1863	(0.0428; 0.3297)
β_{33}	0.0876	(-0.0556; 0.2309)
β_{34}	0.1560	(0.0178; 0.2943)
β_{35}	0.3142	(0.1561; 0.4723)
β_{36}	0.1114	(-0.0289; 0.2518)

6.1.2 Aplicação do Modelo 1

Assumimos efeitos fixos ($\alpha_i = \alpha; i = 1, \dots, 36$), para o modelo de Prentice (1988) com as densidades a priori (2.9) e $\mu_0 = 1.6$, $\sigma_0 = 0.2$, $b_{11} = 1.3$, $b_{12} = 0.9$, $b_{13} = -0.6$, $b_{14} = 0.06$, $b_{15} = 1.3$, $b_{16} = 0.4$, $b_{21} = -1.9$, $b_{22} = -1.2$, $b_{23} = -1.0$, $b_{24} = -1.0$, $b_{25} = -1.3$, $b_{26} = -0.7$, $b_{31} = 0.3$, $b_{32} = 0.1$, $b_{33} = 0.08$, $b_{34} = 0.1$, $b_{35} = 0.3$, $b_{36} = 0.1$ e $\alpha_j = 0.2$, $l = 1, 2, 3; j = 1, 2, \dots, 6$.

Para o método de Monte Carlo em Cadeias de Markov, construímos as distribuições condicionadas a posteriori, como na seção (2.3), e a partir delas geramos amostras dos parâmetros.

Como as distribuições condicionadas dos parâmetros não possuem forma padronizada, as amostras foram geradas através do algoritmo de Metropolis-Hastings, no caso deste algoritmo as amostras geradas podem ter autocorrelação alta. Para diminuir a autocorrelação geramos várias cadeias, com uma grande quantidade de pontos e selecionamos sistematicamente valores de cada cadeia para formar a amostra, ou seja, foram geradas

5 cadeias de 2000 valores cada, onde cada cadeia inicia com um valor diferente, para a amostra desprezamos os 500 primeiros valores de cada cadeia e selecionamos uma amostra final de 750 valores, tomados de 10 em 10 dos 1500 valores restantes de cada cadeia, isto é, a amostra final é formada pelos valores das iterações : 510^a, 520^a, ..., 2000^a de cada uma das 5 cadeias.

Para o início de cada cadeia foram considerados os seguintes valores: -5.00; -2.75; -0.50; 1.75; 4.00.

Na tabela 6.3, apresentamos resumos a posteriori correspondentes aos parâmetros do modelo. Onde encontram-se a média a posteriori, DP o desvio padrão, IC 95% o intervalo de credibilidade, \widehat{R} o fator de redução de escala do critério de convergência Gelman e Rubin (ver Apêndice C) e TAc a taxa de aceitação do algoritmo de Metropolis-Hastings.

Tabela 6.3: Sumário à Posteriori considerando $\alpha_i = \alpha$

Parâmetro	Média	DP	IC 95%	$\sqrt{\widehat{R}}$	TAc
α	1.6255	0.1762	(1.2708; 1.9756)	1.0011	0.6238
β_{11}	1.3177	0.2022	(0.9157; 1.7084)	0.9984	0.8971
β_{12}	0.9018	0.1942	(0.5086; 1.2800)	1.0056	0.8880
β_{13}	-0.5956	0.1955	(-0.9871; -0.2010)	1.0043	0.9070
β_{14}	0.0629	0.2019	(-0.3725; 0.4427)	1.0034	0.8985
β_{15}	1.3072	0.1937	(0.8925; 1.7308)	1.0004	0.9149
β_{16}	0.3999	0.2035	(0.0150; 0.7840)	1.0051	0.9100
β_{21}	-1.9060	0.2082	(-2.3076; -1.4975)	1.0062	0.7535
β_{22}	-1.2081	0.1968	(-1.5937; -0.8167)	1.0000	0.6974
β_{23}	-0.9963	0.2028	(-1.3845; -0.5993)	1.0061	0.7254
β_{24}	-1.0933	0.2083	(-1.4968; -0.6724)	1.0008	0.7062
β_{25}	-1.3101	0.2044	(-1.7068; -0.9329)	1.0018	0.7330
β_{26}	-0.7078	0.1965	(-1.0747; -0.3024)	1.0025	0.7135
β_{31}	0.3057	0.1993	(-0.1013; 0.6896)	1.0034	0.5000
β_{32}	0.1807	0.1992	(-0.2258; 0.5911)	1.0040	0.4162
β_{33}	0.0766	0.2089	(-0.3093; 0.4718)	1.0027	0.4365
β_{34}	0.1471	0.1922	(-0.2292; 0.5129)	1.0042	0.4165
β_{35}	0.3076	0.0965	(0.0973; 0.5050)	1.0014	0.7040
β_{36}	0.1057	0.1992	(-0.3019; 0.4957)	1.0070	0.4310

Pela tabela (6.3) verificamos que a taxa de aceitação dos candidatos a parâmetros, gerados pela densidade normal como sendo valores da distribuição a posteriori desejada, estão entre 41.62% e 91.49%, o que pode ser considerado como alto, indicando que o núcleo

de transição foi bem escolhido. Pelos valores de $\sqrt{\widehat{R}}$ verificamos que a convergência foi atingida e as amostras selecionadas das iterações são *i.i.d.*, pois $\sqrt{\widehat{R}} < 1.1$, portanto o número de iterações é suficiente.

Constatamos que houve a convergência dos valores para uma distribuição de equilíbrio também através dos gráficos que mostram as trajetórias das cadeias geradas ao longo das iterações para cada parâmetro, A seguir damos dois exemplos destes gráficos, e omitimos os demais por conveniência:

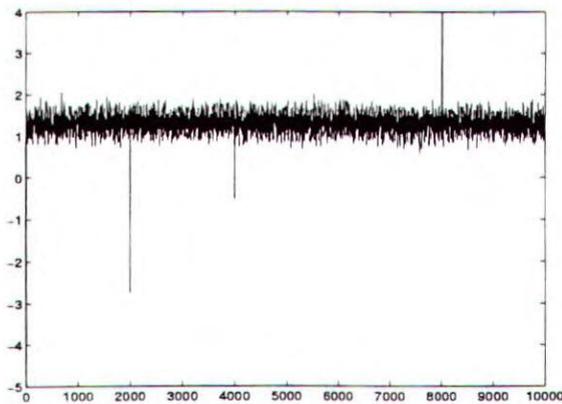


Figura 6.1: Iterações para β_{15}

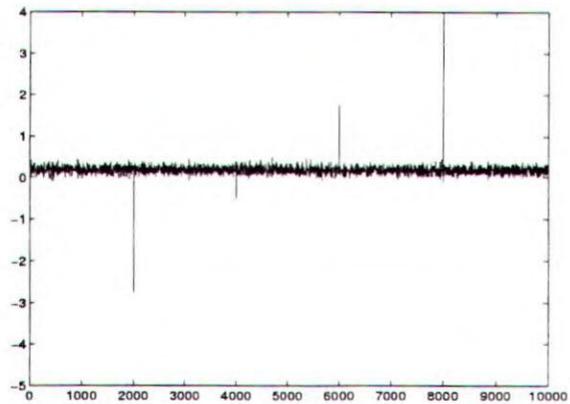


Figura 6.2: Iterações para β_{32}

Através das figuras de (6.1) e (6.2) podemos observar que após um período inicial, os gráficos apresentam o mesmo comportamento, o que mostra a convergência do algoritmo de Metropolis-Hastings com 5 condições iniciais diferentes para os 19 parâmetros.

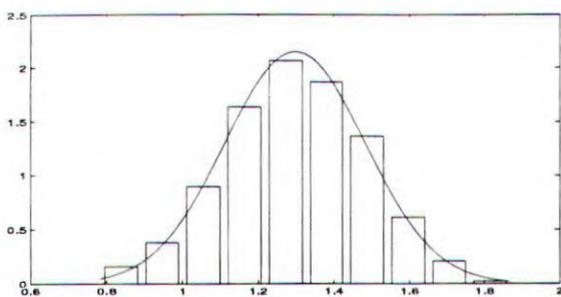


Figura 6.3: Aproximação da densidade a posteriori marginal de β_{15}

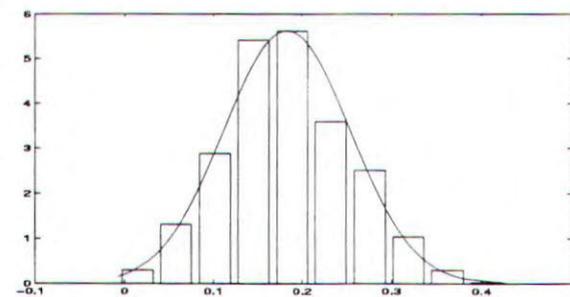


Figura 6.4: Aproximação da densidade a posteriori marginal de β_{32}

As figuras (6.3) a (6.4) são aproximações das densidades a posteriori marginais dos parâmetros do modelo, obtidas a partir das amostras selecionadas, geradas pelo algoritmo de Metropolis-Hastings.

Na tabela 6.4 e na figura 6.5 apresentamos os resíduos de Pearson d_{ij} (5.2), que expressa a medida de discrepância entre os dados e o modelo, dado por,

$$d_{ij} = \frac{y_{ij} - \tilde{p}_{ij}}{\sqrt{\tilde{p}_{ij}\tilde{q}_{ij}}}, \quad (6.3)$$

onde \tilde{p}_{ij} é t_{ij} dado em (6.1) com os parâmetros estimados por Bayes (tabela 6.3) e $\tilde{q}_{ij} = 1 - \tilde{p}_{ij}$, $i = 1, 2, \dots, 36$; $j = 1, 2, \dots, 6$.

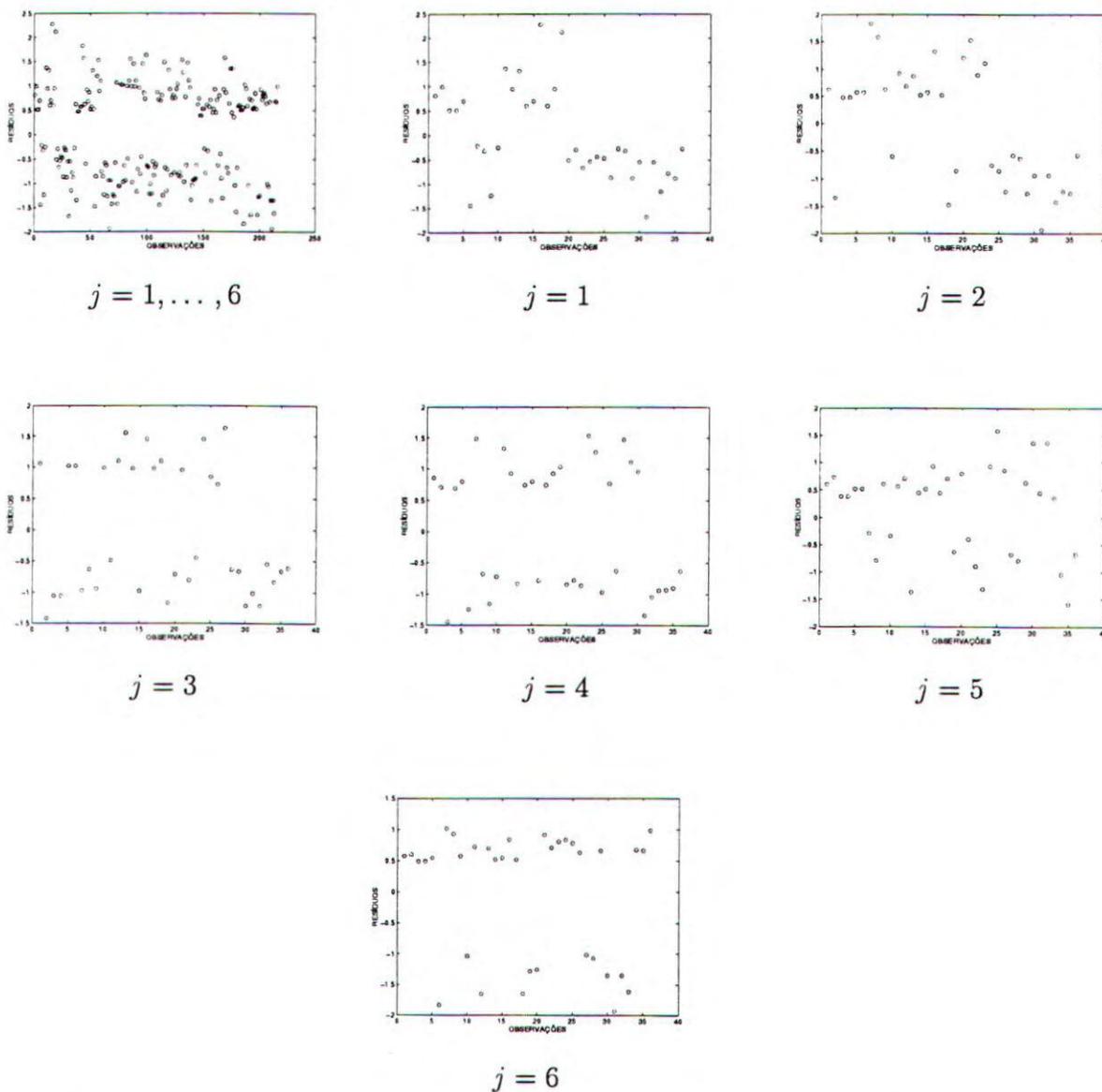


Figura 6.5: Resíduos de Pearson considerando $\alpha_i = \alpha(i = 1, 2, \dots, 36)$

Tabela 6.4: Resíduos de Pearson considerando

$$\alpha_i = \alpha(i = 1, 2, \dots, 36)$$

		J					
	1	2	3	4	5	6	
0.8082	0.6195	1.0616	0.8603	0.6043	0.5752		
0.9875	-1.3486	-1.4232	0.7120	0.7323	0.5995		
0.5110	0.4724	-1.0566	-1.4493	0.3809	0.4908		
0.5110	0.4724	-1.0566	0.6900	0.3809	0.4908		
0.6937	0.5659	1.0217	0.7993	0.5181	0.5456		
-1.4416	0.5659	1.0217	-1.2511	0.5181	-1.8330		
-0.2196	1.8302	-0.9703	1.4856	-0.2933	1.0169		
-0.3233	1.5798	-0.6364	-0.6798	-0.7949	0.9341		
-1.2373	0.6195	-0.9420	-1.1624	0.6043	0.5752		
-0.2559	-0.5980	0.9918	-0.7245	-0.3421	-1.0367		
1.3736	0.9195	-0.4910	1.3243	0.5611	0.7255		
0.9417	0.6780	1.1031	0.9260	0.7048	-1.6491		
1.3252	0.8642	1.5575	-0.8390	-1.3635	0.6992		
0.5954	0.5171	0.9833	0.7426	0.4443	0.5175		
0.6937	0.5659	-0.9787	0.7993	0.5181	0.5456		
2.2783	1.3186	1.4554	-0.7876	0.9250	0.8404		
0.5954	0.5171	0.9833	0.7426	0.4443	0.5175		
0.9417	-1.4748	1.1031	0.9260	0.7048	-1.6491		
2.1205	-0.8584	-1.1752	1.0285	-0.6329	-1.2808		
-0.5114	1.2047	-0.7139	-0.8477	0.7931	-1.2545		
-0.2981	1.5277	0.9546	-0.7798	-0.3990	0.9149		
-0.6713	0.8891	-0.8082	-0.8661	-0.8945	0.7106		
-0.5363	1.1016	-0.4548	1.5342	-1.3102	0.8063		
-0.4389	-0.7584	1.4554	1.2697	0.9250	0.8404		
-0.4716	-0.8584	0.8510	-0.9723	1.5801	0.7808		
-0.8691	-1.2321	0.7301	0.7664	0.8541	0.6320		
-0.2775	-0.5783	1.6326	-0.6316	-0.6815	-1.0154		
-0.3233	-0.6330	-0.6364	1.4709	-0.7949	-1.0705		
-0.8792	-1.2665	-0.6671	1.1074	0.6289	0.6633		
-0.5495	-0.9396	-1.2210	0.9556	1.3549	-1.3503		
-1.6797	-1.9340	-1.0169	-1.3466	0.4443	-1.9325		
-0.5495	-0.9396	-1.2210	-1.0465	1.3549	-1.3503		
-1.1515	-1.4262	-0.5507	-0.9415	0.3537	-1.6153		
-0.7822	-1.2310	-0.8397	-0.9322	-1.0432	0.6740		
-0.8792	-1.2665	-0.6671	-0.9030	-1.5902	0.6633		
-0.2775	-0.5783	-0.6125	-0.6316	-0.6815	0.9848		

6.1.3 Aplicação do Modelo 2

Para uma análise Bayesiana assumindo efeitos aleatórios α_i , $i = 1, 2, \dots, 36$ com uma distribuição $N(0, \sigma_\alpha^2)$ com as densidades a priori (2.14) onde $a = 4$, $b = 3$ e os demais valores b_{ij} e c_{ij} como sendo os mesmos usados no modelo 1.

A partir das distribuições condicionais obtidas da distribuição a posteriori (2.11), geramos 5 cadeias de Gibbs, com 6000 iterações cada, descartamos as 3000 primeiras iterações e a partir das 3000 restantes selecionamos com saltos de 10 iterações obtendo uma amostra de 1500 pontos para cada parâmetro.

As quantidades à posteriori de interesse obtidas a partir das amostras selecionadas encontram-se na tabela 6.5, onde observamos que o fator de redução de escala potencial estão bem próximos de 1.1 ($\sqrt{\widehat{R}} < 1.1$), indicando a possível convergência das amostras geradas.

Tabela 6.5: Sumário à Posteriori considerando $\alpha_i \sim N(0, \sigma_\alpha^2)$

Parâmetro	Média	DP	IC 95%	$\sqrt{\widehat{R}}$	TAc
σ_α^2	0.3267	0.2762	(0.0568; 1.0388)	1.0477	-
β_{11}	1.3715	0.1859	(0.9998; 1.7310)	1.0040	0.7899
β_{12}	0.9660	0.1874	(0.5949; 1.3310)	1.0023	0.8031
β_{13}	-0.5516	0.1914	(-0.9291; -0.1778)	0.9999	0.8505
β_{14}	0.1321	0.1857	(-0.2251; 0.4993)	1.0021	0.7943
β_{15}	1.3635	0.1885	(0.9837; 1.7359)	1.0015	0.8306
β_{16}	0.4574	0.1845	(0.0845; 0.7912)	1.0038	0.8183
β_{21}	-1.2067	0.1614	(-1.5263; -0.8966)	1.0002	0.6757
β_{22}	-0.4186	0.1560	(-0.7417; -0.1178)	1.0031	0.6549
β_{23}	-0.2174	0.1552	(-0.5212; 0.0865)	1.0021	0.6697
β_{24}	-0.3053	0.1529	(-0.6073; -0.0029)	1.0001	0.6253
β_{25}	-0.4614	0.1603	(-0.7772; -0.1434)	1.0011	0.7206
β_{26}	0.1390	0.1558	(-0.1690; 0.4450)	1.0014	0.6733
β_{31}	0.2272	0.0926	(0.0501; 0.4208)	0.9992	0.4773
β_{32}	0.0785	0.0760	(-0.0657; 0.2384)	1.0030	0.3773
β_{33}	-0.0480	0.0764	(-0.2071; 0.0993)	1.0001	0.4341
β_{34}	0.0550	0.0763	(-0.0908; 0.2042)	1.0020	0.3764
β_{35}	0.2052	0.0859	(0.0437; 0.3867)	0.9998	0.3955
β_{36}	0.0056	0.0761	(-0.1498; 0.1538)	0.9993	0.3699

A seguir apresentamos os valores (tabela 6.6) e os gráficos (figura 6.6) do resíduo de Pearson para os dados .

Tabela 6.6: Resíduos de Pearson ($\alpha_i \sim N(0, \sigma_\alpha^2)$) $i = 1, 2, \dots, 36$

J					
1	2	3	4	5	6
0.8437	0.6005	1.0220	0.8410	0.5709	0.5448
1.4171	-0.9715	-1.0085	0.9841	0.9870	0.8078
0.7723	0.6870	-0.7074	-1.0033	0.5401	0.6953
0.6959	0.6191	-0.7851	0.8981	0.4867	0.6265
0.7601	0.5828	1.0566	0.8258	0.5200	0.5483
-0.9550	0.8028	1.4554	-0.8791	0.7163	-1.3239
-0.2234	1.5459	-1.2640	1.3142	-0.3580	0.8426
-0.3214	1.4904	-0.6208	-0.7411	-0.9202	0.8036
-0.8587	0.8289	-0.7089	-0.8614	0.7880	0.7520
-0.2002	-0.5381	1.0132	-0.6256	-0.3173	-0.9519
1.1281	0.7129	-0.5704	0.9909	0.4000	0.5141
1.0574	0.6987	1.1162	0.9671	0.7077	-1.6362
1.3368	0.8019	1.4920	-0.9097	-1.5528	0.6141
0.6972	0.5758	1.1121	0.8256	0.4822	0.5619
0.8517	0.6530	-0.8447	0.9253	0.5827	0.6144
1.8934	1.0524	1.2908	-1.0252	0.6760	0.6103
0.6851	0.5658	1.0927	0.8112	0.4738	0.5521
1.1602	-1.3044	1.2247	1.0611	0.7764	-1.4912
2.9253	-0.6825	-0.9720	1.3209	-0.5185	-1.0436
-0.4263	1.4043	-0.5450	-0.7593	0.8467	-1.1839
-0.2798	1.4323	0.8319	-0.8022	-0.4386	0.8397
-0.5160	1.0245	-0.7386	-0.7305	-0.7969	0.8034
-0.5857	0.9299	-0.4963	1.2625	-1.6887	0.6234
-0.4816	-0.8666	1.4154	1.0696	0.7413	0.6692
-0.3795	-0.7578	0.9266	-0.8405	1.7371	0.8631
-0.6442	-0.9554	0.9465	0.9891	1.0693	0.7920
-0.2558	-0.5753	1.7635	-0.6430	-0.7406	-1.1066
-0.2860	-0.5971	-0.5525	1.5161	-0.8190	-1.1075
-0.7515	-1.1629	-0.5868	1.1927	0.6482	0.6829
-0.3827	-0.7093	-0.9482	1.2861	1.7419	-1.0457
-0.8653	-1.0478	-0.5425	-0.7308	0.7993	-1.0737
-0.3432	-0.6362	-0.8505	-0.6974	1.9421	-0.9379
-0.7825	-0.9907	-0.3333	-0.6881	0.4683	-1.2315
-0.5165	-0.9071	-0.6444	-0.6709	-0.7890	0.8966
-0.6101	-0.9441	-0.4764	-0.6807	-1.2525	0.8412
-0.2555	-0.5745	-0.5662	-0.6420	-0.7395	0.9050

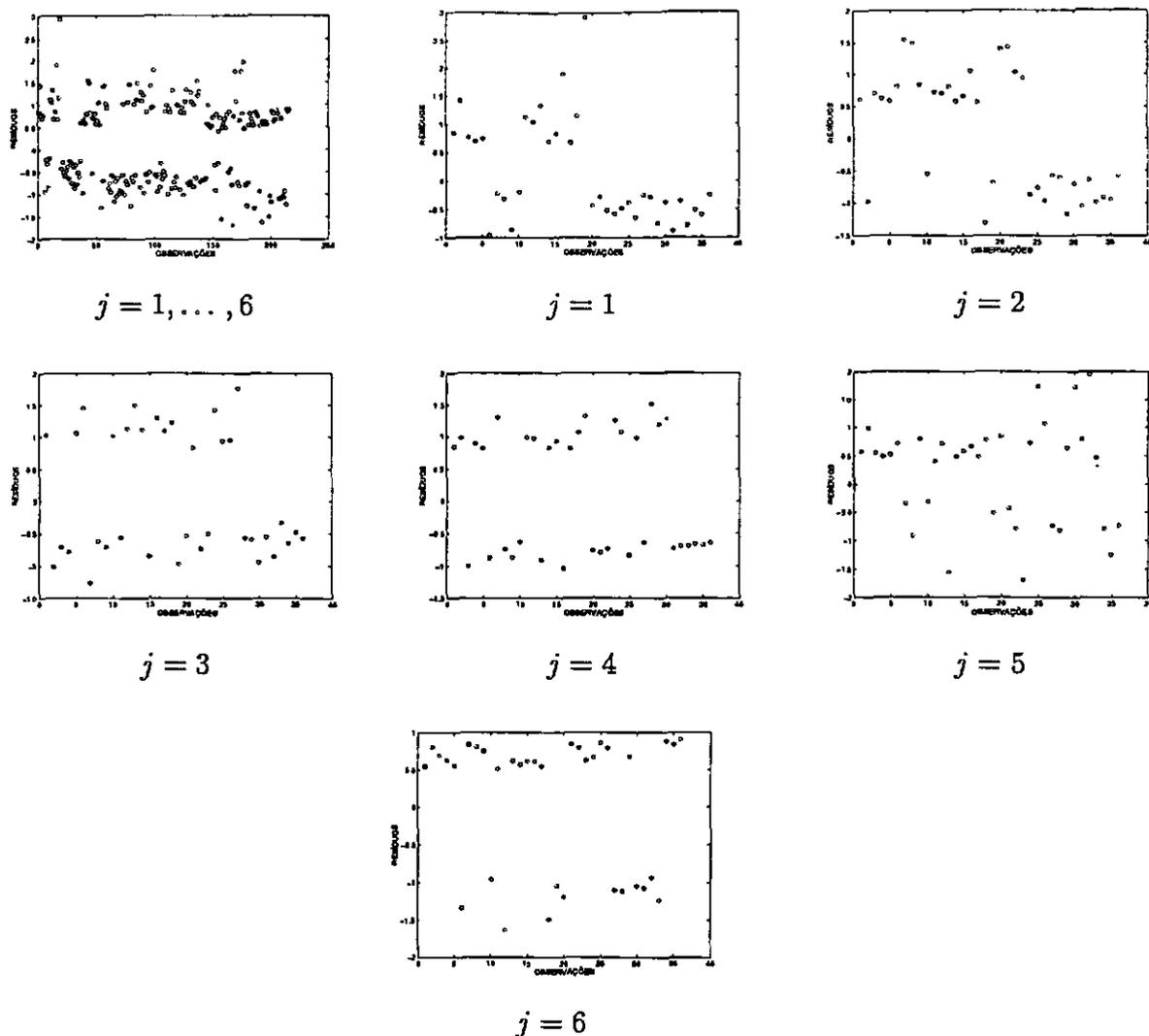


Figura 6.6: Resíduos de Pearson considerando $\alpha_i \sim N(0, \sigma_\alpha^2)$, $i = 1, 2, \dots, 36$

6.1.4 Aplicação do Modelo 3

Assumimos agora uma mistura de duas distribuições normais (2.17) para os efeitos aleatórios α_i , $i = 1, 2, \dots, 36$ e distribuições a priori (2.18) com $d_1 = -0.4$, $d_2 = 0.4$, $e_1 = e_2 = 0.2$, $a_1 = a_2 = 4$, $b_1 = b_2 = 3$, $f = 1$, $g = 1$ e os valores para b_{ij} e c_{ij} os mesmos usados para o modelo 1.

Na tabela 6.7, apresentamos os resumos a posteriori correspondentes ao modelo. Os resultados foram obtidos pelo método MCMC a partir das distribuições condicionais da distribuição a posteriori (2.2), das quais geramos 5 cadeias de 6000 iterações cada e para a amostra, desprezamos as 3000 iterações e a partir deste ponto tomamos de 10 em 10 obtendo assim uma amostra final de 1500 pontos amostrais.

Tabela 6.7: Sumário à Posteriori considerando α_i ; Mistura de Normais

Parâmetro	Média	DP	IC 95%	$\sqrt{\hat{R}}$	TAc
μ_1	-0.6478	0.3881	(-1.4215 ; 0.1078)	1.0287	-
μ_2	1.0367	0.2925	(0.5044 ; 1.6452)	1.0366	-
σ_1^2	0.1317	0.1172	(0.0380 ; 0.4470)	1.0147	-
σ_2^2	0.1308	0.0987	(0.0381; 0.3955)	1.0088	-
p_1	0.3586	0.1477	(0.0890 ; 0.6551)	1.0289	-
β_{11}	1.2975	0.1948	(0.9284 ; 1.6791)	1.0014	0.6664
β_{12}	0.8999	0.1990	(0.5140 ; 1.2726)	1.0010	0.9015
β_{13}	-0.6039	0.2042	(-1.0033 ; -0.2117)	1.0011	0.8766
β_{14}	0.0577	0.1934	(-0.3100 ; 0.4333)	1.0007	0.8829
β_{15}	1.2930	0.1964	(0.9037 ; 1.6887)	0.9999	0.6929
β_{16}	0.3962	0.2001	(-0.0247 ; 0.7842)	1.0011	0.8335
β_{21}	-1.2995	0.2061	(-1.7026 ; -0.8977)	1.0000	0.8500
β_{22}	-0.5030	0.2001	(-0.8949 ; -0.1181)	1.0016	0.8457
β_{23}	-0.3006	0.2015	(-0.7041 ; 0.0911)	1.0052	0.8393
β_{24}	-0.3952	0.2032	(-0.7832 ; -0.0161)	1.0017	0.8764
β_{25}	-0.5020	0.1972	(-0.8984 ; -0.0960)	1.0045	0.8310
β_{26}	0.0599	0.1961	(-0.3277 ; 0.4368)	1.0045	0.7898
β_{31}	0.3047	0.2019	(-0.0871 ; 0.7049)	1.0019	0.8965
β_{32}	0.1725	0.2024	(-0.2293 ; 0.5590)	1.0022	0.7187
β_{33}	-0.0824	0.2047	(-0.4873 ; 0.3255)	1.0033	0.8393
β_{34}	0.1416	0.1938	(-0.2369 ; 0.5188)	1.0018	0.7888
β_{35}	0.3078	0.1909	(-0.0688 ; 0.7014)	1.0024	0.7184
β_{36}	0.1159	0.2024	(-0.2681 ; 0.5214)	1.0031	0.8886

Observamos que houve a convergência dos valores para distribuição de equilíbrio através do fator de redução e escala de Gelman-Rubin (1992), que constam na tabela 6.7, onde para todos os parâmetros constatamos que $\sqrt{\hat{R}} < 1.1$. Além disso observamos os gráficos das iterações de todos os parâmetros os quais não se encontram nesta dissertação. Observamos também uma excelente taxa de aceitação dos valores gerados pelo algoritmo Metropolis-Hastings .

A seguir apresentamos a tabela 6.8 com o resíduo de Pearson para cada observação.

Tabela 6.8: Resíduos de Pearson (mistura de duas distribuições normais para α_i)

		J					
		1	2	3	4	5	6
0.7292	0.5234	0.7775	0.7326	0.4920	0.4802		
1.0291	-1.3546	-1.3195	0.7104	0.6896	0.5731		
0.6363	0.5569	-0.8247	-1.2248	0.4274	0.5561		
0.5220	0.4568	-1.0054	0.6698	0.3506	0.4562		
0.6319	0.4845	0.8176	0.6887	0.4257	0.4572		
-1.0433	0.7350	1.2402	-0.9571	0.6457	-1.4417		
-0.2121	1.7467	-1.9948	1.4402	-0.3116	1.0218		
-0.3549	1.2684	-0.4972	-0.8447	-1.1849	0.6461		
-0.9246	0.7763	-0.8671	-0.9203	0.7298	0.7122		
-0.1507	-0.3809	0.8559	-0.4549	-0.2218	-0.6329		
0.7499	0.4399	-0.5949	0.6354	0.2248	0.2966		
0.9131	0.6135	0.8023	0.8455	0.6171	-1.8279		
1.0799	0.6346	1.2488	-1.1343	-2.0543	0.4783		
0.5442	0.4458	0.8546	0.6436	0.3661	0.4328		
0.6951	0.5330	-1.1118	0.7577	0.4683	0.5030		
1.2183	0.6259	1.2808	-1.6596	0.3638	0.3374		
0.5427	0.4446	0.8522	0.6419	0.3651	0.4316		
1.0104	-1.4732	0.8877	0.9356	0.6828	-1.6519		
2.8697	-0.6760	-1.2454	1.3177	-0.5160	-1.0031		
-0.4908	1.1184	-0.3847	-0.9146	0.6075	-1.6124		
-0.2847	1.4847	0.5498	-0.7920	-0.4197	0.9191		
-0.4560	1.2084	-0.9275	-0.6289	-0.6716	0.9900		
-0.6897	0.7453	-0.4531	1.0436	-2.2939	0.4748		
-0.7434	-1.4469	1.4143	0.6653	0.4017	0.3726		
-0.4361	-0.8460	0.6416	-0.9498	1.5486	0.7966		
-0.8355	-1.2442	0.7264	0.7615	0.8033	0.6066		
-0.2595	-0.6159	2.2669	-0.6701	-0.8650	-1.2437		
-0.2682	-0.5959	-0.3757	1.5664	-0.8955	-1.1698		
-0.8469	-1.3490	-0.6035	1.0459	0.5314	0.5748		
-0.4004	-0.7270	-1.1790	1.2443	1.6842	-1.0486		
-0.7195	-0.8784	-0.4582	-0.6084	0.9349	-0.9047		
-0.2882	-0.5233	-0.8487	-0.5785	2.3399	-0.7548		
-0.8582	-1.1996	-0.2142	-0.7930	0.3477	-1.6348		
-0.4268	-0.7251	-0.7154	-0.5426	-0.6296	1.1622		
-0.5324	-0.8479	-0.3793	-0.6010	-1.1828	0.9144		
-0.2380	-0.5648	-0.4045	-0.6145	-0.7933	0.8768		

A figura 6.7 apresenta a tabela em forma de gráfico, primeiramente com todos os tempos e em seguida um gráfico para cada tempo

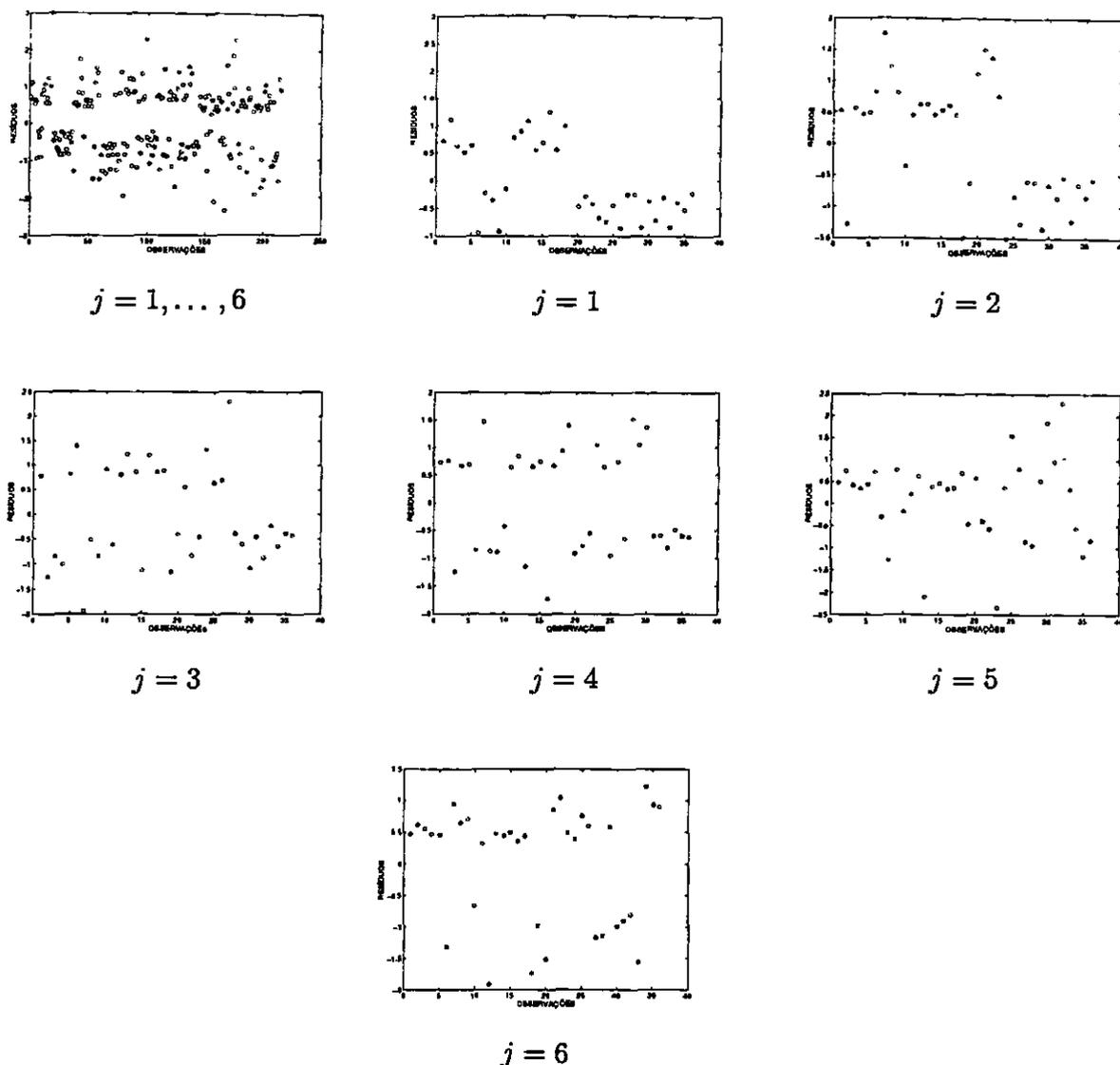


Figura 6.7: Resíduo de Pearson assumindo o mistura de normais para os efeitos aleatórios

6.1.5 Comparação dos Modelos 1, 2 e 3

Na tabela 6.9, temos um resumo dos estimadores de Monte Carlo para a média a posteriori dos parâmetros β_{lj} ($l = 1, 2, 3; j = 1, 2, \dots, 6$) baseados em amostras de Metropolis-Hastings considerando os diferentes modelos com função de ligação logística (modelo 1, modelo 2 e modelo 3).

Tabela 6.9: Médias a posteriori para β_{lj} ($l = 1, 2, 3; j = 1, 2, \dots, 6$)

Parâmetro	EMV	α Fixo	$\alpha_i \sim N(0, \sigma_\alpha^2)$	$\alpha_i \sim$ Mistura de Normais
β_{11}	1.3060	1.3177	1.3715	1.2975
β_{12}	0.9334	0.9018	0.9660	0.8999
β_{13}	-0.6831	-0.5956	-0.5516	-0.6039
β_{14}	0.0606	0.0629	0.1321	0.0577
β_{15}	1.3575	1.3072	1.3635	1.2930
β_{16}	0.4252	0.3999	0.4574	0.3962
β_{21}	-1.9722	-1.9060	-1.2067	-1.2995
β_{22}	-1.2288	-1.2081	-0.4186	-0.5030
β_{23}	-1.0135	-0.9963	-0.2174	-0.3006
β_{24}	-1.0931	-1.0933	-0.3053	-0.3952
β_{25}	-1.3414	-1.3101	-0.4614	-0.5020
β_{26}	-0.7094	-0.7078	0.1390	0.0599
β_{31}	0.3256	0.3057	0.2272	0.3047
β_{32}	0.1863	0.1807	0.0785	0.1725
β_{33}	0.0876	0.0766	-0.0480	-0.0824
β_{34}	0.1560	0.1471	0.0550	0.1416
β_{35}	0.3142	0.3076	0.2052	0.3078
β_{36}	0.1114	0.1057	0.0056	0.1159

Podemos observar nas tabelas 6.3, 6.5 e principalmente na tabela 6.7 a partir dos intervalos de credibilidade 95% para os parâmetros de regressão β_{lj} , $l = 1, 2, 3; j = 1, 2, \dots, 6$ que as covariáveis x_{ij1} (sexo) e x_{ij2} (idade) $i = 1, 2, \dots, 36; j = 1, 2, \dots, 6$ apresentam efeitos significativos na probabilidade de captura nos diferentes tempos, mas a covariável x_{ij3} (peso) não apresenta efeito significativo na probabilidade de captura, pois os intervalos de credibilidade para β_{3j} , $j = 1, 2, \dots, 6$ incluem zero.

Verificamos que nas tabelas 6.6 e 6.8 que a 19ª observação tem o valor resíduo de Pearson d_{ij} alto, sendo um possível *outlier* em específico para $j = 1$, isto fica mais visível quando observamos as figuras 6.6 e 6.7 para $j = 1$. Então na tentativa de melhorarmos os modelos retiramos a 19ª observação do conjunto de dados e obtivemos novos estimadores para os parâmetros dos modelos. Na tabela 6.10 verificamos a performance dos modelos para os dados com e sem o *outlier* (conjunto de dados com 36 e 35 obserções), através do Total de Pearson (5.3),

$$D = \sum_{i=1}^n \sum_{j=1}^J \frac{(y_{ij} - \tilde{p}_{ij})^2}{\tilde{p}_{ij}(1 - \tilde{p}_{ij})}, \quad (6.4)$$

onde \tilde{p}_{ij} é t_{ij} dado em (6.1) com os parâmetros estimados por Bayes e $\tilde{q}_{ij} = 1 - \tilde{p}_{ij}$, $i = 1, 2, \dots, 36; j = 1, 2, \dots, 6$. Também apresentamos a verossimilhança preditiva (5.9),

dada por,

$$\begin{aligned} \hat{p}_2(\hat{y}|M_k) &= \left\{ \frac{1}{m} \sum_{t=1}^m L(d|\alpha_i^{(t)}, \beta_{\sim_1}^{(t)}, \beta_{\sim_2}^{(t)}, \beta_{\sim_3}^{(t)})^{-1} \right\}^{-1} \\ &= \left\{ \frac{1}{m} \sum_{t=1}^m \left[\prod_{i=1}^{36} \prod_{j=1}^6 \frac{\exp\{(\alpha_i^{(t)} + x_{ij1}\beta_{1j}^{(t)} + x_{ij2}\beta_{2j}^{(t)} + x_{ij3}\beta_{3j}^{(t)})y_{ij}\}}{1 + \exp\{\alpha_i^{(t)} + x_{ij1}\beta_{1j}^{(t)} + x_{ij2}\beta_{2j}^{(t)} + x_{ij3}\beta_{3j}^{(t)}\}} \right]^{-1} \right\}^{-1} \quad (6.5) \\ & \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, 6; \quad k = 1, 2, 3; \end{aligned}$$

onde k indica o modelo o qual esta sendo calculada a densidade preditiva e m o tamanho da amostra gerada.

Tabela 6.10: Resultados do Total de Pearson (6.4) e as Densidades Preditivas para os Modelos

	Modelo 1	Modelo 2	Modelo 3
D (36 obs)	207.5838	185.7211	183.3444
D (35 obs)	198.2011	177.0421	175.0432
$\hat{p}_2(\hat{y} M_k)$ (36 obs)	$5.3519 \cdot 10^{-84}$	$2.9348 \cdot 10^{-61}$	$6.7799 \cdot 10^{-59}$
$\hat{p}_2(\hat{y} M_k)$ (35 obs)	$9.6933 \cdot 10^{-77}$	$3.0310 \cdot 10^{-59}$	$6.2787 \cdot 10^{-57}$

Observamos através da tabela 6.10 que para os dados binários correlacionados da tabela 6.1, considerando a função de ligação logística, o modelo de regressão logístico de efeitos aleatórios, considerando mistura de duas distribuições normais, mostra-se mais adequado. Isso também se verifica pela densidade preditiva, observando a tabela 5.1 de calibração do fator de Bayes. Podemos considerar uma forte evidência de que o modelo 3 é o mais adequado.

Além da análise dos dados que apresentamos aqui, também consideramos a possibilidade de haver interações entre as covariáveis com e sem o ponto considerado outlier, mas os resultados não foram significativos, pois os intervalos de credibilidade para estes novos parâmetros incluíam zero, e também houve um sensível aumento no Total de Pearson.

As figuras 6.8 e 6.9 apresentam o histograma das frequências (Fr) dos valores de α_i , $i = 1, 2, \dots, 36$ para o modelo 2 e o modelo 3, respectivamente. Através dos gráficos podemos verificar que modelar estes parâmetros por uma única distribuição pode não ser apropriado.

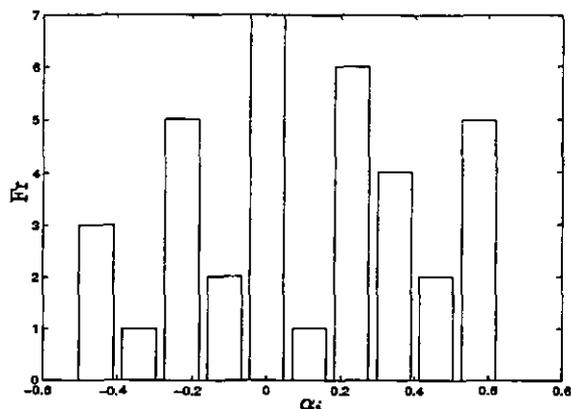


Figura 6.8: Frequências dos α_i Assumindo $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$ $i = 1, 2, \dots, 36$

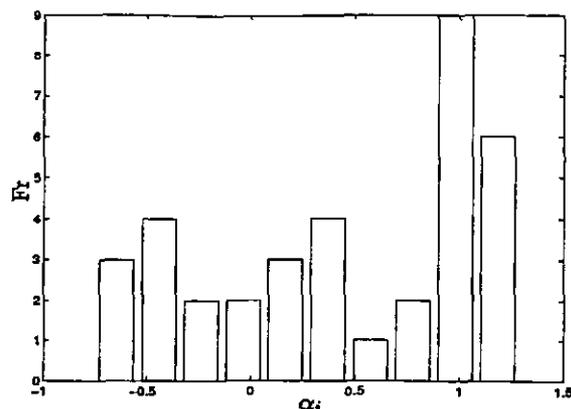


Figura 6.9: Frequências dos α_i Assumindo $\alpha_i, i = 1, 2, \dots, 36$ Mmistura de duas Normais

6.2 Ilustração Numérica para o Capítulo 3

Para exemplificar o capítulo 3 também usamos o conjunto de dados apresentados por Huggins (1991), dados na tabela 6.1.

Sem considerar interações entre as covariáveis a função de ligação probito para os dados da tabela 6.1 será dada por,

$$\theta_{ij} = \Phi\{\alpha_i + x_{ij1}\beta_{1j} + x_{ij2}\beta_{2j} + x_{ij3}\beta_{3j}\}, \quad i = 1, 2, \dots, 36 \quad (6.6)$$

$$j = 1, 2, \dots, 6$$

onde:

α_i : mede a correlação entre as observações ;

β_{1j} : é o efeito da covariável x_{ij1} ;

β_{2j} : é o efeito da covariável x_{ij2} ;

β_{3j} : é o efeito da covariável x_{ij3} .

De (6.6), a função de verossimilhança (3.2) é dada por,

$$L(\underline{\alpha}, \beta_1, \beta_2, \dots, \beta_J | \mathcal{D}) = \prod_{i=1}^{36} \prod_{j=1}^6 \Phi(\alpha_i + x_{ij1}\beta_{1j} + x_{ij2}\beta_{2j} + x_{ij3}\beta_{3j})^{y_{ij}} \cdot [1 - \Phi(\alpha_i + x_{ij1}\beta_{1j} + x_{ij2}\beta_{2j} + x_{ij3}\beta_{3j})]^{1-y_{ij}} \quad (6.7)$$

onde $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{36})$ e \mathcal{D} representa o conjunto de dados (x_{ij}, y_{ij}) , $i = 1, \dots, 36$; $j = 1, \dots, 6$.

6.2.1 EMV para os Parâmetros do Modelo com Ligação Probita

Nesta seção, considerando os dados da tabela 6.1, aplicamos a metodologia apresentada na seção 3.1 de Análise Clássica, baseada na aproximação Normal assintótica dos estimadores de máxima verossimilhança, dos parâmetros envolvidos no modelo considerando efeitos fixos, α_i , $i = 1, 2, \dots, 36$ iguais à α .

A partir do método iterativo de Newton-Raphson obtemos os estimadores de máxima verossimilhança para os parâmetros, juntamente com os respectivos intervalos de confiança assintóticos 95% descritos na tabela 6.11

Tabela 6.11: Estimadores de Máxima Verossimilhança

Parâmetro	EMV	Int. Conf. 95%
α	1.0629	(0.8853 ; 0.8853)
β_{11}	0.7799	(0.2052 ; 0.2052)
β_{12}	0.5650	(-0.0220 ; -0.0220)
β_{13}	-0.4330	(-1.0024 ; -1.0024)
β_{14}	0.0336	(-0.5209 ; -0.5209)
β_{15}	0.8677	(0.2285 ; 0.2285)
β_{16}	0.2403	(-0.3548 ; -0.3548)
β_{21}	-1.2116	(-1.4671 ; -1.4671)
β_{22}	-0.7623	(-0.9872 ; -0.9872)
β_{23}	-0.6351	(-0.8679 ; -0.8679)
β_{24}	-0.6830	(-0.9057 ; -0.9057)
β_{25}	-0.8444	(-1.0842 ; -1.0842)
β_{26}	-0.4452	(-0.6713 ; -0.6713)
β_{31}	0.2012	(0.0991 ; 0.0991)
β_{32}	0.1129	(0.0289 ; 0.0289)
β_{33}	0.0549	(-0.0333 ; -0.0333)
β_{34}	0.0973	(0.0125 ; 0.0125)
β_{35}	0.1968	(0.0987 ; 0.0987)
β_{36}	0.0698	(-0.0148 ; -0.0148)

6.2.2 Aplicação do Modelo 4

Considerando os dados introduzidos por Huggins (1991), e o modelo probito de efeitos fixos, escolhemos os seguintes valores para as densidades a priori: $\mu_0 = 1$, $\sigma_0^2 = 0.4$, $b_{11} = 0.7$, $b_{12} = 0.5$, $b_{13} = -0.4$, $b_{14} = 0.03$, $b_{15} = 0.8$, $b_{16} = 0.2$, $b_{21} = -1.2$, $b_{22} = -0.7$, $b_{23} = -0.6$, $b_{24} = -0.6$, $b_{25} = -0.8$, $b_{26} = -0.4$, $b_{31} = 0.2$, $b_{32} = 0.1$, $b_{33} = 0.05$, $b_{34} = 0.09$, $b_{35} = 0.19$, $b_{36} = 0.06$ e $c_{kj} = 0.4$, $k = 1, 2, 3$; $j = 1, 2, \dots, 6$.

Na tabela 6.12, apresentamos os resumos a posteriori correspondentes aos parâmetros do modelo 4, a partir das distribuições condicionais das quais geramos 5 cadeias de Metropolis-Hastings de 6000 amostras cada, todas as cadeias com condições iniciais diferentes, desprezamos as 3000 primeiras iterações, a partir destas tomamos a 3010^a, 3020^a, ... a fim de diminuir a auto correlação das amostras, formando assim amostras de tamanho 1500 para cada parâmetro. Podemos considerar a convergência para as distribuições de equilíbrio, pois pela tabela 6.12, $\sqrt{\widehat{R}} < 1.1$.

Tabela 6.12: Sumário à Posteriori considerando $\alpha_i = \alpha$

Parâmetro	Média	DP	IC 95%	$\sqrt{\widehat{R}}$	TAc
α	0.9917	0.1692	(0.6545; 1.3228)	1.0097	0.3354
β_{11}	0.7552	0.2334	(0.3195; 1.2041)	1.0003	0.7077
β_{12}	0.5473	0.2323	(0.1149; 1.0113)	1.0005	0.7452
β_{13}	-0.4081	0.2309	(-0.8732; 0.0384)	1.0022	0.7328
β_{14}	0.0254	0.2278	(-0.4064; 0.4601)	1.0008	0.7216
β_{15}	0.8289	0.2355	(0.3816; 1.2899)	1.0005	0.7340
β_{16}	0.2336	0.2456	(-0.2409; 0.7393)	1.0010	0.7620
β_{21}	-1.1850	0.1654	(-1.5141; -0.8638)	0.9993	0.4755
β_{22}	-0.7127	0.1540	(-1.0209; -0.3894)	1.0062	0.4262
β_{23}	-0.6081	0.1519	(-0.9072; -0.3173)	1.0026	0.4423
β_{24}	-0.6372	0.1521	(-0.9418; -0.3360)	1.0029	0.4210
β_{25}	-0.8020	0.1626	(-1.1261; -0.4766)	1.0050	0.4397
β_{26}	-0.3773	0.1686	(-0.6914; -0.0038)	1.0025	0.4740
β_{31}	0.1972	0.0610	(0.0823; 0.3176)	1.0028	0.2327
β_{32}	0.1097	0.0535	(0.0054; 0.2148)	1.0014	0.1951
β_{33}	0.0463	0.0548	(-0.0565; 0.1503)	1.0002	0.1961
β_{34}	0.0904	0.0496	(-0.0048; 0.1927)	1.0050	0.1924
β_{35}	0.1933	0.0591	(0.0794; 0.3144)	1.0036	0.2031
β_{36}	0.0746	0.0602	(-0.0386; 0.2024)	1.0120	0.2261

A seguir apresentamos (tabela 6.13 e na figura 6.10) os resíduos de Pearson d_{ij} (5.2), que expressa a medida de discrepância entre os dados e o modelo, dado por,

$$d_{ij} = \frac{y_{ij} - \tilde{p}_{ij}}{\sqrt{\tilde{p}_{ij}\tilde{q}_{ij}}}, \quad (6.8)$$

onde \tilde{p}_{ij} é θ_{ij} dado em (6.6) com os parâmetros estimados por Bayes (tabela 6.12) e $\tilde{q}_{ij} = 1 - \tilde{p}_{ij}$, $i = 1, 2, \dots, 36$; $j = 1, 2, \dots, 6$.

Tabela 6.13: Resíduos de Pearson considerando $\alpha_i = \alpha$

J					
1	2	3	4	5	6
0.8748	0.6112	1.0980	0.8529	0.5984	0.5656
0.9969	-1.3665	-1.4133	0.6989	0.7352	0.5703
0.5371	0.4600	-1.0175	-1.4609	0.3565	0.4655
0.5371	0.4600	-1.0175	0.6845	0.3565	0.4655
0.7462	0.5570	1.0581	0.7930	0.5072	0.5305
-1.3401	0.5570	1.0581	-1.2610	0.5072	-1.8849
-0.1586	1.7825	-0.9739	1.4475	-0.2539	1.0465
-0.3024	1.4825	-0.6383	-0.7385	-0.8157	0.7817
-1.1432	0.6112	-0.9107	-1.1725	0.5984	0.5656
-0.1995	-0.6154	0.9895	-0.7438	-0.3095	-1.0141
1.4107	0.8738	-0.4680	1.2330	0.5331	0.6067
1.0240	0.6695	1.1394	0.9169	0.7025	-1.6601
1.4086	0.8365	1.6082	-0.8752	-1.3903	0.6434
0.6346	0.5067	1.0197	0.7370	0.4270	0.4972
0.7462	0.5570	-0.9451	0.7930	0.5072	0.5305
2.2732	1.2401	1.4512	-0.8545	0.8999	0.6923
0.6346	0.5067	1.0197	0.7370	0.4270	0.4972
1.0240	-1.4936	1.1394	0.9169	0.7025	-1.6601
2.2658	-0.8802	-1.1719	1.0057	-0.6233	-1.2882
-0.5237	1.1355	-0.7157	-0.9186	0.7704	-1.5362
-0.2476	1.4834	0.9536	-0.8002	-0.3733	0.9290
-0.6038	0.8743	-0.7849	-0.8783	-0.8934	0.7248
-0.5086	1.0412	-0.4306	1.4279	-1.3581	0.6867
-0.4399	-0.8064	1.4512	1.1703	0.8999	0.6923
-0.4413	-0.8802	0.8534	-0.9944	1.6044	0.7763
-0.8568	-1.2503	0.7348	0.7524	0.8593	0.6074
-0.2467	-0.6158	1.6282	-0.6859	-0.6973	-1.2047
-0.3024	-0.6745	-0.6383	1.3541	-0.8157	-1.2793
-0.8328	-1.3061	-0.6462	1.0630	0.6131	0.6046
-0.5253	-0.9608	-1.2163	0.9357	1.3682	-1.3686
-1.5758	-1.9734	-0.9806	-1.3568	0.4270	-2.0112
-0.5253	-0.9608	-1.2163	-1.0687	1.3682	-1.3686
-1.1399	-1.4944	-0.5286	-1.0076	0.3124	-1.9961
-0.7109	-1.2491	-0.8148	-0.9441	-1.0425	0.6819
-0.8328	-1.3061	-0.6462	-0.9408	-1.6311	0.6046
-0.2467	-0.6158	-0.6142	-0.6859	-0.6973	0.8301

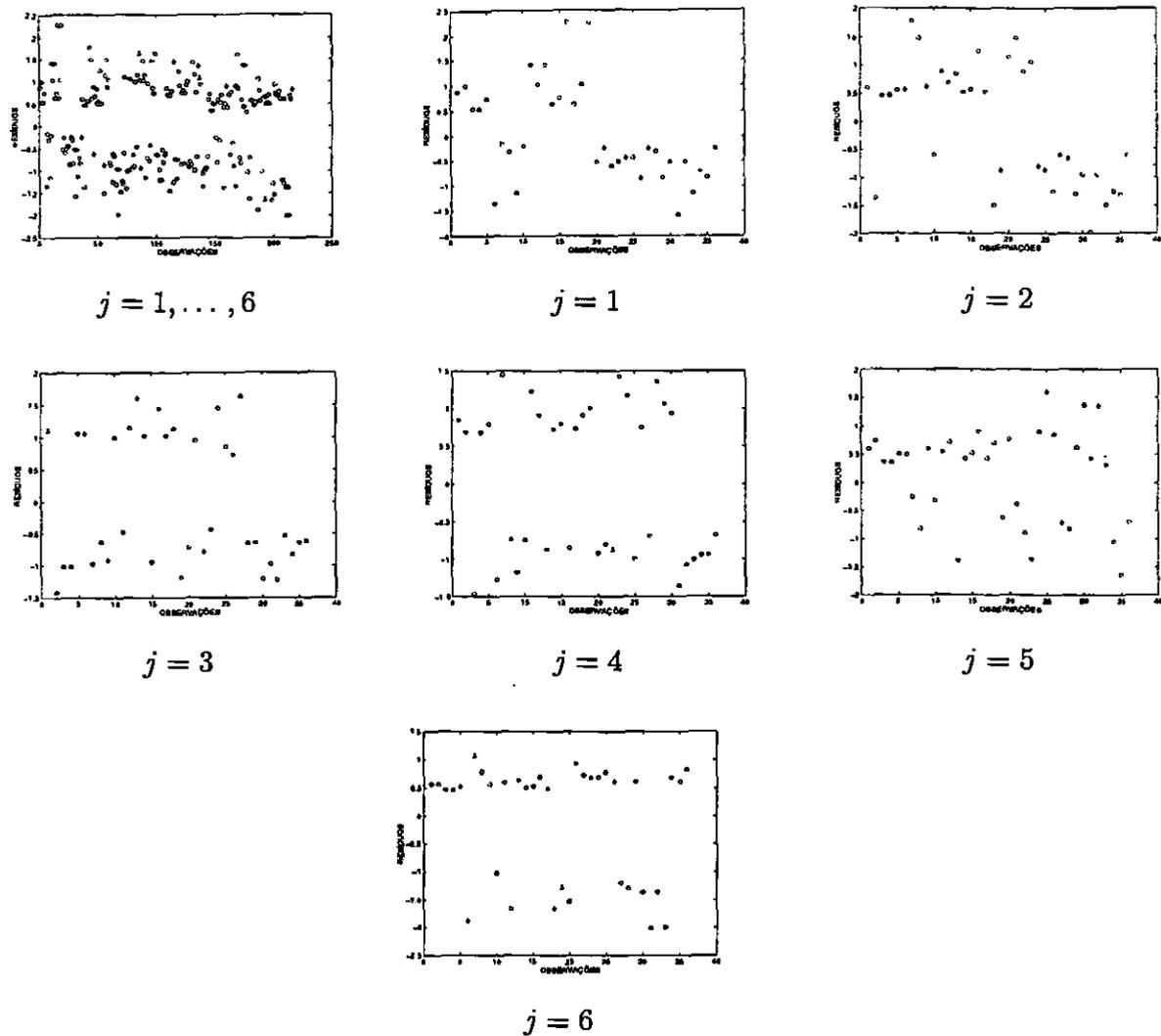


Figura 6.10: Resíduos de Pearson considerando $\alpha_i = \alpha$, $i = 1, 2, \dots, 36$

6.2.3 Aplicação do Modelo 5

Para as densidades a priori (3.6) escolhemos os seguintes valores: $a = 3$, $b = 10$ e os b_{kj} e c_{kj} , $k = 1, 2, 3$; $j = 1, \dots, 6$ como sendo os mesmos usados no modelo 4.

Na tabela 6.14 apresentamos o sumário a posteriori para os parâmetros do modelo 5, a partir das amostras geradas das distribuições condicionais a posteriori obtidas de (3.7). Para obtermos este sumário, geramos 5 cadeias de 6000 iterações e consideramos uma amostra final de 1500 pontos amostrais, que foi suficiente para verificar a convergência observando que para todos os parâmetros $\sqrt{\hat{R}} < 1.1$.

Tabela 6.14: Sumário à Posteriori considerando $\alpha_i \sim N(0, \sigma_\alpha^2)$

Parâmetro	Média	DP	IC 95%	$\sqrt{\hat{R}}$	TAc
σ_α^2	0.1860	0.1526	(0.0240; 0.5871)	1.0204	-
β_{11}	1.1736	0.2356	(0.7242; 1.6318)	1.0071	0.7227
β_{12}	1.0833	0.2413	(0.6179; 1.5696)	1.0010	0.7543
β_{13}	0.1336	0.2310	(-0.3281; 0.5621)	1.0046	0.7045
β_{14}	0.5907	0.2326	(0.1386; 1.0485)	1.0030	0.7223
β_{15}	1.5074	0.2593	(1.0005; 2.0310)	0.9991	0.7614
β_{16}	0.6974	0.2482	(0.1914; 1.1748)	1.0148	0.7656
β_{21}	-0.7696	0.1645	(-1.0881; -0.4615)	1.0042	0.4806
β_{22}	-0.3106	0.1430	(-0.5850; -0.0285)	1.0033	0.4560
β_{23}	-0.2505	0.1384	(-0.5343; 0.0196)	0.9995	0.3329
β_{24}	-0.2780	0.1379	(-0.5403; -0.0118)	1.0085	0.3848
β_{25}	-0.3539	0.1589	(-0.6742; -0.0523)	1.0006	0.5274
β_{26}	0.0688	0.1441	(-0.2118; 0.3603)	1.0163	0.4409
β_{31}	0.1317	0.0668	(0.0021; 0.2708)	1.0044	0.2352
β_{32}	0.0485	0.0565	(-0.0559; 0.1569)	1.0085	0.1979
β_{33}	-0.0093	0.0545	(-0.1201; 0.0967)	1.0024	0.1997
β_{34}	0.0342	0.0542	(-0.0712; 0.1430)	0.9994	0.1934
β_{35}	0.1321	0.0612	(0.0206; 0.2581)	1.0038	0.2240
β_{36}	-0.0082	0.0618	(-0.1272; 0.1121)	1.0032	0.2172

A partir do modelo ajustado pelos parâmetros dados na tabela 6.14 calculamos os resíduos de Pearson d_{ij} (5.2) dados na tabela 6.15.

Tabela 6.15: Resíduos de Pearson ($\alpha_i \sim N(0, \sigma_\alpha^2)$)
 $i = 1, 2, \dots, 36$

		J					
	1	2	3	4	5	6	
0.6658	0.4191	0.8089	0.6108	0.3428	0.3795		
1.3608	-0.9939	-0.9957	0.9916	0.9743	0.7769		
0.6675	0.5247	-0.8792	-1.2841	0.3435	0.5540		
0.5704	0.4445	-1.0217	0.6680	0.2850	0.4703		
0.6057	0.4075	0.8268	0.6026	0.3065	0.3890		
-1.0948	0.6328	1.2378	-1.1002	0.4909	-1.6480		
-0.2141	1.4972	-1.0661	1.3122	-0.3287	0.7312		
-0.2694	1.6483	-0.5265	-0.6128	-0.8227	0.8644		
-0.9908	0.6551	-0.8184	-1.0763	0.5474	0.5992		
-0.1573	-0.4963	1.3003	-0.5635	-0.2507	-0.9865		
0.9548	0.5356	-0.7182	0.8073	0.2302	0.3897		
0.8094	0.4824	0.8756	0.6873	0.4294	-2.4016		
1.0854	0.5935	1.2170	-1.1734	-2.4390	0.4510		
0.5456	0.3927	0.8388	0.5897	0.2703	0.3951		
0.6928	0.4718	-1.0623	0.6893	0.3590	0.4511		
1.8086	1.0112	1.2676	-0.9760	0.6543	0.5792		
0.5445	0.3918	0.8371	0.5885	0.2696	0.3942		
0.9094	-1.8273	0.9833	0.7736	0.4889	-2.1072		
2.9106	-0.6911	-0.8749	1.3446	-0.5023	-1.1251		
-0.3794	1.5197	-0.4928	-0.6414	0.9201	-1.0915		
-0.2787	1.4009	0.9643	-0.7950	-0.4158	0.7508		
-0.5988	0.8275	-0.7825	-0.8960	-1.0969	0.6301		
-0.6749	0.7351	-0.5770	1.0686	-2.5133	0.4955		
-0.4620	-0.8389	1.4993	1.2081	0.7742	0.6876		
-0.3966	-0.7819	1.0126	-0.8406	1.7471	0.7869		
-0.6694	-0.9697	0.9829	1.0047	1.0675	0.7608		
-0.1975	-0.5099	2.1581	-0.5217	-0.6511	-1.0288		
-0.2218	-0.5195	-0.4485	1.9052	-0.7098	-1.0002		
-0.8354	-1.4208	-0.6632	1.0158	0.4579	0.5670		
-0.3736	-0.6938	-0.8383	1.3552	1.8443	-1.0792		
-0.9248	-1.2420	-0.6051	-0.8596	0.5887	-1.2355		
-0.3172	-0.6028	-0.7311	-0.6420	2.1384	-0.9427		
-0.7834	-1.1274	-0.3615	-0.7324	0.3337	-1.3086		
-0.5855	-1.1065	-0.6826	-0.8110	-1.0741	0.7228		
-0.6435	-1.0972	-0.5056	-0.7614	-1.6555	0.7396		
-0.1939	-0.5029	-0.4568	-0.5145	-0.6425	0.9845		

Na figura 6.11 apresentamos os resíduos de Pearson dados na tabela 6.15.

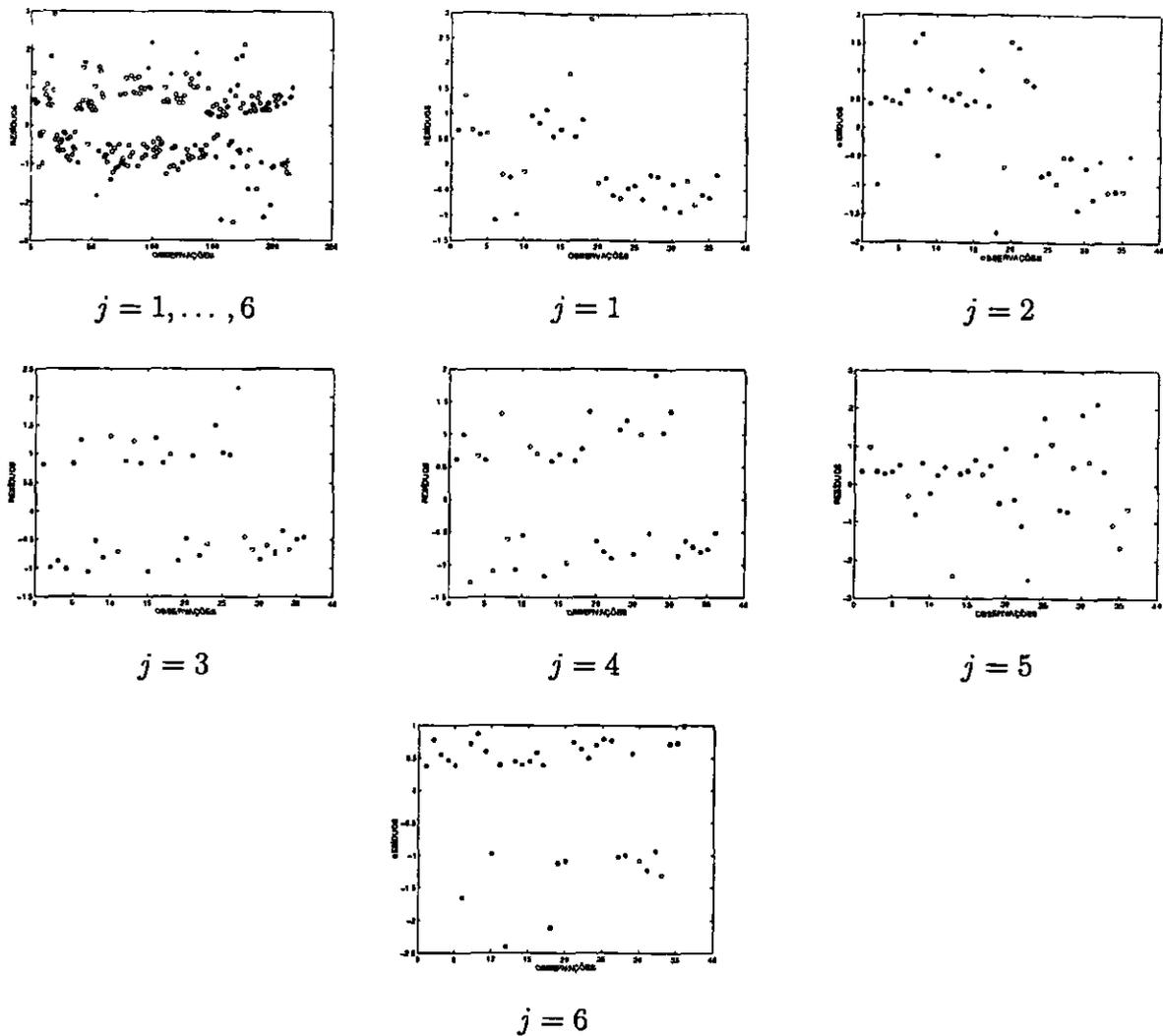


Figura 6.11: Resíduos de Pearson considerando $\alpha_i = \alpha$, $i = 1, 2, \dots, 36$

6.2.4 Aplicação do Modelo 6

Para análise Bayesiana do modelo probito assumindo uma mistura de distribuições normais para os efeitos aleatórios α_i , $i = 1, 2, \dots, 36$, consideramos os seguintes valores para as densidades a priori (3.9): $d_1 = -0.1$, $d_2 = 0.3$, $e_1 = e_2 = 0.2$, $a_1 = a_2 = 4$, $b_1 = b_2 = 3$, $f = 1$, $g = 1$ e os valores para b_{lj} e c_{lj} , $l = 1, 2, 3$; $j = 1, 2, \dots, 6$ os mesmos usados nos modelos 4 e 5.

Os resumos a posteriori são apresentados na tabela 6.16, a seguir.

Tabela 6.16: Sumário à Posteriori considerando α_i Mistura de Normais

Parâmetro	Média	DP	IC 95%	$\sqrt{\widehat{R}}$	TAc
μ_1	-0.1307	0.4743	(-0.9566; 0.8289)	1.0090	-
μ_2	0.4897	0.4070	(-0.5162; 1.1677)	1.0115	-
σ_1^2	0.1385	0.1051	(0.0427; 0.4204)	1.0337	-
σ_2^2	0.1308	0.0987	(0.0381; 0.3955)	1.0088	-
p_1	0.4337	0.2342	(0.0500; 0.9203)	1.0252	-
β_{11}	1.1476	0.2547	(0.6207; 1.6483)	1.0032	0.7317
β_{12}	1.0615	0.2451	(0.6107; 1.5566)	1.0025	0.7583
β_{13}	0.1067	0.2412	(-0.3927; 0.5662)	1.0046	0.6738
β_{14}	0.5789	0.2370	(0.1395; 1.0646)	1.0002	0.7036
β_{15}	1.5021	0.2772	(0.9593; 2.0489)	1.0005	0.7662
β_{16}	0.6976	0.2620	(0.2127; 1.2363)	1.0021	0.7723
β_{21}	-0.8815	0.1769	(-1.2300; -0.5425)	1.0093	0.4776
β_{22}	-0.4126	0.1669	(-0.7349; -0.0875)	1.0020	0.4378
β_{23}	-0.3726	0.1617	(-0.6837; -0.0561)	1.0037	0.4233
β_{24}	-0.4006	0.1607	(-0.7106; -0.0956)	1.0056	0.4460
β_{25}	-0.4473	0.1840	(-0.7950; -0.0842)	1.0017	0.4917
β_{26}	-0.0157	0.1852	(-0.3690; 0.3860)	1.0034	0.4758
β_{31}	0.0930	0.0676	(0.0164; 0.2802)	1.0006	0.2359
β_{32}	0.0574	0.0578	(-0.0631; 0.1721)	1.0004	0.1951
β_{33}	0.0040	0.0614	(-0.1222; 0.1206)	1.0039	0.2030
β_{34}	0.0552	0.0547	(-0.0461; 0.1647)	1.0013	0.1989
β_{35}	0.1493	0.0659	(0.0265; 0.2860)	1.0017	0.2288
β_{36}	0.0013	0.0719	(-0.1481; 0.1374)	1.0030	0.2312

Na tabela 6.16 podemos observar pelo fator de redução de escala de $\sqrt{\widehat{R}}$ que houve a convergência para todos os parâmetros do modelo. Para conseguirmos a convergência dos parâmetros geramos 5 cadeias de 6000 iterações e desprezamos as 3000 primeiras iterações considerando as amostras 3010^a, 3020^a,

Abaixo apresentamos os resíduos de Pearson referentes ao modelo 6 para cada observação.

Tabela 6.17: Resíduos de Pearson considerando α_i
com Mistura de Normais

J					
1	2	3	4	5	6
0.5054	0.3327	0.6817	0.5108	0.2660	0.2879
1.1820	-1.1963	-1.1835	0.8294	0.7983	0.6330
0.6667	0.4965	-0.9097	-1.3534	0.3124	0.5066
0.5067	0.3698	-1.1784	0.5646	0.2229	0.3779
0.4687	0.3168	0.6836	0.4899	0.2288	0.2897
-1.1326	0.6281	1.2585	-1.0882	0.4799	-1.7159
-0.4064	1.2545	-1.2092	1.2227	-0.3799	0.6052
-0.2047	1.5669	-0.5388	-0.6504	-0.9175	0.7870
-1.0463	0.6621	-0.7881	-1.0361	0.5494	0.5865
-0.2493	-0.5124	1.3202	-0.5257	-0.2451	-1.0260
1.1412	0.4724	-0.7848	0.6994	0.1748	0.3147
0.5876	0.3805	0.7318	0.5750	0.3360	-3.1919
0.9978	0.5101	1.0913	-1.3328	-3.0010	0.3632
0.4333	0.3008	0.6845	0.4688	0.1950	0.2910
0.5702	0.3934	-1.2177	0.5948	0.2900	0.3617
2.2502	0.8889	1.1361	-1.1417	0.5282	0.4794
0.4420	0.3075	0.6969	0.4780	0.1999	0.2975
0.7017	-2.1618	0.8694	0.6871	0.4111	-2.5981
2.2095	-0.7596	-0.9313	1.2945	-0.5436	-1.2498
-0.2486	1.4773	-0.4987	-0.6860	0.8259	-1.1751
-0.4550	1.2021	0.8635	-0.8504	-0.4716	0.6359
-0.7048	0.8789	-0.7059	-0.7890	-1.0129	0.6510
-0.5752	0.6882	-0.5954	0.9951	-2.9372	0.4303
-0.3718	-0.9640	1.3277	1.0223	0.6224	0.5670
-0.5443	-0.8997	0.9078	-0.9148	1.5406	0.6741
-0.8103	-1.1791	0.8213	0.8400	0.8720	0.6132
-0.1293	-0.4729	2.4142	-0.4835	-0.6412	-1.0043
-0.1432	-0.4925	-0.4114	1.9902	-0.7183	-0.9963
-0.7666	-1.4357	-0.6503	1.0112	0.4326	0.5336
-0.4653	-0.7568	-0.8891	1.3017	1.7078	-1.1902
-0.7761	-1.0546	-0.4971	-0.7227	0.6833	-1.0828
-0.3519	-0.5895	-0.6962	-0.5988	2.2218	-0.9346
-0.4054	-0.9155	-0.2733	-0.6255	0.3799	-1.1320
-0.5895	-0.9314	-0.5474	-0.6416	-0.8923	0.8357
-0.4793	-0.9165	-0.3996	-0.6292	-1.4318	0.8465
-0.1237	-0.4586	-0.4012	-0.4689	-0.6230	1.0237

Na figura 6.12 temos os resíduos de Pearson referentes ao modelo 6 para melhor

visualização.

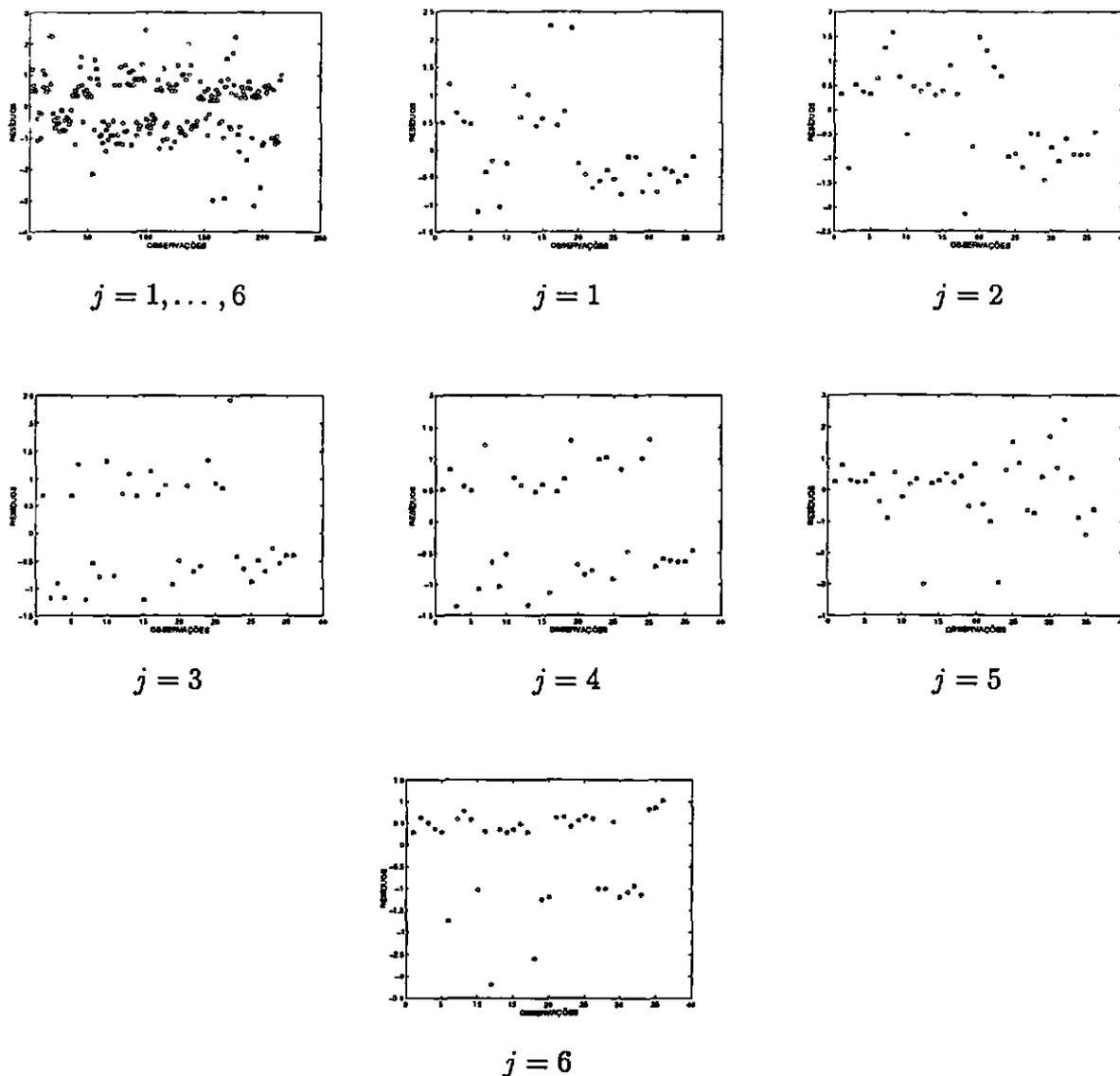


Figura 6.12: Resíduos de Pearson considerando $\alpha_i = \alpha$, $i = 1, 2, \dots, 36$

6.2.5 Comparação dos Modelos 4, 5 e 6

Para observar o comportamento dos parâmetros sob os diferentes modelos produzimos a tabela 6.18 que contém um resumo dos estimadores de MCMC para a média a posteriori dos parâmetros β_{lj} ($l = 1, 2, 3$; $j = 1, 2, \dots, 6$) dos modelos 4, 5 e 6 que consideram a função de ligação probito generalizada.

Tabela 6.18: Sumário à Posteriori para β_j

Parâmetro	EMV	α fixo	$\alpha_i \sim N(0, \sigma_\alpha^2)$	$\alpha_i \sim$ Mistura de Normais
β_{11}	0.7799	0.7328	1.1736	1.1476
β_{12}	0.5650	0.5383	1.0833	1.0615
β_{13}	-0.4330	-0.4261	0.1336	0.1067
β_{14}	0.0336	0.0426	0.5907	0.5789
β_{15}	0.8677	0.8219	1.5074	1.5021
β_{16}	0.2403	0.2355	0.6974	0.6976
β_{21}	-1.2116	-1.1880	-0.7696	-0.8815
β_{22}	-0.7623	-0.7160	-0.3106	-0.4126
β_{23}	-0.6351	-0.6065	-0.2505	-0.3726
β_{24}	-0.6830	-0.6385	-0.2780	-0.4006
β_{25}	-0.8444	-0.8069	-0.3539	-0.4473
β_{26}	-0.4452	-0.3769	0.0688	-0.0157
β_{31}	0.2012	0.1983	0.1317	0.0930
β_{32}	0.1129	0.1103	0.0485	0.0574
β_{33}	0.0549	0.0462	-0.0093	0.0040
β_{34}	0.0973	0.0897	0.0342	0.0552
β_{35}	0.1968	0.1934	0.1321	0.1493
β_{36}	0.0698	0.0691	-0.0082	0.0013

Na tabela 6.19, apresentamos o Total de Pearson (5.3) e as densidades preditivas para verificação da performance dos modelos.

Tabela 6.19: Resultados do Total de Pearson (6.4) e as Densidades Preditivas para os Modelos

	Modelo 4	Modelo 5	Modelo 6
D (36 obs)	209.4069	192.7197	191.9622
$\hat{p}_2(\hat{y} M_k)$ (36 obs)	$1.2669 \cdot 10^{-67}$	$4.3601 \cdot 10^{-64}$	$3.3753 \cdot 10^{-62}$

Pelos valores de D podemos constatar uma melhor performance do modelo de efeitos aleatórios com uma mistura de duas distribuições normais, modelo 6 sobre os modelos 4 e 5.

Calculando o fator de Bayes (5.11), temos,

- modelos 6 e 4 - $B_{64} = 2.6642 \cdot 10^5$
- modelos 6 e 5 - $B_{65} = 77.4134$
- modelos 5 e 4 - $B_{54} = 3.4416 \cdot 10^3$

verificamos uma forte evidência a favor do modelo 6 sobre os modelos 4 e 5 (tabela 5.1).

A seguir temos nas figura 6.13 e 6.14, histogramas que mostram as frequências dos α_i , $i = 1, 2, \dots, 36$ para os modelos 5 e 6 respectivamente.

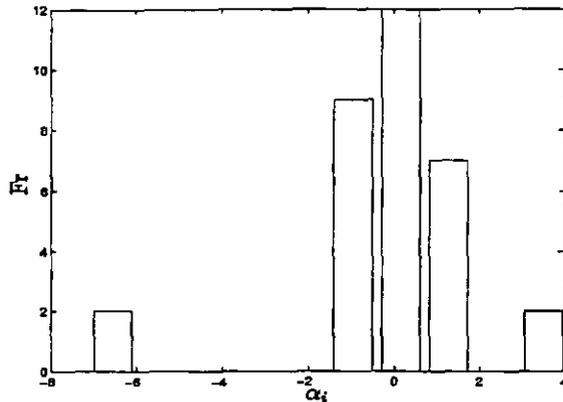


Figura 6.13: Frequências dos α_i Assumindo $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$ $i = 1, 2, \dots, 36$

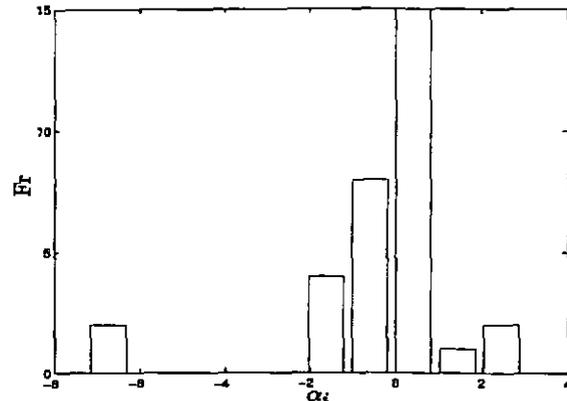


Figura 6.14: Frequências dos α_i Assumindo α_i , $i = 1, 2, \dots, 36$ Mistura de duas Normais

6.3 Ilustração Numérica para o Capítulo 4

Para ilustrar o capítulo 4 usamos o conjunto de dados (tabela 6.20) introduzido por Chib e Greenberg (1998), que contém medidas binárias repetidas do estado de bronquite para 537 crianças em idades de 7,8,9 e 10 anos de Stuebenville, Ohio. O objetivo do estudo é modelar a probabilidade do estado de bronquite com o passar do tempo, em função de uma variável indicadora binária, que representa, mães que possuem o hábito de fumar durante o primeiro ano do estudo.

Tabela 6.20: Frequência de Crianças com Bronquite

Mãe fumante					Mãe não fumante				
Idade da criança				Frequência	Idade da criança				Frequência
7	8	9	10		7	8	9	10	
0	0	0	0	237	0	0	0	0	118
0	0	0	1	10	0	0	0	1	6
0	0	1	0	15	0	0	1	0	8
0	0	1	1	4	0	0	1	1	2
0	1	0	0	16	0	1	0	0	11
0	1	0	1	2	0	1	0	1	1
0	1	1	0	7	0	1	1	0	6
0	1	1	1	3	0	1	1	1	4
1	0	0	0	24	1	0	0	0	7
1	0	0	1	3	1	0	0	1	3
1	0	1	0	3	1	0	1	0	3
1	0	1	1	2	1	0	1	1	1
1	1	0	0	6	1	1	0	0	4
1	1	0	1	2	1	1	0	1	2
1	1	1	0	5	1	1	1	0	4
1	1	1	1	11	1	1	1	1	7

Temos então:

$N = 537$: número de crianças observadas;

$n = 32$: número de grupos de frequência f_i , $i = 1, 2, \dots, 32$;

$J = 4$: número de tempos, $j = 1, 2, 3, 4$;

y_{ij} : a resposta para a i -ésima observação no j -ésimo tempo, onde:

$$y_{ij} = \begin{cases} 1 & \text{com bronquite,} \\ 0 & \text{sem bronquite.} \end{cases}$$

$x_{ij} = x_{ij1}$: covariada de regressão, onde:

$$x_{ij} = \begin{cases} 1 & \text{se a mãe é fumante,} \\ 0 & \text{se a mãe não é fumante.} \end{cases}$$

Assim Y_{ij} (0 ou 1), trata-se de resposta binária com distribuição de Bernoulli(θ_{ij}), ou seja,

$$Pr[Y_{ij} = y_{ij} | \alpha_i, \beta_j, x_{ij}] = \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \quad (6.9)$$

$$i = 1, 2, \dots, 32; \quad j = 1, 2, 3, 4.$$

onde θ_{ij} tem duas possibilidades:

a) $\theta_{ij} = \frac{\exp(\alpha_i + x_{ij1}\beta_{1j})}{1 + \exp(\alpha_i + x_{ij1}\beta_{1j})} \quad i = 1, 2, \dots, 32; \quad j = 1, 2, 3, 4$ (Logístico)

b) $\theta_{ij} = \Phi(\alpha_i + x_{ij1}\beta_{1j}) \quad i = 1, 2, \dots, 32; \quad j = 1, 2, 3, 4$ (Probit)

Para o modelo que usa a função de ligação logística, a função de verossimilhança é dada por,

$$L(\underline{\alpha}, \beta_1 | \mathcal{D}) = \prod_{i=1}^{32} \prod_{j=1}^4 \left[\frac{\exp\{(\alpha_i + x_{ij1}\beta_{1j})y_{ij}\}}{1 + \exp(\alpha_i + x_{ij1}\beta_{1j})} \right]^{f_i} \quad (6.10)$$

onde $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{32})$ e $f_i, i = 1, 2, \dots, 32$ a frequência de cada grupo.

A função de verossimilhança para o modelo de função de ligação probito é dada por,

$$L(\underline{\alpha}, \beta_1 | \mathcal{D}) = \prod_{i=1}^{32} \prod_{j=1}^4 \left[\Phi(\alpha_i + x_{ij1}\beta_{1j})^{y_{ij}} [1 - \Phi(\alpha_i + x_{ij1}\beta_{1j})]^{1-y_{ij}} \right]^{f_i} \quad (6.11)$$

onde $\underline{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_{32})$ e $f_i, i = 1, 2, \dots, 32$ a frequência de cada grupo.

Para exemplificar o capítulo 4 faremos as seguintes considerações:

i) Para função de ligação logística:

Modelo 7: Modelo logístico de efeitos fixos;

Modelo 8: Modelo logístico de efeitos aleatórios considerando $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$,
 $i = 1, 2, \dots, 32$;

Modelo 9: Modelo logístico de efeitos aleatórios com uma mistura de duas distribuições Normais.

i) Para função de ligação probito:

Modelo 10: Modelo probito de efeitos fixos;

Modelo 11: Modelo probito de efeitos aleatórios considerando $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$,
 $i = 1, 2, \dots, 32$;

Modelo 12: Modelo probito de efeitos aleatórios com uma mistura de duas distribuições Normais.

6.3.1 EMV para os Parâmetros do Modelo com Ligação Logística Considerando Réplicas

A tabela 6.21 apresenta os estimadores de máxima verossimilhança para os parâmetros sob o modelo 7.

Tabela 6.21: Estimadores de máxima verossimilhança

Parâmetro	EMV	Int. Conf. (95%)
α	-1.8212	(-1.9393; -1.9393)
β_{11}	0.2054	(-0.1801; -0.1801)
β_{12}	0.4876	(0.1348; 0.1348)
β_{13}	0.3527	(-0.0148; -0.0148)
β_{14}	-0.0021	(-0.4163; -0.4163)

6.3.2 Aplicação do Modelo 7

Considerando as distribuições a priori (4.7) com $\mu_0 = -1.8$, $\sigma_0 = 0.3$, $b_{11} = 0.2$, $b_{12} = 0.4$, $b_{13} = 0.3$, $b_{14} = 0$, $c_{1j} = 0.3$, $j = 1, 2, 3, 4$, geramos 5 cadeias cada uma com 6000 iterações e tomamos as 3010^a, 3020^a, ... iterações de cada cadeia para formar a amostra. Os resumos a posteriori são apresentados na tabela 6.22.

Tabela 6.22: Sumários a posteriori

Parâmetro	Média	DP	IC(95%)	$\sqrt{\hat{R}}$	TAc
α	-1.7928	0.2968	(-2.3935; -1.2271)	1.0052	0.9998
β_{11}	0.2022	0.3014	(-0.3978; 0.7859)	1.0026	0.9997
β_{12}	0.3990	0.3021	(-0.1904; 0.9741)	0.9999	0.9989
β_{13}	0.2953	0.3009	(-0.2579; 0.9181)	0.9999	0.9996
β_{14}	-0.0043	0.2925	(-0.5554; 0.5906)	1.0003	0.9998

Na tabela 6.23 apresentamos os resíduos de Pearson para o modelo 7.

Tabela 6.23: Resíduos de Pearson assumindo

$$\alpha_i = \alpha(i = 1, 2, \dots, 32)$$

		J			
1	2	3	4	4	
-0.4080	-0.4080	-0.4080	-0.4080	-0.4080	
-0.4080	-0.4080	-0.4080	2.4508	2.4508	
-0.4080	-0.4080	2.4508	-0.4080	-0.4080	
-0.4080	-0.4080	2.4508	2.4508	2.4508	
-0.4080	2.4508	-0.4080	-0.4080	-0.4080	
-0.4080	2.4508	-0.4080	2.4508	2.4508	
-0.4080	2.4508	2.4508	-0.4080	-0.4080	
-0.4080	2.4508	2.4508	2.4508	2.4508	
2.4508	-0.4080	-0.4080	-0.4080	-0.4080	
2.4508	-0.4080	-0.4080	2.4508	2.4508	
2.4508	-0.4080	2.4508	-0.4080	-0.4080	
2.4508	-0.4080	2.4508	2.4508	2.4508	
2.4508	2.4508	-0.4080	-0.4080	-0.4080	
2.4508	2.4508	-0.4080	2.4508	2.4508	
2.4508	2.4508	2.4508	-0.4080	-0.4080	
2.4508	2.4508	2.4508	2.4508	2.4508	
-0.4514	-0.4981	-0.4730	-0.4072	-0.4072	
-0.4514	-0.4981	-0.4730	2.4560	2.4560	
-0.4514	-0.4981	2.1144	-0.4072	-0.4072	
-0.4514	-0.4981	2.1144	2.4560	2.4560	
-0.4514	2.0075	-0.4730	-0.4072	-0.4072	
-0.4514	2.0075	-0.4730	2.4560	2.4560	
-0.4514	2.0075	2.1144	-0.4072	-0.4072	
-0.4514	2.0075	2.1144	2.4560	2.4560	
2.2151	-0.4981	-0.4730	-0.4072	-0.4072	
2.2151	-0.4981	-0.4730	2.4560	2.4560	
2.2151	-0.4981	2.1144	-0.4072	-0.4072	
2.2151	-0.4981	2.1144	2.4560	2.4560	
2.2151	2.0075	-0.4730	-0.4072	-0.4072	
2.2151	2.0075	-0.4730	2.4560	2.4560	
2.2151	2.0075	2.1144	-0.4072	-0.4072	
2.2151	2.0075	2.1144	2.4560	2.4560	

A seguir apresentamos os gráficos dos resíduos de Pearson d_{ij} .

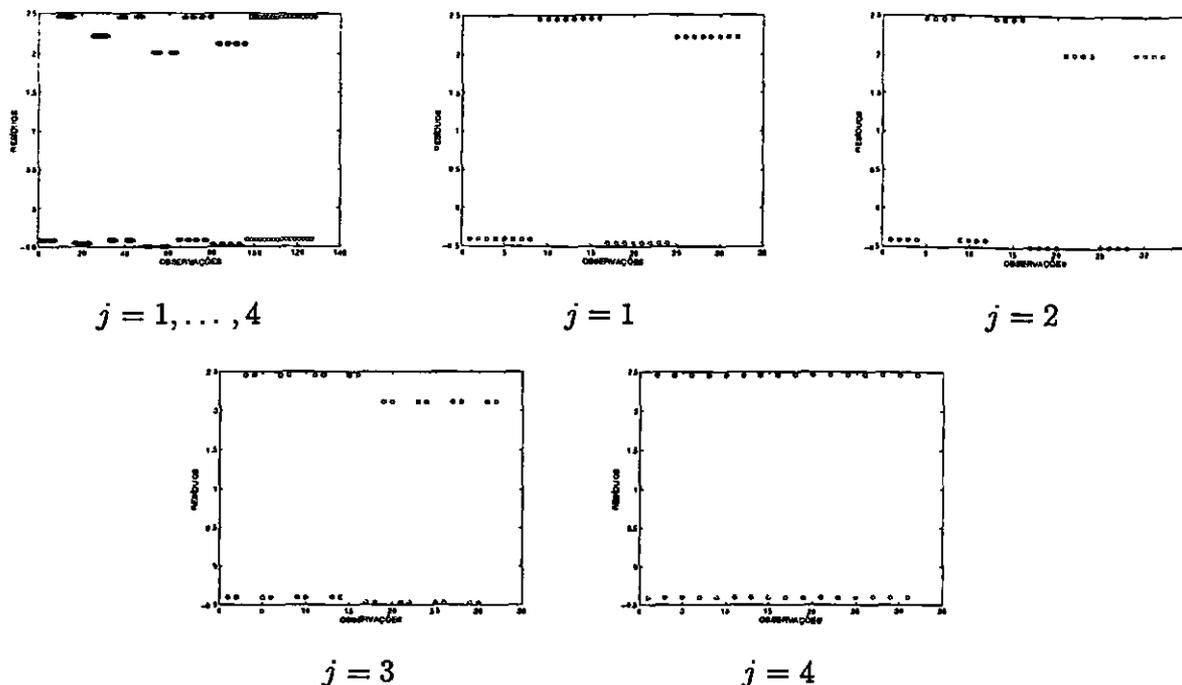


Figura 6.15: Resíduos de Pearson considerando $\alpha_i = \alpha(i = 1, 2, \dots, 32)$

6.3.3 Aplicação do Modelo 8

Para o modelo 8, modelo logístico de efeitos aleatórios com distribuição normal, consideramos os seguintes valores para as distribuições a priori (4.9): $a = 3$, $b = 8$, para b_{1j} e c_{1j} ($j = 1, 2, 3, 4$) os mesmos valores usados no modelo de efeitos fixos, modelo 7. Para obtermos o sumário a posteriori (Tabela 6.24) geramos 5 cadeias de 6000 amostras e desprezamos as 3000 primeiras, e assim selecionamos 1500 pontos amostrais com saltos de 10 iterações, sendo suficiente para que $\sqrt{\widehat{R}} < 1.1$ para todos os parâmetros do modelo.

Tabela 6.24: Sumário à Posteriori considerando $\alpha_i \sim N(0, \sigma_\alpha^2)$

Parâmetro	Média	DP	IC 95%	$\sqrt{\widehat{R}}$	TAc
σ_α^2	4.6162	1.5449	(2.4822; 8.2327)	1.0112	-
β_{11}	0.1935	0.1976	(-0.1788; 0.5750)	1.0038	0.9976
β_{12}	0.4032	0.1986	(0.0128; 0.7987)	1.0021	0.9526
β_{13}	0.2940	0.2040	(-0.1144; 0.7125)	0.9988	0.4333
β_{14}	-0.0058	0.2006	(-0.4119; 0.3814)	1.0012	0.7542

A seguir apresentamos os valores (tabela 6.25) e os gráficos (figura 6.16) do resíduo de Pearson para os dados .

Tabela 6.25: Resíduos de Pearson ($\alpha_i \sim N(0, \sigma_\alpha^2)$)
 $i = 1, 2, \dots, 36$

J			
1	2	3	4
-0.0286	-0.0286	-0.0286	-0.0286
-0.5952	-0.5952	-0.5952	1.6801
-0.6128	-0.6128	1.6317	-0.6128
-0.9690	-0.9690	1.0320	1.0320
-0.5931	1.6860	-0.5931	-0.5931
-0.9542	1.0480	-0.9542	1.0480
-0.9123	1.0961	1.0961	-0.9123
-1.6409	0.6094	0.6094	0.6094
1.6746	-0.5971	-0.5971	-0.5971
1.0252	-0.9754	-0.9754	1.0252
1.0346	-0.9665	1.0346	-0.9665
0.6314	-1.5838	0.6314	0.6314
1.0255	1.0255	-0.9751	-0.9751
0.6271	0.6271	-1.5946	0.6271
0.5940	0.5940	0.5940	-1.6836
0.1309	0.1309	0.1309	0.1309
-0.0445	-0.0494	-0.0468	-0.0402
-0.5540	-0.6153	-0.5826	1.9941
-0.5878	-0.6527	1.6180	-0.5320
-0.9381	-1.0419	1.0137	1.1776
-0.6707	1.3426	-0.7053	-0.6071
-0.9166	0.9824	-0.9638	1.2053
-0.8979	1.0029	1.0592	-0.8127
-1.6326	0.5515	0.5825	0.6767
1.7896	-0.6205	-0.5876	-0.5058
1.0590	-1.0487	-0.9930	1.1699
1.0459	-1.0618	0.9946	-0.8655
0.6706	-1.6562	0.6377	0.7408
1.0462	0.9420	-1.0051	-0.8652
0.6312	0.5683	-1.6660	0.6973
0.6112	0.5504	0.5812	-1.4810
0.1590	0.1432	0.1512	0.1757

Os valores apresentados na tabela 6.25 são colocados em forma de gráfico na figura 6.16.

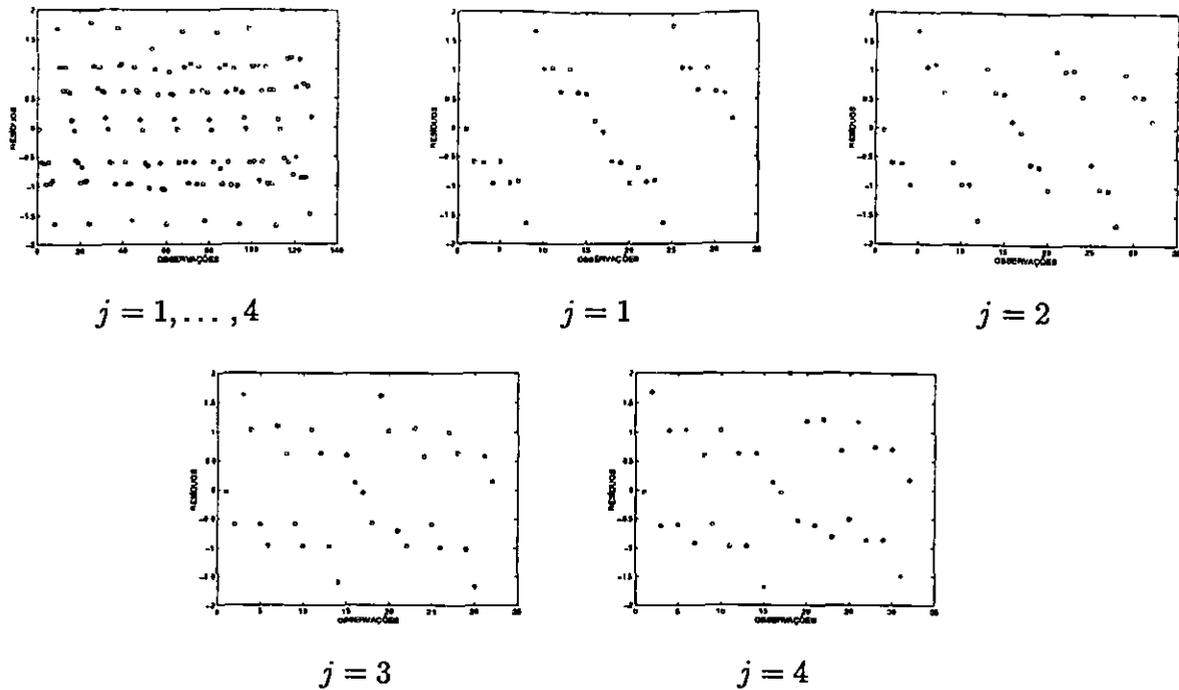


Figura 6.16: Resíduos de Pearson considerando $\alpha_i \sim N(0, \sigma_\alpha^2), i = 1, 2, \dots, 32$

6.3.4 Aplicação do Modelo 9

Assumimos uma mistura de duas distribuições normais (4.10) para os efeitos aleatórios $\alpha_i, i = 1, 2, \dots, 32$ e distribuições a priori (4.11) com $d_1 = -6.7, e_1 = 0.4, d_2 = 0.5, e_2 = 1.2, a_1 = 8, b_1 = 4, a_2 = 3, b_2 = 0.3, f = 1, g = 1$ e os valores para b_{ij} e c_{ij} os mesmos usados para o modelo 7. Foram feitas 5 cadeias de 6000 iterações cada, dessas descartamos as 3000 primeiras, e assim foram tomadas amostras de 10 em 10, portanto as inferências foram feitas sobre amostras de 1500 observações.

Tabela 6.26: Sumário à Posteriori considerando α_i com Mistura de Normais

Parâmetro	Méd.	STD	IC 95%	$\sqrt{\hat{R}}$	TAc
μ_2	-7.2261	0.5277	(-8.3457; -6.2234)	1.0032	-
μ_2	0.0924	0.2339	(-0.3536; 0.5713)	1.0014	-
σ_1^2	0.0448	0.3570	(0.0167; 0.0724)	1.0936	-
σ_2^2	1.3606	0.4320	(0.7312; 2.3887)	1.0041	-
p_1	0.0868	0.0473	(0.0183; 0.1964)	1.0006	-
β_{11}	0.1459	0.1705	(-0.1988; 0.4813)	0.9993	0.7519
β_{12}	0.4860	0.1740	(0.1313; 0.8213)	1.0022	0.7039
β_{13}	0.3197	0.1733	(-0.0301; 0.6526)	1.0015	0.7869
β_{14}	-0.1112	0.1764	(-0.4703; 0.2269)	0.9999	0.6261

Do modelo ajustado pelos parâmetros dados na tabela 6.26 calculamos os resíduos de Pearson d_{ij} (5.2) dados na tabela 6.27.

Tabela 6.27: Resíduos de Pearson (mistura de duas distribuições normais para α_i)

		J	
1	2	3	4
-0.0266	-0.0266	-0.0266	-0.0266
-0.6014	-0.6014	-0.6014	1.6629
-0.5934	-0.5934	1.6853	-0.5934
-1.0018	-1.0018	0.9983	0.9983
-0.5950	1.6807	-0.5950	-0.5950
-1.0172	0.9830	-1.0172	0.9830
-1.0046	0.9955	0.9955	-1.0046
-1.5859	0.6306	0.6306	0.6306
1.7014	-0.5878	-0.5878	-0.5878
0.9928	-1.0073	-1.0073	0.9928
0.9865	-1.0136	0.9865	-1.0136
0.6598	-1.5157	0.6598	0.6598
0.9864	0.9864	-1.0138	-1.0138
0.6681	0.6681	-1.4968	0.6681
0.6270	0.6270	0.6270	-1.5950
0.2257	0.2257	0.2257	0.2257
-0.0288	-0.0341	-0.0314	-0.0253
-0.6061	-0.7185	-0.6611	1.8762
-0.5929	-0.7029	1.5461	-0.5214
-1.0311	-1.2222	0.8891	1.1029
-0.5807	1.4527	-0.6334	-0.5107
-1.0410	0.8104	-1.1355	1.0924
-0.9816	0.8595	0.9340	-0.8632
-1.5903	0.5305	0.5765	0.7151
1.6912	-0.7009	-0.6450	-0.5200
0.9988	-1.1868	-1.0921	1.1358
1.0024	-1.1826	0.9189	-0.8773
0.7028	-1.6867	0.6443	0.7992
1.0011	0.8446	-1.0896	-0.8784
0.6484	0.5470	-1.6824	0.7373
0.6314	0.5327	0.5789	-1.3927
0.2605	0.2198	0.2388	0.2962

Na figura 6.17, apresentamos os gráficos dos resíduos de Pearson d_{ij}

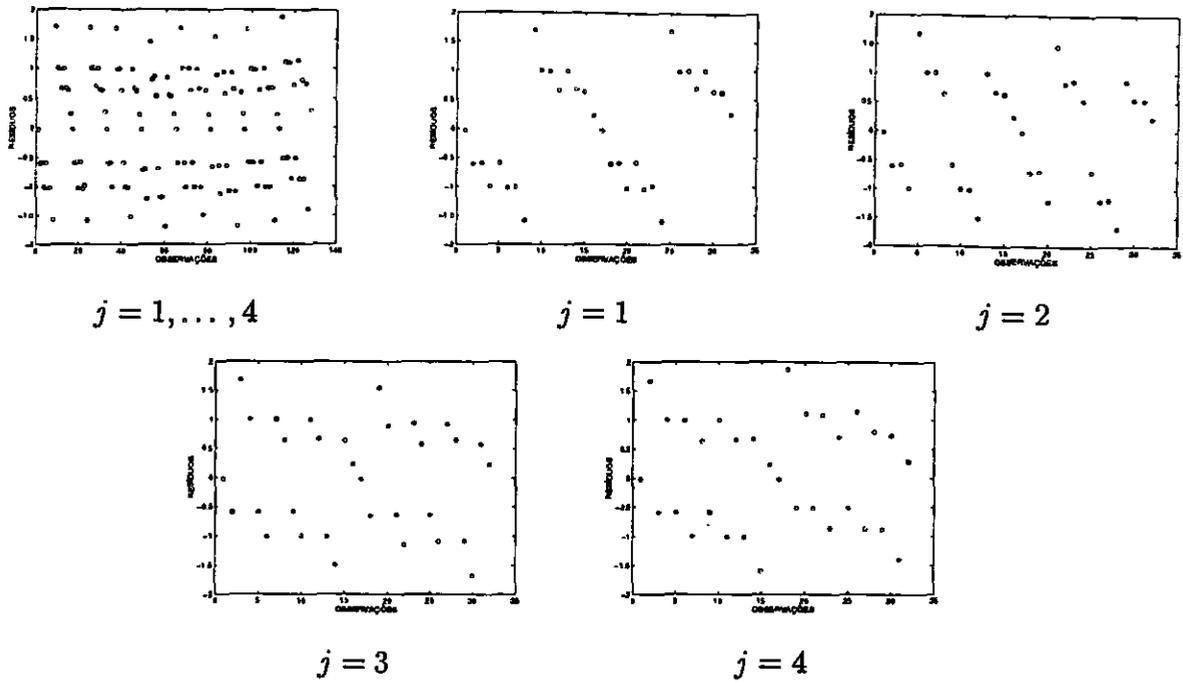


Figura 6.17: Resíduos de Pearson assumindo uma mistura de normais para os efeitos aleatórios

6.3.5 Comparação dos Modelos 7, 8 e 9

Na tabela 6.28, temos um resumo dos estimadores de Monte Carlo para a média a posteriori dos parâmetros β_{1j} ($j = 1, 2, 3, 4$) baseados em amostras geradas pelo algoritmo Metropolis-Hastings considerando os diferentes modelos com função de ligação logística (modelos 7, 8 e 9), onde pode ser observado uma variação dos estimadores entre os diferentes modelos.

Tabela 6.28: Sumário à Posteriori para β_{1j} , $j = 1, 2, 3, 4$

Parâmetro	EMV	α fixo	$\alpha_i \sim N(0, \sigma_\alpha^2)$	$\alpha_i \sim$ Mistura de Normais
β_{11}	0.2054	0.2022	0.1935	0.1459
β_{12}	0.4876	0.3990	0.4032	0.4860
β_{13}	0.3527	0.2953	0.2940	0.3197
β_{14}	-0.0021	-0.0043	-0.0058	-0.1112

Para compararmos a eficiência dos modelos, na tabela 6.29 temos o Total de Pearson dado por (5.3) e também os valores das densidades preditivas nos pontos observados, que por dificuldades numéricas nestes modelos que envolvem réplicas, tivemos que optar pela aproximação (5.8) para a densidade preditiva no lugar da aproximação de média harmônica dada por (5.9). Isso é devido a algumas frequências terem muito peso sobre

a verossimilhança, o que computacionalmente leva algumas verossimilhança para infinito ou para zero. Utilizando a estimativa (5.8) a verossimilhança preditiva é dada por,

$$\begin{aligned} \hat{p}_1(\hat{y}|M_k) &= \frac{1}{m} \sum_{t=1}^m L(d|\underset{\sim}{\alpha}^{(t)}, \underset{\sim}{\beta}_1^{(t)}) \\ &= \frac{1}{m} \sum_{t=1}^m \left[\prod_{i=1}^{32} \prod_{j=1}^4 \left(\frac{\exp\{(\alpha_i^{(t)} + x_{ij1}\beta_{1j}^{(t)})y_{ij}\}}{1 + \exp\{\alpha_i^{(t)} + x_{ij1}\beta_{1j}^{(t)}\}} \right)^{f_i} \right] \end{aligned} \quad (6.12)$$

onde $k = 7, 8, 9$, indica o modelo o qual esta sendo calculada a densidade preditiva e m o tamanho da amostra gerada.

Tabela 6.29: Resultados do Total de Pearson (6.4) e para as Densidades Preditivas para os Modelos Calculados nos Valores Observados

	Modelo 7	Modelo 8	Modelo 9
D (32 obs)	359.7745	110.3901	109.9357
$\hat{p}_1(\hat{y} M_k)$ (32 obs)	-	$2.0669 \cdot 10^{-180}$	$1.4616 \cdot 10^{-175}$

Notamos pela tabela 6.29, uma grande diferença do Total de Pearson entre o modelo 7 e os modelos 8 e 9, mostrando que os modelos de efeitos aleatórios são melhores havendo uma superioridade para o modelo 9. Superioridade esta que fica muito acentuada quando observamos as densidades preditivas, pois $B_{98} = 7.0715 \cdot 10^4$. No modelo 7 as verossimilhanças foram muito baixas impossibilitando o seu cálculo.

Nas figuras 6.18 e 6.19 apresentamos dois gráficos de frequências para os α_i , $i = 1, 2, \dots, 32$ para os modelos de efeitos aleatórios, modelos 8 e 9 respectivamente.

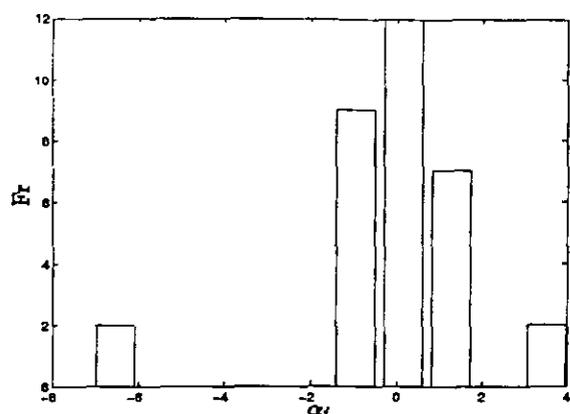


Figura 6.18: Frequências dos α_i Assumindo $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$ $i = 1, 2, \dots, 32$

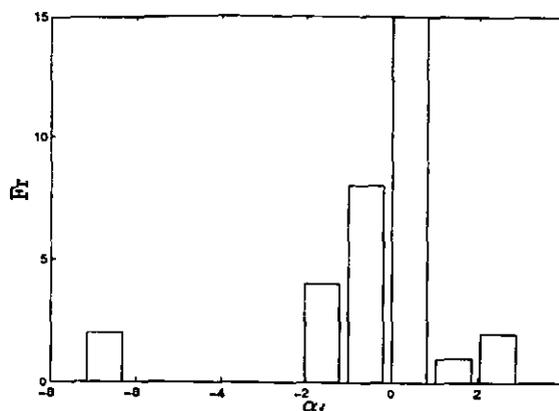


Figura 6.19: Frequências dos α_i Assumindo $\alpha_i, i = 1, 2, \dots, 32$ Mistura de duas Normais

6.3.6 EMV para os Parâmetros do Modelo com Ligação Probito Considerando Réplicas

Na tabela 6.30 apresentamos os estimadores de máxima verossimilhança e os intervalos de confiança assintóticos para os parâmetros do modelo probito assumindo efeitos fixos e réplicas.

Tabela 6.30: Estimadores de máxima verossimilhança

Parâmetro	EMV	Int. Conf. (95%)
α	-1.0835	(-1.1463; -1.0208)
β_{11}	0.1125	(-0.1079; 0.3330)
β_{12}	0.2721	(0.0221; 0.5221)
β_{13}	0.1951	(-0.0389; 0.4292)
β_{14}	-0.0011	(-0.2075; 0.2053)

6.3.7 Aplicação do Modelo 10

Considerando os dados introduzidos por Chib e Greenberg (1998), e o modelo probito de efeitos fixos com réplicas, escolhemos os seguintes valores para as densidades a priori $\mu_0 = -1$, $\sigma_0 = 0.3$, $b_{11} = 0.1$, $b_{12} = 0.2$, $b_{13} = 0.1$, $b_{14} = 0$, e $c_{1j} = 0.3$, $j = 1, 2, 3, 4$.

Na tabela 6.31, apresentamos os resumos a posteriori correspondentes aos parâmetros do modelo 10, a partir das distribuições condicionais das quais geramos 5 cadeias de

Metropolis-Hastings de 6000 pontos cada, todas as cadeias com condições iniciais diferentes, desprezamos as 3000 primeiras iterações, a partir destas tomamos as 3015^a, 3030^a, ... a fim de diminuir a auto correlação entre as amostras, formando assim amostras de 1000 pontos para cada parâmetro. Podemos considerar a convergência para as distribuições de equilíbrio, pois pela tabela 6.31 $\sqrt{\widehat{R}} < 1.1$.

Tabela 6.31: Sumário à Posteriori considerando $\alpha_i = \alpha$

Parâmetro	Média	DP	IC 95%	$\sqrt{\widehat{R}}$	TAc
α	-1.0036	0.2917	(-1.5528; -0.4330)	1.0010	0.9994
β_{11}	0.1241	0.2835	(-0.4636; 0.6657)	0.9996	0.9545
β_{12}	0.1636	0.3156	(-0.4764; 0.7805)	1.0024	0.9592
β_{13}	0.0867	0.3159	(-0.5396; 0.6894)	1.0034	0.9524
β_{14}	0.0209	0.2838	(-0.5301; 0.5662)	1.0049	0.9506

A seguir apresentamos (figura 6.20 e na tabela 6.32) os resíduos de Pearson d_{ij} (5.2), que expressa a medida de discrepância entre os dados e o modelo 10.

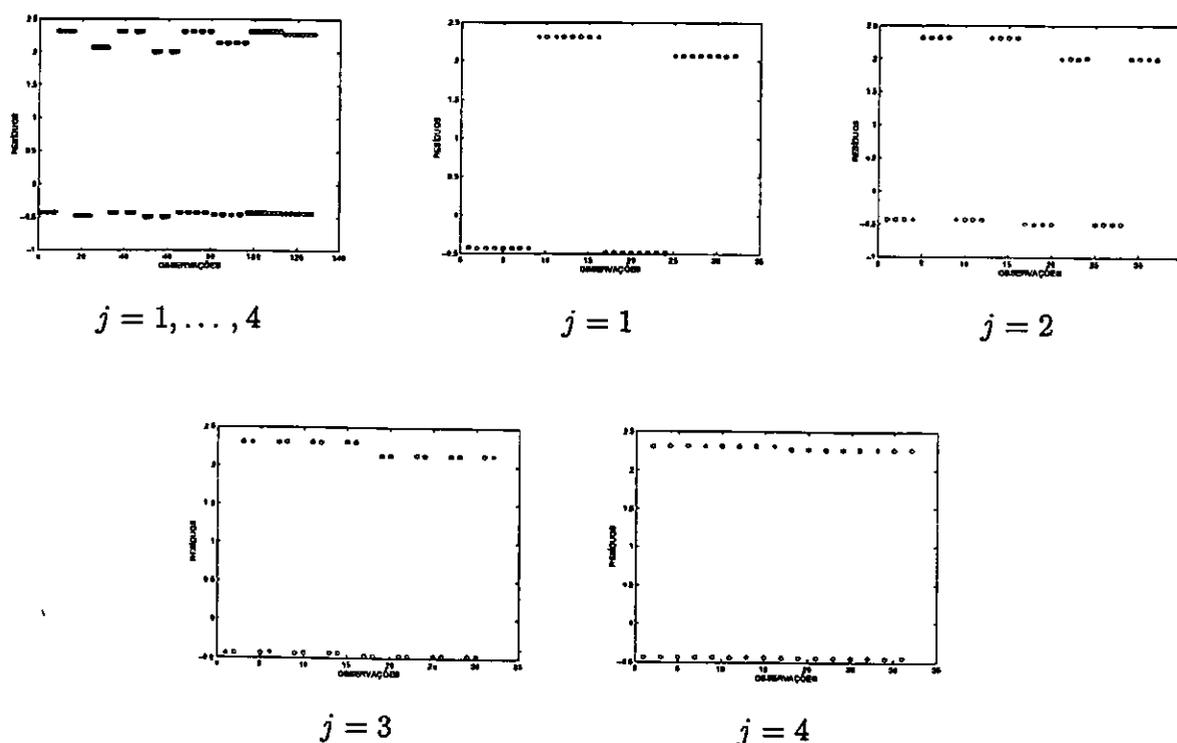


Figura 6.20: Resíduos de Pearson considerando $\alpha_i = \alpha (i = 1, 2, \dots, 32)$

Tabela 6.32: Resíduos de Pearson assumindo

$$\alpha_i = \alpha(i = 1, 2, \dots, 32)$$

		J			
		1	2	3	4
-0.4328	-0.4328	-0.4328	-0.4328	-0.4328	-0.4328
-0.4328	-0.4328	-0.4328	-0.4328	-0.4328	2.3103
-0.4328	-0.4328	-0.4328	2.3103	2.3103	-0.4328
-0.4328	-0.4328	-0.4328	2.3103	2.3103	2.3103
-0.4328	2.3103	-0.4328	-0.4328	-0.4328	-0.4328
-0.4328	2.3103	-0.4328	-0.4328	-0.4328	2.3103
-0.4328	2.3103	2.3103	2.3103	2.3103	-0.4328
-0.4328	2.3103	2.3103	2.3103	2.3103	2.3103
2.3103	-0.4328	-0.4328	-0.4328	-0.4328	-0.4328
2.3103	-0.4328	-0.4328	-0.4328	-0.4328	2.3103
2.3103	-0.4328	2.3103	2.3103	2.3103	-0.4328
2.3103	-0.4328	2.3103	2.3103	2.3103	2.3103
2.3103	2.3103	-0.4328	-0.4328	-0.4328	-0.4328
2.3103	2.3103	-0.4328	-0.4328	-0.4328	2.3103
2.3103	2.3103	2.3103	2.3103	2.3103	-0.4328
2.3103	2.3103	2.3103	2.3103	2.3103	2.3103
-0.4836	-0.5007	-0.4679	-0.4679	-0.4679	-0.4411
-0.4836	-0.5007	-0.4679	-0.4679	-0.4679	2.2671
-0.4836	-0.5007	2.1373	2.1373	2.1373	-0.4411
-0.4836	-0.5007	2.1373	2.1373	2.1373	2.2671
-0.4836	1.9972	-0.4679	-0.4679	-0.4679	-0.4411
-0.4836	1.9972	-0.4679	-0.4679	-0.4679	2.2671
-0.4836	1.9972	2.1373	2.1373	2.1373	-0.4411
-0.4836	1.9972	2.1373	2.1373	2.1373	2.2671
2.0677	-0.5007	-0.4679	-0.4679	-0.4679	-0.4411
2.0677	-0.5007	-0.4679	-0.4679	-0.4679	2.2671
2.0677	-0.5007	2.1373	2.1373	2.1373	-0.4411
2.0677	-0.5007	2.1373	2.1373	2.1373	2.2671
2.0677	1.9972	-0.4679	-0.4679	-0.4679	-0.4411
2.0677	1.9972	-0.4679	-0.4679	-0.4679	2.2671
2.0677	1.9972	2.1373	2.1373	2.1373	-0.4411
2.0677	1.9972	2.1373	2.1373	2.1373	2.2671

6.3.8 Aplicação do Modelo 11

Para as densidades a priori (4.9) escolhemos os seguintes valores: $a = 3$, $b = 4$ e os b_{1j} e c_{1j} , $j = 1, 2, 3, 4$ como sendo os mesmos usados no modelo 10.

Na tabela 6.33 apresentamos o sumário a posteriori para os parâmetros do modelo 5, a partir das amostras geradas das distribuições condicionais a posteriori. Para obtermos este sumário geramos 5 cadeias de 6000 iterações e consideramos uma amostra final de 1000 pontos amostrais, que foi suficiente para verificar a convergência observando que para todos os parâmetros $\sqrt{\widehat{R}} < 1.1$.

Tabela 6.33: Sumário à Posteriori considerando $\alpha_i \sim N(0, \sigma_\alpha^2)$

Parâmetro	Média	DP	IC 95%	$\sqrt{\widehat{R}}$	TAc
σ_α^2	1.0688	0.1390	(0.8163; 1.2866)	1.0553	1.0000
β_{11}	0.0034	0.1899	(-0.3839; 0.3583)	1.0148	0.5117
β_{12}	0.2969	0.1986	(-0.0930; 0.6460)	1.0158	0.5002
β_{13}	0.1391	0.1941	(-0.2427; 0.5039)	1.0126	0.5235
β_{14}	-0.1976	0.1859	(-0.5840; 0.1451)	1.0053	0.4471

A partir do modelo ajustado pelos parâmetros dados na tabela 6.33 calculamos os resíduos de Pearson d_{ij} (5.2) dados na tabela 6.34.

Table 6.34: Resíduos de Pearson ($\alpha_i \sim N(0, \sigma_\alpha^2)$)
 $i = 1, 2, \dots, 36$

J			
1	2	3	4
-0.0294	-0.0294	-0.0294	-0.0294
-0.5890	-0.5890	-0.5890	1.6978
-0.5823	-0.5823	1.7173	-0.5823
-0.9913	-0.9913	1.0088	1.0088
-0.5846	1.7107	-0.5846	-0.5846
-1.0077	0.9924	-1.0077	0.9924
-1.0054	0.9946	0.9946	-1.0054
-1.6286	0.6140	0.6140	0.6140
0.8729	-1.1456	-1.1456	-1.1456
0.9920	-1.0081	-1.0081	0.9920
0.9911	-1.0090	0.9911	-1.0090
0.6311	-1.5846	0.6311	0.6311
1.0014	1.0014	-0.9986	-0.9986
0.6319	0.6319	-1.5825	0.6319
0.5979	0.5979	0.5979	-1.6724
0.1137	0.1137	0.1137	0.1137
-0.0265	-0.0432	-0.0334	-0.0185
-0.5768	-0.7357	-0.6462	2.0626
-0.5636	-0.7195	1.5828	-0.4733
-0.9750	-1.2331	0.9203	1.2046
-0.5413	1.4444	-0.6074	-0.4539
-0.9785	0.8081	-1.0904	1.2003
-0.9607	0.8231	0.9340	-0.8179
-1.6251	0.4782	0.5486	0.7264
1.7620	-0.7244	-0.6361	-0.4768
1.0445	-1.2107	-1.0670	1.2269
1.0405	-1.2153	0.9337	-0.8182
0.6885	-1.8580	0.6153	0.8107
1.0450	0.8264	-1.0664	-0.8146
0.6511	0.5075	-1.7207	0.7675
0.6125	0.4759	0.5460	-1.3829
0.1496	0.1028	0.1263	0.1897

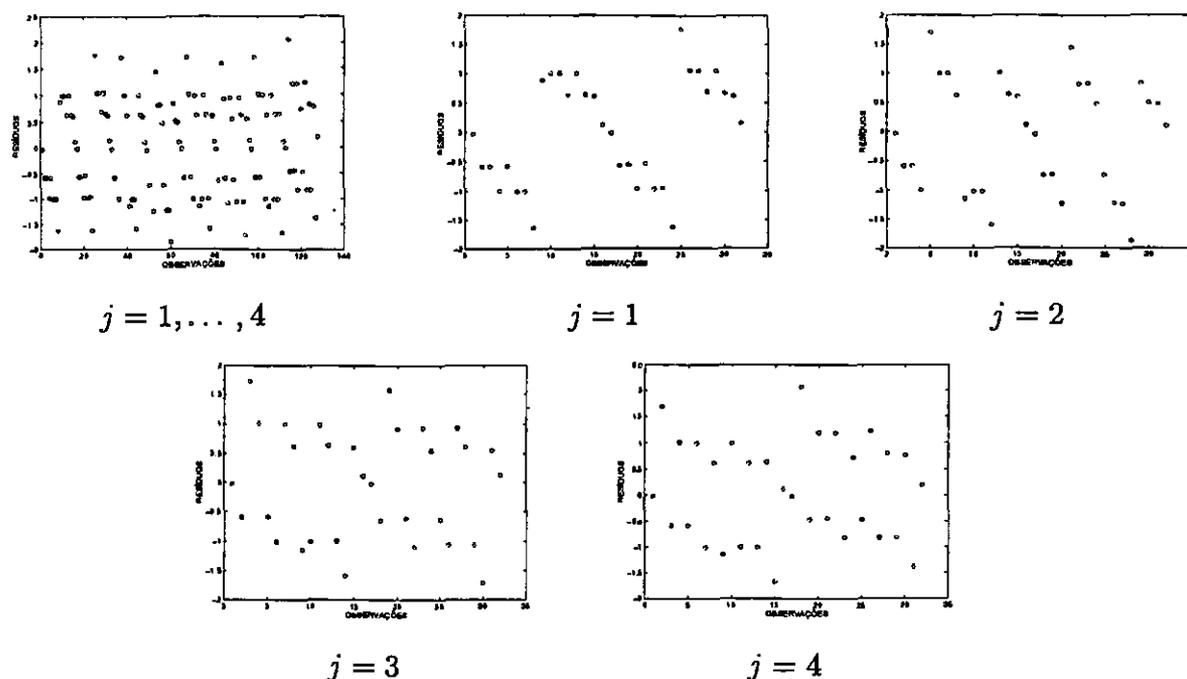


Figura 6.21: Resíduos de Pearson considerando $\alpha_i \sim N(0, \sigma_\alpha^2)$, $i = 1, 2, \dots, 32$

6.3.9 Aplicação do Modelo 12

Numa análise Bayesiana do modelo probito assumindo uma mistura de distribuições normais com réplicas para os efeitos aleatórios α_i , $i = 1, 2, \dots, 36$, consideramos os seguintes valores para as densidades a priori (4.11): $d_1 = -0.1$, $d_2 = 0.3$, $e_1 = e_2 = 0.2$, $a_1 = a_2 = 4$, $b_1 = b_2 = 3$, $f = 1$, $g = 1$ e os valores para b_{lj} e c_{lj} , $l = 1, 2, 3$; $j = 1, 2, \dots, 6$ os mesmos usados nos modelos 4 e 5.

Os resumos a posteriori são apresentados na tabela 6.35, a seguir.

Tabela 6.35: Sumário à Posteriori considerando α_i com Mistura de Normais

Parâmetro	Média	DP	IC 95%	$\sqrt{\hat{R}}$	TAc
μ_2	-0.5460	0.4709	(-1.5895; 0.2569)	0.9993	1.0000
μ_2	0.2107	0.3444	(-0.4219; 0.9332)	1.0022	1.0000
σ_1^2	1.0390	0.1449	(0.8085; 1.2871)	1.0023	1.0000
σ_2^2	1.0164	0.1431	(0.8070; 1.2822)	0.9993	1.0000
p_1	0.4360	0.2747	(0.0318; 0.9570)	0.9999	1.0000
β_{11}	0.0222	0.1868	(-0.3522; 0.3843)	1.0058	0.5091
β_{12}	0.3155	0.1868	(-0.0393; 0.6887)	1.0085	0.4730
β_{13}	0.1686	0.1871	(-0.2096; 0.5404)	1.0073	0.5096
β_{14}	-0.1679	0.1776	(-0.5121; 0.1715)	1.0056	0.4635

Na tabela 6.36 apresentamos o resíduo de Pearson d_{ij} .

Tabela 6.36: Resíduos de Pearson (mistura de duas distribuições normais para α_i)

J			
1	2	3	4
-0.0213	-0.0213	-0.0213	-0.0213
-0.5778	-0.5778	-0.5778	1.7306
-0.5846	-0.5846	1.7107	-0.5846
-0.9940	-0.9940	1.0060	1.0060
-0.5804	1.7228	-0.5804	-0.5804
-1.0080	0.9921	-1.0080	0.9921
-1.0022	0.9978	0.9978	-1.0022
-1.6277	0.6144	0.6144	0.6144
1.7178	-0.5822	-0.5822	-0.5822
1.0128	-0.9874	-0.9874	1.0128
1.0180	-0.9824	1.0180	-0.9824
0.6506	-1.5370	0.6506	0.6506
0.9985	0.9985	-1.0015	-1.0015
0.6433	0.6433	-1.5545	0.6433
0.5934	0.5934	0.5934	-1.6853
0.1077	0.1077	0.1077	0.1077
-0.0265	-0.0432	-0.0340	-0.0189
-0.5714	-0.7289	-0.6460	2.0629
-0.5486	-0.7011	1.6106	-0.4647
-0.9370	-1.1845	0.9495	1.2429
-0.5384	1.4522	-0.6096	-0.4557
-0.9424	0.8394	-1.0591	1.2358
-0.9513	0.8314	0.9352	-0.8170
-1.6089	0.4834	0.5492	0.7270
1.7914	-0.7129	-0.6315	-0.4732
1.0552	-1.1981	-1.0652	1.2287
1.0474	-1.2071	0.9319	-0.8200
0.6996	-1.8268	0.6198	0.8163
1.0600	0.8385	-1.0603	-0.8101
0.6490	0.5059	-1.7421	0.7584
0.6186	0.4810	0.5465	-1.3819
0.1419	0.0970	0.1179	0.1782

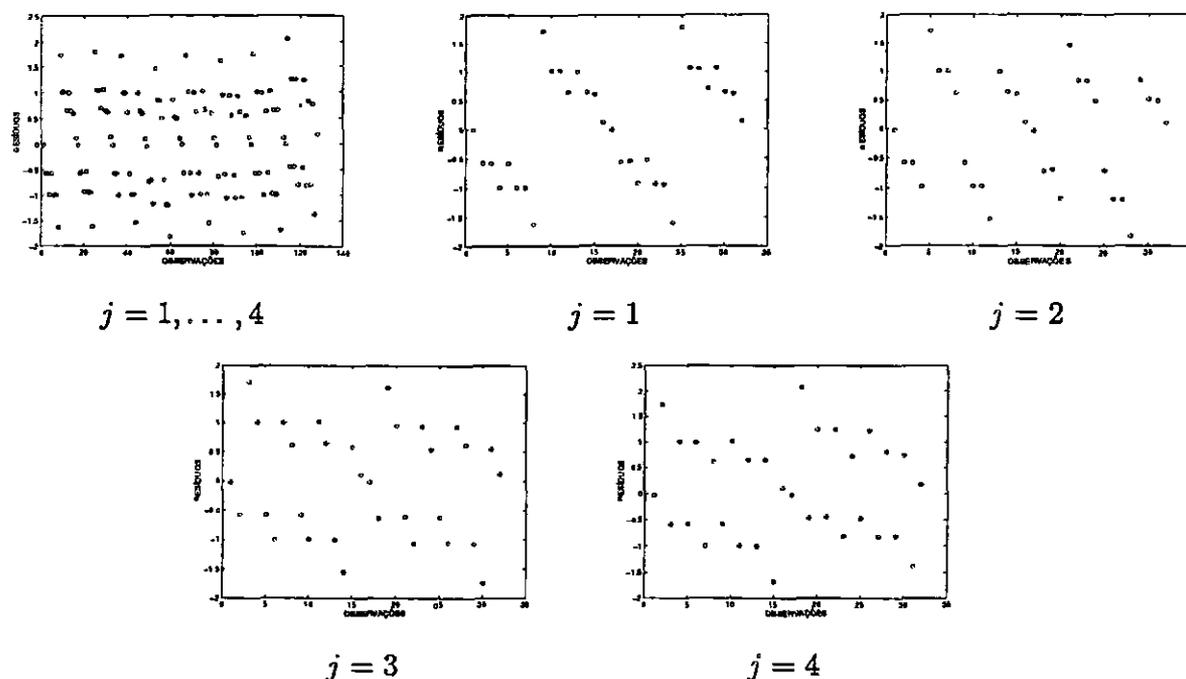


Figura 6.22: Resíduos de Pearson assumindo uma mistura de normais para os efeitos aleatórios

6.3.10 Comparação dos Modelos 10, 11 e 12

Na tabela 6.37 apresentamos os estimadores a posteriori para os parâmetros dos modelos 10, 11, 12 e os estimadores de máxima verossimilhança.

Tabela 6.37: Sumário à Posteriori para β_{1j} , $j = 1, 2, 3, 4$

Parâmetro	EMV	α fixo	$\alpha_i \sim N(0, \sigma_\alpha^2)$	α_i Mistura de Normais
β_{11}	0.1125	0.1241	0.0034	0.0222
β_{12}	0.2721	0.1636	0.2969	0.3155
β_{13}	0.1951	0.0867	0.1391	0.1686
β_{14}	-0.0011	0.0209	-0.1976	-0.1679

Na tabela 6.38, apresentamos o Total de Pearson (5.3) e as densidades preditivas calculadas nas observações para checagem da performance dos modelos.

Tabela 6.38: Resultados do Total de Pearson (6.4) e as Densidades Preditivas para os Modelos

	Modelo 10	Modelo 11	Modelo 12
D (32 obs)	327.7585	112.8056	112.0882
$\hat{p}_1(\hat{y} M_k)$ (32 obs)	–	$8.9319 \cdot 10^{-175}$	$1.0092 \cdot 10^{-174}$

Observamos que o Total de Pearson não revela diferença entre os modelos 11 e 12, mas as densidades preditivas apresentam uma grande diferença entre os modelos evidenciando uma melhor performance do modelo probito de efeitos aleatórios com uma mistura de duas distribuições normais.

A seguir temos nas figura 6.23 e 6.24, histogramas que mostram as frequências dos α_i , $i = 1, 2, \dots, 36$ para os modelos 5 e 6 respectivamente.

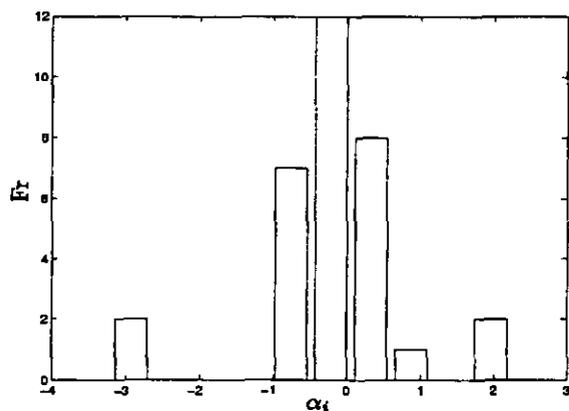


Figura 6.23: Frequências dos α_i ; Assumindo $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_\alpha^2)$ $i = 1, 2, \dots, 32$

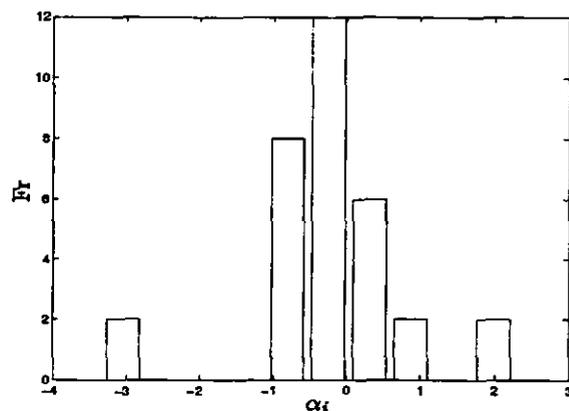


Figura 6.24: Frequências dos α_i ; Assumindo α_i , $i = 1, 2, \dots, 32$ Mistura de duas Normais

Capítulo 7

CONCLUSÕES E CONSIDERAÇÕES FUTURAS

Nesta dissertação, observamos que o uso de mistura de distribuições normais para modelos de efeitos aleatórios para dados binários correlacionados, considerando as funções de ligações logito e probito apresentam uma grande flexibilidade para o ajuste do modelo aos dados reais.

Usualmente, em experimentos envolvendo dados binários correlacionados, existem várias covariáveis de interesse, o que acarreta em um grande número de parâmetros e neste caso verificamos que o uso de métodos Bayesianos com técnicas de simulação de Monte Carlo em Cadeias de Markov (MCMC), são apropriadas para a obtenção de sumários a posteriori de interesse.

Fazendo uma comparação entre as funções de ligação logística e probito podemos verificar através do Total de Pearson e das densidades preditivas que a função de ligação logística apresentou uma melhor performance para os dados considerados entre os modelos de mesma ordem, tanto para dados replicados quanto para dados sem réplicas. Do ponto de vista computacional constatamos que a função de ligação probito apresenta maiores dificuldades em sua implementação pelo fato de convergir para 0 ou 1 (zero ou um) quando se trata valores absolutos menores que os necessários que isso ocorra na função de ligação

logística. Na figura 7.1, observamos esse comportamento.

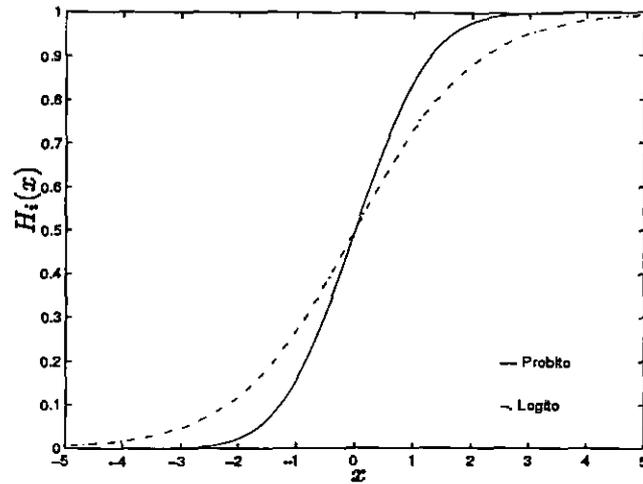


Figura 7.1: Comparação do Comportamento entre as Funções de Ligação Logito e Probit

Na figura 7.1 temos que $H_i(x)$, $i = 1, 2$, é dado por;

$$H_1(x) = \Phi(x) \text{ (Probit)}$$

$$H_2(x) = \frac{e^x}{1+e^x} \text{ (Logito)}$$

Para desenvolvimento futuro, podemos considerar o uso dos modelos condicionais e multivariados como uma ligação *t-de-Student* sob o enfoque Bayesiano, assumindo os aspectos computacionais considerados neste trabalho.

Apêndice A

ALGORITMO GIBBS SAMPLING

O algoritmo Gibbs sampling foi introduzido formalmente por Geman e Geman (1984), quando propuseram um esquema de amostragem de distribuições explorando as distribuições condicionais completas através de um algoritmo iterativo que define uma cadeia de Markov.

O algoritmo Gibbs sampling é essencialmente um esquema iterativo de amostragem de uma cadeia de Markov cujo núcleo de transição é formado pelas distribuições condicionais completas.

Descrição do método:

Suponha que a distribuição de interesse é $\pi(\theta)$ onde $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. Considere ainda que as densidades condicionais completas a posteriori $\pi_i(\theta_i | \theta_{-i})$, $i = 1, \dots, p$ estão disponíveis sob forma padronizada. Podemos então gerar as amostras como segue:

(i) inicializemos o contador de iterações da cadeia $j = 1$ e arbitremos valores iniciais $\theta^{(0)} = \theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)}$.

(ii) obtemos um novo valor $\theta^{(j)} = \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_p^{(j)}$ a partir de $\theta^{(j-1)}$ através de sucessivas gerações de valores

$$\theta_1^{(j)} \sim \pi(\theta_1 | \theta_2^{(j-1)}, \dots, \theta_p^{(j-1)})$$

$$\theta_2^{(j)} \sim \pi(\theta_2 | \theta_1^{(j)}, \theta_3^{(j-1)}, \dots, \theta_p^{(j-1)})$$

⋮

$$\theta_p^{(j)} \sim \pi(\theta_p | \theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_{p-1}^{(j)})$$

(iii) mudamos o contador j para $j + 1$ e retornamos a (ii) até a convergência.

Geman e Geman (1984) mostram que os valores gerados para θ_i , $i = 1, \dots, p$, convergem em distribuição para uma variável aleatória da distribuição $\pi(\theta)$ com $j \rightarrow \infty$.

Apêndice B

ALGORITMO DE METROPOLIS-HASTINGS

Este algoritmo é usado quando as densidades condicionais não são dadas sob forma padronizada conhecida, como de uma Normal, Gama ou outras.

Para descrevermos este método consideremos uma distribuição π , com densidades condicionais não padronizadas, da qual desejamos gerar uma amostra através de uma cadeia de Markov. Nesse caso, devemos construir um núcleo de transição $p(\theta, \phi)$ de forma que π seja a distribuição de equilíbrio da cadeia.

Uma forma simples de fazermos isso é através de cadeias onde o núcleo p satisfaça a condição de reversibilidade $\pi(\theta)p(\theta, \phi) = \pi(\phi)p(\phi, \theta)$, $\forall(\theta, \phi)$, sendo a condição suficiente para que π seja a distribuição de equilíbrio da cadeia.

O núcleo $p(\theta, \phi)$ é constituído de dois elementos um núcleo de transição $q(\theta, \phi)$ e uma probabilidade $\alpha(\theta, \phi)$ de forma que:

$$p(\theta, \phi) = q(\theta, \phi)\alpha(\theta, \phi), \quad \text{se } \phi \neq \theta.$$

Portanto, o núcleo de transição define uma densidade $p(\theta, \cdot)$ para todos os valores diferentes de θ . Consequentemente, resta uma probabilidade positiva da cadeia ficar em θ dada por,

$$p(\theta, \theta) = 1 - \int q(\theta, \phi)\alpha(\theta, \phi)\partial\phi.$$

Essas duas formas podem ser agrupadas na expressão geral:

$$p(\theta, A) = \int_A q(\theta, \phi)\alpha(\theta, \phi)\partial\phi + I(\theta \in A) \left[1 - \int q(\theta, \phi)\alpha(\theta, \phi)\partial\phi \right]. \quad (\text{B.1})$$

Logo, o núcleo de transição define uma distribuição mista para o novo estado ϕ da cadeia. Para $\phi \neq 0$, essa distribuição tem densidade e para $\phi = 0$, essa distribuição atribui uma probabilidade positiva.

A expressão mais comumente citada para a probabilidade de aceitação é

$$\alpha(\theta, \phi) = \min \left\{ 1, \frac{\pi(\phi)q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)} \right\}. \quad (\text{B.2})$$

Assim, a simulação de uma amostra π usando a cadeia de Markov definida pela transição (B.1) pode ser esquematizada da seguinte forma:

- (i) inicializamos o contador de iterações da cadeia $j = 1$ e arbitremos o valor inicial $\theta^{(0)}$;
- (ii) movemos a cadeia para um novo valor ϕ gerado da densidade $q(\theta^{(j-1)}, \cdot)$;
- (iii) calculamos a probabilidade de aceitação do movimento $\alpha(\theta^{(j-1)}, \phi)$ dada por (B.2). Se o movimento for aceito $\theta^{(j)} = \phi$. Se não for aceito, $\theta^{(j)} = \theta^{(j-1)}$ e a cadeia não se move;
- (iv) mudamos o contador de j para $j + 1$ e retornamos a (ii) até a convergência.

Realizamos a etapa (iii) após a geração de uma quantidade uniforme u independentes de todas as outras variáveis. Se $u \leq \alpha$, o movimento é aceito e se $u > \alpha$ o movimento não é permitido. O núcleo de transição q define uma proposta de movimento que pode ou não ser confirmado por α . O sucesso do método depende de taxas de aceitação não muito baixas e de propostas fáceis de simular. O método substitui uma geração difícil de π por várias gerações propostas de (q) .

Apêndice C

O CRITÉRIO DE CONVERGÊNCIA DE GELMAN E RUBIN (1992)

Esta é uma forma simples de monitorar a convergência dos valores gerados para uma distribuição de equilíbrio.

O critério de convergência de Gelman e Rubin (1992) é baseado em técnicas de análise de variância. Para o algoritmo devemos considerar pelo menos duas cadeias paralelas com valores iniciais, diferentes, amostrados de uma distribuição bem comportada. Após as cadeias atingirem estacionariedade digamos na t -ésima iteração, consideramos as realizações $(\theta_j, \theta_{j+h}, \dots, \theta_{j+nh})$, $j \geq t$ como uma amostra aleatória da distribuição desejada. Devemos assumir h razoavelmente grande de tal forma que 2 valores sucessivos de θ sejam independentes.

Se cada sequência tem tamanho $2n$, descartamos as primeiras n iterações e consideramos as n restantes. Para cada parâmetro de interesse, calculamos:

- a variância entre as cadeias:

$$E = \frac{n}{m-1} \sum_{i=1}^m (\bar{\theta}_i - \bar{\theta})^2;$$

- a variância dentro das cadeias:

$$D = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (\theta_i^{(j)} - \bar{\theta}_i)^2$$

onde: $\bar{\theta}_i$ é a média das observações da i -ésima cadeia; e

$\bar{\theta}$ é a média dos $\bar{\theta}_i$ com $i = 1, 2, \dots, m$.

Então, a variância da distribuição desejada pode ser estimada por $\hat{\sigma}^2 = \frac{n-1}{n}D + \frac{1}{n}E$ e a média por $\hat{\mu} = \bar{\theta}$. Podemos melhorar na estimativa mais precisa da distribuição desejada, admitindo para a variabilidade amostral das estimativas, μ e $\hat{\sigma}^2$. O resultado é uma distribuição t de Student aproximada para θ com centro $\hat{\mu}$, escala $\sqrt{\hat{V}} = \sqrt{\hat{\sigma}^2 + E/mn}$ e graus de liberdade $df = 2\hat{V}^2/\widehat{var}(\hat{V})$, onde

$$\begin{aligned} \widehat{var}(\hat{V}) = & \left(\frac{n-1}{n}\right)^2 \frac{1}{m} \widehat{var}(s_i^2) + \left(\frac{m+1}{mn}\right)^2 \frac{2}{m-1} E + \\ & + 2 \frac{(m+1)(n-1)m}{mn^2} \frac{1}{n} [\widehat{cov}(s_i^2, \bar{\theta}_i^2) - 2\bar{\theta} \widehat{cov}(s_i^2, \bar{\theta}_i)], \end{aligned}$$

onde as variâncias e covariâncias estimadas são obtidas dos m valores amostrais de $\bar{\theta}_i$ e s_i^2 ; $df \rightarrow \infty$ quando $n \rightarrow \infty$.

Monitoramos a convergência da simulação iterativa, calculando o fator de redução potencial de escala estimado $\sqrt{\hat{R}} = \sqrt{(\hat{V}/D)df/(df-2)}$. À medida que n cresce ($n \rightarrow \infty$), R converge para 1. \hat{R} é a razão do estimador da variância atual para a variância dentro da cadeia com um fator que avalia a variância adicional da distribuição 't de Student'.

Desde que \hat{R} esteja próximo de 1, admitimos que a convergência ocorreu e as amostras das iterações selecionadas são independentes e identicamente distribuídas (ver, Gelman & Rubin, 1992).

Apêndice D

PROGRAMA COMPUTACIONAL PARA O MODELO 3

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%  
% MCMC considerando o Modelo de efeitos aleat\'}rios com mistura,  
% de tres covariaveis sem interacao, para Dados de Huggis (1991),  
% dados Binarios Longitudinais com J=6 e n=36  
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%  
  
clear  
t0=clock;  
  
x11=[1 0 1 1 1 1 0 0 1 0 1 1 1 1 1 0 1 1 0 0 0 1 1 0 0 0 0 0 1 0 1 0 1  
1 1 0];  
x22=[1 1 1 1 1 1 1 3 1 1 3 1 2 1 1 3 1 1 1 3 1 1 3 3 1 1 3 3 2 1 1 1 3  
1 2 3];  
x3k=[12 15 15 15 13 13 5 20 12 6 21 11 15 14 13 22 14 11 10 23 7 8 19  
22 10 14 19 20 16 11 14 11 24 9 16 19];  
  
x33=x3k-14;  
  
t1=[1 1 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0  
0 0 0];  
t2=[1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0  
0 0 0];  
t3=[1 0 0 0 1 1 0 0 0 1 0 1 1 1 0 1 1 1 0 0 1 0 0 1 1 1 1 0 0 0 0 0  
0 0 0];  
t4=[1 1 0 1 1 0 1 0 0 0 1 1 0 1 1 0 1 1 1 0 0 0 1 1 0 1 0 1 1 1 0 0 0  
0 0 0];  
t5=[1 1 1 1 1 1 0 0 1 0 1 1 0 1 1 1 1 1 0 1 0 0 0 1 1 1 0 0 1 1 1 1  
0 0 0];  
t6=[1 1 1 1 0 1 1 1 0 1 0 1 1 1 1 1 0 0 0 1 1 1 1 1 1 0 0 1 0 0 0 0  
1 1 1];
```

```
x1=[x11' x11' x11' x11' x11' x11'];
x2=[x22' x22' x22' x22' x22' x22'];
x3=[x33' x33' x33' x33' x33' x33'];
t=[t1' t2' t3' t4' t5' t6'];

[lin,col]=size(t);

% Contadores
c(1:lin)=0.*ones(1,lin);
c1(1:col)=0.*ones(1,col);
c2(1:col)=0.*ones(1,col);
c3(1:col)=0.*ones(1,col);

m=5          % Numero de cadeias
int=30001    % Numero de iteracoes

% condicoes iniciais

al_o=linspace(-1,1,m);
B1_o=linspace(-5,4,m);
p1_o=linspace(.2,.8,m);
mi1_o=linspace(-.6,-.3,m);
mi2_o=linspace(.1,.8,m);
sig1_o=linspace(.02,.8,m);
sig2_o=sig1_o;

for i=2:int          % LOOP para iteracoes

f=1; g=1;           % f para p1; g para p1
a1=4; b1=3;         % priori p/ sig1
a2=a1; b2=b1;       % priori p/ sig2
d1=-0.4; e1=0.2;   % priori p/ mi1
d2=0.4; e2=0.2;    % priori p/ mi2

if rem(i-2,6000)==0
    alf(i-1,1:lin)=al_o(1+(i-2)/6000).*ones(1,lin);
    B1(i-1,1:col)=B1_o(1+(i-2)/6000).*ones(1,col);
    B2(i-1,1:col)=B1_o(1+(i-2)/6000).*ones(1,col);
    B3(i-1,1:col)=B1_o(1+(i-2)/6000).*ones(1,col);

    p1(i-1)=p1_o(1+(i-2)/6000);
    mi1(i-1)=mi1_o(1+(i-2)/6000);
    mi2(i-1)=mi2_o(1+(i-2)/6000);
    sig1(i-1)=sig1_o(1+(i-2)/6000);
    sig2(i-1)=sig2_o(1+(i-2)/6000);
end
end
```

```

%*****  CALCULOS BASICOS E GERACAO DA z1  *****

for is=1:lin

if abs(alf(i-1,is)-mi1(i-1)) < 20*sqrt(sig1(i-1))
    nor1=normpdf(alf(i-1,is),mi1(i-1),sqrt(sig1(i-1)));
else
    nor1=10(-6);
end

if abs(alf(i-1,is)-mi2(i-1)) < 20*sqrt(sig2(i-1))
    nor2=normpdf(alf(i-1,is),mi2(i-1),sqrt(sig2(i-1)));
else
    nor2=10(-6);
end

    nu(is)=p1(i-1)*nor1;
    de(is)=nu(is)+(1-p1(i-1))*nor2;
    h1(is)=nu(is)/de(is);
Z1(is)=binornd(1,h1(is));
end
    z1=Z1';
    z2=1-z1;
    v1=sum(z1);
    v2=sum(z2);

%***** Gibb's para p1 *****

p1(i)=betarnd(f+v1,g+v2);

m1=zeros(lin,col);
m2=zeros(lin,col);
m3=zeros(lin,col);

for is=1:lin

m1(is,:)=B1(i-1,:).*x1(is,:);
m2(is,:)=B2(i-1,:).*x2(is,:);
m3(is,:)=B3(i-1,:).*x3(is,:);

end

%***** Metropolis para alpha *****

for s=1:lin
    MU(s)=(z1(s)*mi1(i-1)*sig2(i-1)+z2(s)*mi2(i-1)*sig1(i-1))/...
        (z1(s)*sig2(i-1)+z2(s)*sig1(i-1));
    SIG(s)=(sig1(i-1)*sig2(i-1))/(z1(s)*sig2(i-1)+z2(s)*sig1(i-1));
end

```

```

all=normrnd(MU(s),sqrt(SIG(s)));

psio_al=exp(alf(i-1,s)*sum(t(s,:))-sum(log(1+exp(alf(i-1,s)+m1(s,:)...
      + m2(s,)+m3(s,)))));
psin_al=exp(all*sum(t(s,:))-sum(log(1+exp(all+m1(s,)+ m2(s,)+m3(s,)))));

p_all=min(psin_al/psio_al,1);

% Crit\ '{e}rio de aceitacao

u=rand;
if u < p_all
    alf(i,s)=all;
    c(s)=c(s)+1;
else
    alf(i,s)=alf(i-1,s);
end
end

al=[alf(i,:)'];

%***** Gibbs para mi1 *****

MI1=(d1*sig1(i-1)+e1*sum(al.*z1))/ ...
      (sig1(i-1)+e1*v1);
SG1=(e1*sig1(i-1))/(sig1(i-1)+e1*v1);

mi1(i)=normrnd(MI1,sqrt(SG1));

%***** Gibbs para mi2 *****

MI2=(d2*sig2(i-1)+e2*sum(al.*z2))/ ...
      (sig2(i-1)+e2*v2);
SG2=(e2*sig2(i-1))/(sig2(i-1)+e2*v2);

mi2(i)=normrnd(MI2,sqrt(SG2));

%***** Gibbs para sigma2_1 *****

Aa1=a1+v1/2;
Bb1=(1/b1+sum(z1.*(al-mi1(i)).^2)/2)^(-1);

sig1(i)=1/(gamrnd(Aa1,Bb1));

%***** Gibbs para sigma2_2 *****

Aa2=a2+v2/2;

```

```

Bb2=(1/b2+sum(z2.*(al-mi2(i)).^2)/2)^(-1);

sig2(i)=1/(gamrnd(Aa2,Bb2));

%***** Metropolis para b1 *****

bb11=[1.3 0.9 -0.6 0.06 1.3 0.4];

for s=1:col
    bn1=normrnd(bb11(s),0.2);
    m1(:,s)=B1(i-1,s).*x1(:,s);
    m2(:,s)=B2(i-1,s).*x2(:,s);
    m3(:,s)=B3(i-1,s).*x3(:,s);

    mn1(:,s)=bn1.*x1(:,s);

    psio_b1=exp(sum(m1(:,s).*t(:,s))-sum(log(1+exp(al+m1(:,s)+ ...
        m2(:,s)+m3(:,s)))));
    psin_b1=exp(sum(mn1(:,s).*t(:,s))-sum(log(1+exp(al+mn1(:,s)+ ...
        m2(:,s)+m3(:,s)))));

    p_b1=min(psin_b1/psio_b1,1);

% Crit\{e}rio de aceitacao

u=rand;
if u < p_b1
    B1(i,s)=bn1;
    c1(s)=c1(s)+1;
else
    B1(i,s)=B1(i-1,s);
end
end

%***** Metropolis para b2 *****

bb22=[-1.9 -1.2 -1 -1 -1.3 -0.7];

for s=1:col
    bn2=normrnd(bb22(s),0.2);

    m1(:,s)=B1(i,s).*x1(:,s);
    m2(:,s)=B2(i-1,s).*x2(:,s);
    m3(:,s)=B3(i-1,s).*x3(:,s);

    mn2(:,s)=bn2.*x2(:,s);

    psio_b2=exp(sum(m2(:,s).*t(:,s))-sum(log(1+exp(al+m1(:,s)+ ...

```

```

        m2(:,s)+m3(:,s)))));
psin_b2=exp(sum(mn2(:,s).*t(:,s))-sum(log(1+exp(al+m1(:,s)+ ...
        mn2(:,s)+m3(:,s))))));

p_b2=min(psin_b2/psio_b2,1);

% Crit\ '{e}rio de aceitacao

u=rand;
if u < p_b2
    B2(i,s)=bn2;
    c2(s)=c2(s)+1;
else
    B2(i,s)=B2(i-1,s);
end
end

%***** Metropolis para b3 *****

bb33=[0.3 0.1 0.08 0.1 0.3 0.1];

for s=1:col
    bn3=normrnd(bb33(s),0.2);

    m1(:,s)=B1(i,s).*x1(:,s);
    m2(:,s)=B2(i,s).*x2(:,s);
    m3(:,s)=B3(i-1,s).*x3(:,s);

    mn3(:,s)=bn3.*x3(:,s);

    psio_b3=exp(sum(m3(:,s).*t(:,s))-sum(log(1+exp(al+m1(:,s)+ ...
        m2(:,s)+m3(:,s))))));
    psin_b3=exp(sum(mn3(:,s).*t(:,s))-sum(log(1+exp(al+m1(:,s)+ ...
        m2(:,s)+mn3(:,s))))));

    p_b3=min(psin_b3/psio_b3,1);

% Crit\ '{e}rio de aceitacao

u=rand;
if u < p_b3
    B3(i,s)=bn3;
    c3(s)=c3(s)+1;
else
    B3(i,s)=B3(i-1,s);
end
end

```

```

end % p/ iteracoes

res=[mi1' mi2' sig1' sig2' p1' alf B1 B2 B3];

% Selecao das Amostras

g=1;
for j=3001:6000:30001
    for k=j:j+2999
        if rem(k-j+1,10)==0
            res_s(g,:)=res(k,:);
            g=g+1;
        end
    end
end
amostra=length(res_s(:,1))

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

save m3n36 res
save m3n36_s res_s

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Inferencias

M_res=mean(res_s);
S_res=std(res_s);
Me_res=median(res_s);
IC95_resin=prctile(res_s,2.5);
IC95_ressu=prctile(res_s,97.5);

%***** Analise de Convergencia Gelman & Rubin *****

[L,C]=size(res_s);

for j=1:C
    for s=1:m
        M(:,s)=res_s(1+((s-1)*(L))/m:(s*(L))/m,j);
    end
    GR(j)=cc_gr(M);
end

Tempo=etime(clock,t0)
Tempo
ta=[30000 30000 30000 30000 30000 c c1 c2 c3]./int;
ind=[[1:5] [1:36] [1:6] [1:6] [1:6]];
sol=[ind' M_res' S_res' IC95_resin' IC95_ressu' GR' ta']

```


Referências Bibliográficas

- [1] AMEMIYA, T. *Advance Econometrics*. Harvard University Press. Boston, 1985.
- [2] ASHFORD, J.R.; SOWDEN, R.R. Multivariate Probit Analysis. *Biometrics*, n.26, p.535-546, 1970.
- [3] ALBERT, I.; JAIS, J.P. Gibbs Sampler For The Logistic Model in the Analysis of Longitudinal Binary Data. *Statistics in Medicine*, n.17, p.2905-2921, 1998.
- [4] BICKEL, P.J.; DOKSUM, K.A. *Mathematical Statistics*. San Francisco, Holden-Day, 1977.
- [5] CHEN, M.H.; SCHMEISER, B. Performance of the Gibbs, Hit-and-Run, and Metropolis Samplers. *The Journal of Computational and Graphical Statistics*, n.2, p.251-272, 1993.
- [6] COX, D.R. *The Analysis of Binary Data*. London, Chapman and Hall, 1970.
- [7] CHIB, S.; GREENBERG, E. Bayesian Analysis of Multivariate Probit Models. *Technical Report, Olin School of Business, Washington University*. 1995.
- [8] CHIB, S.; GREENBERG, E. Analysis of Multivariate Probit Models. *Biometrika*, n.85, p.347-361, 1998.
- [9] DEY, D.K.; GELFAND, A.E.; SWARTZ, T.B.; VLACHOS, P.K. Simulation Based Model Checking for Hierarchical Models. *Technical Report*, s.n., 95-29, 1995. (Department of Statistics, University of Connecticut).
- [10] DEY, D.K.; CHEN, M.-H. Bayesian Analysis of Correlated Binary Data Models. *Technic Report, Dep. of Statistics, University of Connecticut, U.S.A*, s.d., 1996.
- [11] GELFAND, A.E.; SMITH, A.F.M. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, n.85, p.398-409, 1990.
- [12] GELFAND, A.E.; DEY, D.K.; CHANG, H. Model Determining Using Predictive Distributions with Implementation Via Sampling-Based Methods. *Proceedings of the Fourth Valencia International Meeting on Bayesian Statistics, Oxford University Press*, p.147-167, 1992.
- [13] GELMAN, A.; RUBIN, B.D. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, n.4, p.457-511, 1992.

- [14] GEMAN, S.; GEMAN, D. Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, n.6, p.721-741, 1984.
- [15] GOURIEROUX, C.; MONFORT, A. *Statistics and Econometric Models*. England, Cambridge University Press, 1995.
- [16] HASEMAN, J.K.; KUPPER, L.J. Analysis of Dichotomus Response Data from Certain Toxicological Experimentes. *Biometrics*, n.35, p.281-293, 1979.
- [17] HOSMER, D.W.; LEMESHOW, S. *Applied Logistic Regression*. New York, John Wiley and Sons, 1989.
- [18] HUGGINS, R.M. 'Some Practical' Aspects of a Conditional Likelihood Approach to Capture Experiments. *Biometrics*, n.47, p.725-732, 1991.
- [19] KASS, R.E. e RAFTERY, A.E. Bayes Factors. *Journal of the American Statistical Association*, n.90, p.773-795, 1995.
- [20] LIANG, K.Y.; ZEGER, S.L. Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, n.73, p.13-22, 1986.
- [21] NEWTON, M.A.; RAFTERY, A.E. Approximate Bayesian Inference by the Weighted Likelihood Bootstrap. *Journal of the Royal Statistical Society, series B*, n.56, p.3-48, 1994.
- [22] OCHI, Y. e PRENTICE, R.L. Likelihood Inference in a Correlated Probit Regression Model. *Biometrics*, n.71, p.521-543, 1984.
- [23] PRENTICE, R.L. Correlated Binary Regression with Covariate Specific to Each Binary Observation. *Biometrics*, n.44, p.1033-1048, 1988.
- [24] RAFTERY, A.E. Hypothesis Testing and Model Selection. In *Markov Chain Monte Carlo in practice*, W. Gilks, S. Richardson and D.J. Spiegelhalter, eds., London, Chapman and Hall, p.163-187, 1996.
- [25] RAO, C.R. *Linear Statistical Inference and its Applications*. 2.ed. New York, Wiley, 1973.
- [26] SOARES, J.F.; COLOSIMO, E.A. *Métodos Estatísticos na Pesquisa Clínica*. Ribeirão Preto, SP, USP, 1995.
- [27] TANNER, M.; WONG, W. The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, n.82, p.528-550, 1987.
- [28] ZEGER, S.L.; LIANG, K.Y. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, n.42, p.121-130, 1986.