

Uso de Métodos Clássicos e Bayesianos em Modelos de Regressão Beta

Diomedes Pael Reitman

Orientador: Prof. Dr. Carlos Alberto Ribeiro Diniz

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar, como parte dos requisitos para obtenção do título de Mestre em Estatística.

São Carlos
Maio - 2007

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

R379um

Reitman, Diomedes Pael.

Uso de métodos clássicos e bayesianos em modelos de regressão beta / Diomedes Pael Reitman. -- São Carlos : UFSCar, 2007.

160 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2007.

1. Análise de regressão. 2. Inferência bayesiana. 3. MCMC. I. Título.

CDD: 519.536 (20ª)

Aos meu pais, que com muito sacrifício e persistência colocaram-me no caminho do saber.

À minha esposa Aline, que nas horas críticas me impulsionou e me ensinou que o amanhã se conquista hoje e, principalmente por se manter sempre forte e zelar pela saúde e bem-estar de nossa filha, nos momentos que estive ausente.

À Stefany (minha “Tetéia”), que alegra minha vida e me fez acreditar que os limites individuais são apenas obstáculos a serem superados continuamente em nossa vida.

"A vitória pertence àquele que for mais perseverante".

Roque Schneider

"Para o triunfo do mal,
Basta que os bons não façam nada".

Edmund Burke

Agradeço,

Primeiramente a DEUS, por permitir que eu conquistasse mais essa etapa.

Ao meu pai, Diomedes, e à minha querida mãe, Leane, por todo o incentivo que sempre me deram, sem o qual eu não teria conseguido esta etapa tão importante.

Um agradecimento muito especial à Aline de Almeida Reitman e Stefany de Almeida Reitman, pelo amor, carinho, compreensão e constante apoio nas horas mais difíceis.

Aos meus familiares, que mesmo à distância, sempre me incentivaram.

A amiga e eterna Professora, Palmira, por ter me descoberto na infância e me incentivado, em todos os sentidos.

Aos professores do Departamento de Estatística da UFSCar, por suas contribuições a minha formação como Estatístico.

A Carlos Alberto Ribeiro Diniz, Professor Doutor do Departamento de Estatística da Universidade de São Carlos – UFSCar, pela orientação, pela amizade e confiança depositada.

Ao Governo do Estado de São Paulo, pelos recursos financeiros do programa Bolsa – Mestrado, disponibilizados desde o início dos meus estudos de pós – graduação, sem os quais este trabalho não seria concretizado.

Finalmente, aos participantes da banca examinadora por suas valiosas contribuições.

Resumo

Este trabalho compreende um estudo de um modelo de Regressão Beta adequado para situações em que a variável resposta é medida de forma contínua no intervalo $(0, 1)$ como, por exemplo, dados de taxas ou proporções. As inferências desenvolvidas foram baseadas nas metodologias Clássica e Bayesiana. É apresentada uma discussão ampla sobre a aplicação do modelo de regressão beta a conjuntos de dados reais, o caso *Charter Schools*.

Palavras-chave: Regressão beta, Distribuição beta, MCMC.

Abstract

This work involves a study of a regression model appropriated for situations which the response variable is measured in a continuous scale in the $(0, 1)$ interval, as, for instance, taxes or proportions. The developed inferences were based on the Classic and Bayesian methodology. A discussion about the application of the beta regression model is presented.

Key-words: Beta regression model, Beta distribution, MCMC.

Sumário

| | | |
|----------|--|----------|
| 1 | Introdução | 1 |
| 1.1 | Organização da Dissertação | 2 |
| 2 | Família de Distribuições Beta | 3 |
| 2.1 | Introdução | 3 |
| 2.2 | Estimação dos Parâmetros | 5 |
| 2.2.1 | Método de Máxima Verossimilhança | 5 |
| 3 | O Modelo de Regressão Beta sob o enfoque Clássico | 8 |
| 3.1 | Introdução | 8 |
| 3.2 | Funções de Ligação | 10 |
| 3.3 | Estimação | 11 |
| 3.4 | O modelo reparametrizado | 12 |
| 3.5 | Função Escore e Matriz de Informação | 14 |
| 3.6 | Estimação dos Parâmetros via Função Escore | 18 |
| 3.6.1 | Processo Iterativo de Newton – Raphson | 18 |
| 3.6.2 | Processo Iterativo de Escore de Fisher | 19 |
| 3.6.3 | Método BFGS | 19 |
| 3.7 | Testes de Hipóteses | 20 |
| 3.7.1 | Teste da Razão de Verossimilhança | 20 |
| 3.7.2 | Teste Escore | 20 |
| 3.7.3 | Teste de Wald | 21 |
| 3.8 | Técnicas de Diagnóstico | 22 |
| 3.8.1 | Resíduo Componente do Desvio | 23 |

| | | |
|----------|--|-----------|
| 3.8.2 | Resíduos Ponderados e Resíduos Padronizados Ponderados | 24 |
| 3.8.3 | Gráficos de Probabilidade Meio – normal com Envelopes | 26 |
| 3.8.4 | <i>Leverage</i> Generalizado | 27 |
| 3.8.5 | Influência | 29 |
| 4 | Avaliação de Resultados Assintóticos em Amostras Finitas | 30 |
| 4.1 | Introdução | 30 |
| 4.2 | Detalhes Metodológicos | 30 |
| 4.3 | Comportamento dos Estimadores | 31 |
| 4.4 | Tamanho dos Testes | 34 |
| 5 | Regressão Beta: Procedimentos Práticos em Estimação | 37 |
| 5.1 | Procedimentos Gerais | 37 |
| 5.2 | Implementação em Pacotes Estatísticos | 39 |
| 6 | Modelo de Regressão Beta sob o Enfoque Bayesiano | 41 |
| 6.1 | Abordagem Bayesiana do Modelo | 41 |
| 6.2 | Informação <i>a priori</i> | 41 |
| 6.3 | Função de Verossimilhança | 42 |
| 6.3.1 | Estimação Bayesiana com Prioris vagas | 45 |
| 6.3.2 | Estimação Bayesiana com Prioris Informativas | 49 |
| 7 | Aplicação: O Caso <i>Charter Schools</i> | 54 |
| 7.1 | Introdução | 54 |
| 7.2 | Estatística e Marketing: O Marketing na Educação | 54 |
| 7.3 | Inferência Clássica | 57 |
| 7.4 | Estimação dos Parâmetros | 58 |
| 7.4.1 | Modelo 1 | 58 |
| 7.4.2 | Modelo 2 | 63 |
| 7.4.3 | Modelo 3 | 68 |
| 7.4.4 | Modelo 4 | 72 |
| 7.4.5 | Modelo 5 | 75 |
| 7.5 | Comparação entre os cinco modelos | 80 |

| | | |
|----------|--|------------|
| 7.5.1 | Comparações entre o modelo de regressão Beta ajustados com a função logito e com um modelo de regressão linear | 81 |
| 7.6 | Simulações | 82 |
| 7.7 | Inferência Bayesiana | 83 |
| 7.8 | Estimação dos Parâmetros via Métodos Bayesianos | 85 |
| 7.8.1 | Modelo 6 | 86 |
| 7.8.2 | Modelo 7 | 94 |
| 7.8.3 | Modelo 8 | 104 |
| 8 | Considerações finais | 119 |
| | Referências Bibliográficas e Apêndices | 121 |
| .1 | Referências | 121 |
| .2 | Apêndices | 124 |
| .2.1 | Apêndice A: Programa em R para avaliação das qualidades assintóticas dos estimadores | 124 |
| .2.2 | Apêndice B: Programa em R para inferência Clássica usando as funções de ligação Logito e Probita | 137 |
| .2.3 | Apêndice C: Análise Clássica em SAS | 156 |
| .2.4 | Apêndice D: Programas em WinBugs para a Análise Bayesiana | 157 |
| .2.5 | Apêndice E: Programas em R para calcular o BIC | 160 |

Capítulo 1

Introdução

Em algumas situações da modelagem estatística a variável resposta é uma variável contínua e restrita ao intervalo $(0,1)$. Nesses casos o modelo de regressão usual pode não ser apropriado uma vez que é possível encontrar valores ajustados excedendo o limite inferior ou superior do intervalo de valores para a variável resposta. Uma alternativa de solução seria transformar a variável resposta de tal forma que esta assuma valores em toda a reta. Em seguida, modelamos a média da resposta transformada. No entanto, esse enfoque apresenta algumas desvantagens, como, por exemplo, o fato que os parâmetros do modelo não podem ser facilmente interpretados em termos da resposta original (ver Paolino (2001)), dependendo da transformação.

Ferrari e Cribari-Neto (2004) propuseram um modelo de regressão para situações em que a variável resposta, Y , é medida de forma contínua no intervalo $(0,1)$, isto é, $0 < Y < 1$. O modelo proposto é baseado na suposição que a resposta tem distribuição beta, utilizando uma parametrização da família beta, que é indexada pela média e pela dispersão.

Em nosso trabalho, o interesse reside em situações em que se deseja modelar proporções ou taxas, através de uma estrutura de regressão, como uma função de um conjunto de covariáveis. E como se sabe, a distribuição beta é muito versátil e flexível para modelar proporções, pois sua densidade pode assumir diferentes formas dependendo dos valores dos dois parâmetros que indexam a distribuição.

O principal objetivo deste trabalho é discutir o modelo de regressão beta no que diz respeito à sua definição, resultados de inferências, sob os enfoques clássico e bayesiano,

e aplicações a dados reais. Adicionalmente, apresentamos um estudo de simulação que procura avaliar a qualidade das aproximações utilizadas nas inferências sobre os parâmetros em amostras finitas e comparar as duas metodologias utilizadas.

1.1 Organização da Dissertação

A presente dissertação de mestrado está dividida em oito capítulos. No segundo capítulo desenvolvemos um estudo sobre a família de distribuições beta. No terceiro capítulo apresentamos um estudo sobre o modelo de regressão beta sob o enfoque clássico. Neste capítulo discutimos algumas funções de ligação, a função escore e a matriz de informação, a estimação dos parâmetros de regressão, testes de hipóteses, intervalos de confiança e algumas medidas de diagnóstico, propostos em Ferrari e Cribari-Neto (2004).

No quarto avaliamos, através de simulações de Monte Carlo, os resultados assintóticos em amostras finitas.

No quinto capítulo apresentamos alguns procedimentos computacionais de estimação para o modelo de regressão beta considerando vários pacotes estatísticos.

No sexto capítulo discutimos um estudo sobre o modelo de regressão beta sob o enfoque bayesiano, envolvendo prioris vagas e informativas, estimação dos parâmetros do modelo de regressão, testes de hipóteses, intervalos de credibilidade e algumas medidas de diagnóstico.

No sétimo capítulo ilustramos as metodologias desenvolvidas nos capítulos anteriores através da análise de um conjunto de dados reais e comparamos os modelos ajustados. O conjunto de dados se refere a um estudo da participação de mercado das "*Charter Schools*" em todos os estados dos Estados Unidos.

Finalmente, no oitavo capítulo apresentamos algumas considerações finais.

Capítulo 2

Família de Distribuições Beta

2.1 Introdução

Neste capítulo discutimos um estudo sobre a distribuição beta envolvendo definições, propriedades, inferências e aplicações, conforme as propostas de Ferrari e Cribari-Neto (2004).

A família de distribuições beta é composta de todas as distribuições com função densidade de probabilidade (f.d.p.) da forma:

$$f(y, p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{(p-1)} (1-y)^{(q-1)}, \quad 0 < y < 1, \quad (2.1)$$

sendo, $p > 0$, $q > 0$ e $\Gamma(p)$ é a função gama avaliada no ponto p , ou seja,

$$\Gamma(p) = \int_0^{\infty} y^{(p-1)} e^{-y} dy. \quad (2.2)$$

A distribuição beta é um modelo probabilístico importante para variáveis aleatórias que tem seus valores restritos ao intervalo $(0, 1)$.

Este modelo não apresenta a estrutura dos modelos de locação – escala, pois ambos p e q são parâmetros de forma. Um fato interessante da distribuição beta é que

$$Y \sim Beta(p, q) \rightarrow (1 - Y) \sim Beta(q, p). \quad (2.3)$$

Além disso, se $Y_1 \sim Gama(p, b)$, $Y_2 \sim Gama(q, b)$ e Y_1 e Y_2 são independentes, então

$Y_1/(Y_1 + Y_2) \sim \text{Beta}(p, q)$.

A função de distribuição acumulada (f.d.a.) de uma variável aleatória com distribuição $\text{Beta}(p, q)$ é dada por:

$$F(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^y z^{p-1} (1-z)^{q-1} dz, \quad 0 < y < 1. \quad (2.4)$$

A inversa da razão de funções gamas na expressão (2.4),

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)},$$

é chamada função beta e a integral em (2.4),

$$B_y(y; p, q) = \int_0^y z^{p-1} (1-z)^{q-1} dz,$$

é denominada função beta incompleta. Note que $B_y(p, q) \rightarrow B(p, q)$, quando $y \rightarrow 1$. A f.d.a. de Y pode assim ser expressa como a razão da função beta incompleta pela função beta, ou seja, $f(y; p, q) = B_y(y; p, q) / B(p, q)$ e por (2.5) segue que, $f(y; p, q) = f(1-y; q, p)$.

O r -ésimo momento populacional de Y em relação à origem é dado por,

$$\mu_r^k(Y) = \frac{B(p+r, q)}{B(p, q)} = \frac{\Gamma(p+r)\Gamma(p+q)}{\Gamma(p)\Gamma(p+q+r)} = \frac{p^{[r]}}{(p+q)^{[r]}},$$

se r é um valor inteiro, em que, $x^{[r]} = x(x+1)\dots(x+r-1)$ é o fatorial crescente. Em particular, o valor esperado da variável aleatória Y é dado por

$$E(Y) = \frac{p}{p+q}, \quad (2.5)$$

e sua variância é

$$\text{Var}(Y) = \frac{pq}{(p+q)^2(p+q+1)}. \quad (2.6)$$

O coeficiente de variação de Y é dado por

$$cv = \frac{\sqrt{\text{Var}(Y)}}{E(Y)} = \sqrt{\frac{q}{p(p+q+1)}}.$$

Essa medida caracteriza a dispersão da distribuição de Y em termos relativos à sua média.

A distribuição beta possui um valor modal quando $p + q > 2$ e $p \geq 1$ dado por

$$y_m = \frac{p - 1}{p + q - 2}.$$

O coeficiente de assimetria é dado por

$$y_1 = \frac{2(q - p)}{p + q + 2} \sqrt{\frac{p + q + 1}{pq}}.$$

Esta medida caracteriza como e quanto a distribuição se afasta da condição de simetria .

O coeficiente de curtose é dado por

$$y_2 = \frac{3(p + q + 1) [2(p + q)^2 + pq(p + q - 6)]}{pq(p + q + 2)(p + q + 3)},$$

e é uma medida que procura caracterizar o formato da distribuição quanto ao seu achatamento.

2.2 Estimação dos Parâmetros

A estimação dos parâmetros para a distribuição beta pode ser feita através do método de máxima verossimilhança, em que os estimadores são obtidos a partir da maximização da função de verossimilhança.

2.2.1 Método de Máxima Verossimilhança

Seja Y_1, \dots, Y_n uma amostra aleatória de tamanho n da variável aleatória Y com função densidade (ou de probabilidade) beta definida em (2.1), com $\theta = (p, q) \in \Theta = \mathbb{R}_+ \times \mathbb{R}_+$, sendo Θ o espaço paramétrico. A função de verossimilhança de θ correspondente à amostra observada y_1, \dots, y_n é dada por

$$L(\theta) = L(p, q) = \prod_{k=1}^n \frac{y_k^{(p-1)} (1 - y_k)^{(q-1)}}{B(p, q)} = [B(p, q)]^{-n} \prod_{k=1}^n y_k^{(p-1)} \prod_{k=1}^n (1 - y_k)^{(q-1)}. \quad (2.7)$$

Aplicando o logaritmo natural em ambos os lados da equação (2.7), propriedade que

não muda o resultado do EMV, temos

$$l(p, q) = \ln L(p, q) = -n \ln B(p, q) + (p-1) \sum_{k=1}^n \ln(y_k) + (q-1) \sum_{k=1}^n \ln(1-y_k).$$

As equações de máxima verossimilhança são obtidas derivando $l(p, q)$ em relação a cada parâmetro e igualando o resultado a zero. Assim, temos as equações

$$\Psi(\hat{p}) - \Psi(\hat{p} + \hat{q}) = \frac{1}{n} \sum_{k=1}^n \ln(y_k) \quad (2.8)$$

e

$$\Psi(\hat{q}) - \Psi(\hat{p} + \hat{q}) = \frac{1}{n} \sum_{k=1}^n \ln(1-y_k), \quad (2.9)$$

em que $\Psi(\lambda)$ é a função digama definida como

$$\Psi(\lambda) = \frac{d \ln \Gamma(\lambda)}{d\lambda} = \frac{\Gamma'(\lambda)}{\Gamma(\lambda)}, \quad \lambda > 0.$$

Soluções analíticas explícitas das equações de máxima verossimilhança não podem ser obtidas para os parâmetros da distribuição beta. Neste caso, os estimadores de máxima verossimilhança são facilmente obtidos por meio de métodos numéricos de maximização, como por exemplo, o método score de Fisher.

Ferrari e Cribari-Neto (2004), propuseram iniciar os processos iterativos utilizando como valores iniciais para \hat{p} e \hat{q} os valores

$$\hat{p} = \frac{1}{2} \left\{ \frac{1 - \prod_{k=1}^n (1-y_k)^{\frac{1}{n}}}{1 - \prod_{k=1}^n (y_k)^{\frac{1}{n}} - \prod_{k=1}^n (1-y_k)^{\frac{1}{n}}} \right\}$$

e

$$\hat{q} = \frac{1}{2} \left\{ \frac{1 - \prod_{k=1}^n (y_k)^{\frac{1}{n}}}{1 - \prod_{k=1}^n (y_k)^{\frac{1}{n}} - \prod_{k=1}^n (1-y_k)^{\frac{1}{n}}} \right\}.$$

Estes valores iniciais nos permitem encontrar soluções de (2.8) e (2.9) com uma con-

vergência bastante rápida.

Sob condições gerais de regularidade (ver Mood *et. al*), quando o tamanho da amostra é grande, temos que:

$$\begin{pmatrix} \hat{p} \\ \hat{q} \end{pmatrix} \stackrel{a}{\approx} N_2 \left(\begin{pmatrix} p \\ q \end{pmatrix}; K^{-1}(p, q) \right),$$

onde \hat{p} e \hat{q} são os estimadores de máxima verossimilhança para p e q , respectivamente. A matriz de variâncias e covariâncias assintótica dos estimadores de máxima verossimilhança é dada por:

$$K^{-1}(p, q) = n^{-1} [\Psi'(p) \Psi'(q) - \Psi'(p+q) \{\Psi'(p) + \Psi'(q)\}]^{-1} \\ \times \begin{bmatrix} \Psi'(q) - \Psi'(p+q) & \Psi'(p+q) \\ \Psi'(p+q) & \Psi'(p) - \Psi'(p+q) \end{bmatrix},$$

em que $\Psi'(\lambda)$ é a função trigama, definida por

$$\Psi'(\lambda) = \frac{d\Psi(\lambda)}{d\lambda} = \frac{d^2 \ln \Gamma(\lambda)}{d\lambda^2} = \frac{\Gamma''(\lambda) \Gamma(\lambda) - [\Gamma'(\lambda)]^2}{[\Gamma(\lambda)]^2}, \quad \lambda > 0.$$

Desta forma, os estimadores de máxima verossimilhança são assintoticamente não-viciados e consistentes.

É importante notar que \hat{p} e \hat{q} dependem dos dados apenas através de estatísticas suficientes (ver Mood *et. al*)

$$\left(\prod_{k=1}^n (y_k), \prod_{k=1}^n (1 - y_k) \right),$$

ou equivalentemente,

$$\left(\sum_{k=1}^n \ln(y_k), \sum_{k=1}^n \ln(1 - y_k) \right).$$

Capítulo 3

O Modelo de Regressão Beta sob o enfoque Clássico

3.1 Introdução

A densidade da distribuição beta definida em (2.1) é parametrizada por dois parâmetros p e q . Mas, em muitas situações práticas é comum que estes dois parâmetros sejam funções de outros parâmetros e de covariáveis. Como, por exemplo, a taxa de participação de mercado de uma empresa em um determinado segmento pode seguir distribuição beta com parâmetros p e q com p e q sendo função direta do número de consumidores da região, da renda per capita, do número de concorrentes entre outros. Neste caso a presença das covariáveis podem interferir fortemente na determinação dos parâmetros p e q . Assim, para uma análise de regressão é usual definir um modelo que contenha um parâmetro de precisão (dispersão) e modelar a média da resposta em termos de outros parâmetros e de variáveis explicativas.

Com o objetivo de obter uma estrutura de regressão para a média da resposta juntamente com um parâmetro de precisão, uma parametrização diferente da densidade beta foi sugerida por Ferrari e Cribari–Neto (2004). Sejam

$$\mu = \frac{p}{p + q} \tag{3.1}$$

e

$$\phi = p + q, \quad (3.2)$$

isto é,

$$p = \mu\phi$$

e

$$q = (1 - \mu)\phi.$$

Segue de (2.5) e (2.6) que:

$$E(Y) = \mu$$

e

$$Var(Y) = \frac{V(\mu)}{1 + \phi},$$

em que, $V(\mu) = \mu(1 - \mu)$, de tal forma que μ é a média da variável Y e ϕ pode ser interpretado como um parâmetro de precisão no sentido que, para μ fixado, quanto maior o valor de ϕ , menor a variância de Y . A densidade de Y pode ser escrita, na nova parametrização como

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{(\mu\phi-1)} (1-y)^{[(1-\mu)\phi-1]}, \quad 0 < y < 1, \quad (3.3)$$

em que $0 < \mu < 1$ e $\phi > 0$.

Assumimos que a variável resposta esteja restrita ao intervalo unitário padrão $(0, 1)$. No entanto, o modelo acima pode ser utilizado em situações mais gerais, em que a resposta é restrita ao intervalo (a, b) e a e b são constantes conhecidas, $a < b$. Neste caso, modela-se $\frac{Y-a}{b-a}$ em vez de modelar Y diretamente.

Sejam Y_1, \dots, Y_n variáveis aleatórias independentes, em que cada Y_k , $k = 1, \dots, n$, tem densidade como em (3.3), com média μ_k e parâmetro de precisão desconhecido ϕ . O modelo de regressão beta é definido por (3.3) e pelo componente sistemático:

$$g(\mu_k) = \sum_{i=1}^n x_{ik}\beta_i = x_k^T \boldsymbol{\beta} = \eta_k,$$

sendo $\beta = (\beta_1, \beta_2, \dots, \beta_k)$ um vetor de parâmetros de regressão desconhecidos e $x_k^T = (x_{k1}, \dots, x_{kp})^T$ um vetor de observações em k covariáveis ($k < n$) que são assumidas conhecidas. Finalmente, $g(\cdot)$ é uma função de ligação estritamente monótona e duas vezes diferenciável em $(0, 1)$.

3.2 Funções de Ligação

Existem algumas possíveis escolhas para a função de ligação $g(\cdot)$. Por exemplo, podemos utilizar a especificação logito:

$$g(\mu) = \log \left\{ \frac{\mu}{1 - \mu} \right\},$$

ou a função probito:

$$g(\mu) = \Phi^{-1}(\mu),$$

em que $\Phi(\cdot)$ é a função de distribuição acumulada de uma variável aleatória normal padrão, ou ainda a ligação complemento log-log:

$$g(\mu) = -\log\{-\log(\mu)\},$$

entre outras. Todas essas três funções de ligação são contínuas e estritamente crescentes em $(0, 1)$.

Com a função de ligação logito, temos, para o k -ésimo indivíduo,

$$\begin{aligned} g(\mu_k) &= \log \left\{ \frac{\mu_k}{1 - \mu_k} \right\} = x_k^T \beta \Rightarrow \\ &\Rightarrow \frac{\mu_k}{1 - \mu_k} = \exp(x_k^T \beta) \Rightarrow \\ &\Rightarrow \mu_k = \frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)}, \end{aligned} \tag{3.4}$$

A expressão (3.4) é a função inversa de $g(\mu_k)$.

3.3 Estimação

Paolino (2001) mostrou que uma aproximação normal padrão, para casos em que a resposta está restrita ao intervalo unitário $(0, 1)$, e a estimação dos parâmetros feita via mínimos quadrados podem, geralmente, conduzir a uma avaliação dos efeitos das covariáveis bastante imprecisa. E, por esta razão, Paolino (2001) propôs a distribuição Beta como melhor alternativa para modelar situações que envolvem proporções, além de propor o método de máxima verossimilhança, para a estimação dos parâmetros do modelo.

O método de estimação de Paolino (2001) inicia-se considerando equações em termos da média e da dispersão dadas por

$$E(Y) = \frac{\exp(x^T \beta)}{1 + \exp(x^T \beta)} \quad (3.5)$$

e

$$\phi_Y = \exp(\mathbf{z}^T \boldsymbol{\gamma}),$$

no qual x^T e \mathbf{z}^T são os vetores das observações sobre as covariáveis independentes (geralmente \mathbf{z}^T é um subconjunto de covariáveis de x^T) e $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ são vetores dos parâmetros do modelo de regressão. A dispersão é relatada pela variância, como

$$Var(Y) = \frac{E(Y)(1 - E(Y))}{\phi_Y + 1}.$$

Nas situações onde vetores de observações $(Y_k, x_{k1}, \dots, x_{kk})$ estejam disponíveis, os dois parâmetros da distribuição Beta, p_k e q_k , para cada Y_k , $k = 1, \dots, n$ podem, então ser especificados como

$$p_k = \frac{(E(Y_k))^2 (1 - E(Y_k))}{Var(Y_k)} - E(Y_k) \quad (3.6)$$

e

$$q_k = \frac{(E(Y_k))(1 - E(Y_k))^2}{Var(Y_k)} - (1 - E(Y_k)). \quad (3.7)$$

Estas quantidades são então substituídas em uma função de log – verossimilhança

proveniente da f.d.p. da distribuição Beta,

$$\ln L(p_k, q_k) = \sum_{k=1}^n \{ \ln \Gamma(p_k + q_k) - \ln \Gamma(p_k) - \ln \Gamma(q_k) + (p_k - 1) \ln(y_k) + (q_k - 1) \ln(1 - y_k) \}. \quad (3.8)$$

Em seguida, substituímos $E(Y_k)$, presente na função de verossimilhança através de p_k e q_k , pela expressão (3.5). A função de verossimilhança é agora função dos coeficientes da regressão β . A partir deste ponto o processo de estimação é o usual, de máxima verossimilhança.

Este método produz resultados similares ao método de aproximação normal padrão, mas com grande aumento na facilidade de interpretação, além de trazer benefícios que permitem especificar a equação da média e da dispersão utilizando diferentes subconjuntos de covariáveis.

3.4 O modelo reparametrizado

A densidade (2.1), dada por

$$f(y_k) = \frac{\Gamma(p_k + q_k)}{\Gamma(p_k) \Gamma(q_k)} y_k^{(p_k-1)} (1 - y_k)^{(q_k-1)},$$

com as reparametrizações (3.1) e (3.2), considerando μ_k , é reescrita como

$$f(y_k; \mu_k, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu_k \phi) \Gamma(\phi(1 - \mu_k))} y_k^{(\mu_k \phi - 1)} (1 - y_k)^{(\phi(1 - \mu_k) - 1)} \quad (3.9)$$

e, desse modo, o logaritmo da função de verossimilhança é então

$$l(\beta, \phi) = \sum_{k=1}^n l_k(\mu_k, \phi),$$

sendo $l_k(\mu_k, \phi)$ dada na expressão (3.8), desse modo,

$$\begin{aligned} l(\beta, \phi) &= n \ln [\Gamma(\phi)] - \sum_{k=1}^n \ln \left[\Gamma \left(\frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \phi \right) \right] - \\ &\quad - \sum_{k=1}^n \ln \left[\Gamma \left(\phi \left(1 - \frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \right) \right) \right] + \\ &\quad + \sum_{k=1}^n \left\{ \left(\frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \phi - 1 \right) \ln [y_k] \right\} + \\ &\quad + \sum_{k=1}^n \left\{ \left[\phi \left(1 - \frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \right) - 1 \right] \ln (1 - y_k) \right\}, \end{aligned}$$

ou ainda,

$$\begin{aligned} l(\beta, \phi) &= n \ln [\Gamma(\phi)] - \ln \left[\prod_{k=1}^n \Gamma \left(\frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \phi \right) \right] - \\ &\quad - \ln \left[\prod_{k=1}^n \Gamma \left(\phi \left(1 - \frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \right) \right) \right] + \\ &\quad + \sum_{k=1}^n \left\{ \left(\frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \phi - 1 \right) \ln [y_k] \right\} + \\ &\quad + \sum_{k=1}^n \left\{ \left[\phi \left(1 - \frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \right) - 1 \right] \ln (1 - y_k) \right\}, \end{aligned}$$

que pode ser escrita como,

$$\begin{aligned} l(\beta, \phi) &= n \ln [\Gamma(\phi)] - \ln \left[\prod_{k=1}^n \Gamma \left(\frac{\phi \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right] - \\ &\quad - \ln \left[\prod_{k=1}^n \Gamma \left(\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right] + \\ &\quad + \sum_{k=1}^n \left\{ \left(\frac{(\phi - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \ln [y_k] \right\} + \\ &\quad + \sum_{k=1}^n \left\{ \left[\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} - 1 \right] \ln (1 - y_k) \right\}. \end{aligned}$$

A estimação é feita através da função escore, apresentada a seguir.

3.5 Função Escore e Matriz de Informação

Nesta seção apresentamos a função escore e a matriz de informação de Fisher para o vetor de parâmetros β e para o parâmetro de precisão ϕ .

O logaritmo natural da função de verossimilhança baseada em uma amostra de n observações independentes é

$$l(\boldsymbol{\beta}, \phi) = \sum_{k=1}^n l_k(\mu_k, \phi),$$

$l_k(\mu_k, \phi)$ como em (3.8), com $\mu_k = g^{-1}(\eta_k)$, dependendo da função de ligação empregada. A função escore é obtida derivando o logaritmo da função de verossimilhança com respeito aos parâmetros desconhecidos. Segue que, para $k=1, 2, \dots, n$,

$$\frac{\partial l(\boldsymbol{\beta}, \phi)}{\partial \beta_i} = \sum_{k=1}^n \left\{ \frac{\partial l_k(\mu_k, \phi)}{\partial \mu_k} \frac{d\mu_k}{d\eta_k} \frac{\partial \eta_k}{\partial \beta_i} \right\}, \quad (3.10)$$

com

$$\frac{\partial l_k(\mu_k, \phi)}{\partial \mu_k} = \phi \left\{ \log \left(\frac{y_k}{1 - y_k} \right) - [\Psi(\mu_k \phi) - \Psi((1 - \mu_k) \phi)] \right\},$$

em que $\Psi(\cdot)$ é uma função digama. Note que,

$$\frac{d\mu_k}{d\eta_k} = \frac{1}{g'(\mu_k)}$$

e

$$\frac{\partial \eta_k}{\partial \beta_i} = x_{ki}.$$

Antes de obter a expressão final para (3.10), sejam, $y_k^* = \log \left(\frac{y_k}{1 - y_k} \right)$ e $\mu_k^* = -[\Psi(\mu_k \phi) - \Psi((1 - \mu_k) \phi)]$.

Assim, a expressão (3.10) pode ser reescrita como

$$\frac{\partial l_k(\mu_k, \phi)}{\partial \mu_k} = \phi \sum_{k=1}^n (y_k^* - \mu_k^*)$$

e finalmente, reduz-se a

$$\frac{\partial l(\beta, \phi)}{\partial \beta_k} = \sum_{k=1}^n \left\{ \phi \sum_{k=1}^n (y_k^* - \mu_k^*) \frac{1}{g'(\mu_k)} x_{ki} \right\}. \quad (3.11)$$

Note também que, assumindo condições usuais de regularidade, o valor esperado de $(\partial l_k(\mu_k, \phi) / \partial \mu_k)$ é igual a zero. Mas,

$$E \left[\frac{\partial l_k(\mu_k, \phi)}{\partial \mu_k} \right] = 0 \Leftrightarrow E[\phi(y_k^* - \mu_k^*)] = 0 \Leftrightarrow E[y_k^*] = \mu_k^*,$$

ou seja, o valor esperado da variável transformada y_k^* iguala-se a μ_k^* .

A função escore para β , o vetor coluna $U_\beta(\beta, \phi)$ de dimensão k , pode ser escrita na seguinte forma matricial

$$U_\beta(\beta, \phi) = \phi X^T V (y^* - \mu^*)$$

com $V = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$, $y^* = (y_1^*, \dots, y_n^*)^k$ e $\mu^* = (\mu_1^*, \dots, \mu_n^*)^T$.

Analogamente, a derivada de 1ª ordem do logaritmo da função de verossimilhança em relação a ϕ é dada por:

$$\frac{\partial l(\beta, \phi)}{\partial \phi} = \sum_{k=1}^n \frac{\partial l_k(\mu_k, \phi)}{\partial \phi},$$

como,

$$\begin{aligned} \frac{\partial l(\beta, \phi)}{\partial \phi} &= \mu_k \left\{ \log \left(\frac{y_k}{1 - y_k} \right) - [\Psi(\mu_k \phi) - \Psi((1 - \mu_k) \phi)] \right\} + \\ &+ \log(1 - y_k) - \Psi((1 - \mu_k) \phi) + \Psi(\phi). \end{aligned}$$

Então, a função escore para o parâmetro de precisão ϕ é dada por

$$U_\phi(\beta, \phi) = \sum_{k=1}^n \{ \mu_k (y_k^* - \mu_k^*) + \log(1 - y_k) - [\Psi((1 - \mu_k) \phi) - \Psi(\phi)] \}.$$

As derivadas de 2ª ordem de $l(\beta, \phi)$ com respeito a β_i e β_j são dadas por:

$$\begin{aligned} \frac{\partial l^2(\beta, \phi)}{\partial \beta_i \partial \beta_j} &= \sum_{k=1}^n \left\{ \frac{\partial}{\partial \mu_k} \left(\frac{\partial l_k(\mu_k, \phi)}{\partial \mu_k} \cdot \frac{d\mu_k}{d\eta_k} \right) \frac{d\mu_k}{d\eta_k} \frac{\partial \eta_k}{\partial \beta_i} x_{ki} \right\} = \\ &= \sum_{k=1}^n \left\{ \frac{\partial l_k^2(\mu_k, \phi)}{\partial \mu_k^2} \frac{d\mu_k}{d\eta_k} + \frac{\partial l_k(\mu_k, \phi)}{\partial \mu_k} \frac{\partial}{\partial \mu_k} \left(\frac{d\mu_k}{d\eta_k} \right) \right\} \frac{d\mu_k}{d\eta_k} x_{ki} x_{kj}. \end{aligned}$$

Como $E\left(\frac{\partial l_k(\mu_k, \phi)}{\partial \mu_k}\right) = 0$, temos,

$$E\left(\frac{\partial l^2(\beta, \phi)}{\partial \beta_i \partial \beta_j}\right) = \sum_{k=1}^n E\left(\frac{\partial l_k^2(\mu_k, \phi)}{\partial \mu_k^2}\right) \left(\frac{d\mu_k}{d\eta_k}\right)^2 x_{ki} x_{kj}.$$

E temos

$$\frac{\partial l_k^2(\mu_k, \phi)}{\partial \mu_k^2} = -\phi^2 [\Psi'(\mu_k \phi) - \Psi'((1 - \mu_k) \phi)],$$

em que $\Psi'(\cdot)$ é a função trigama definida anteriormente e assim

$$E\left(\frac{\partial l_k^2(\mu_k, \phi)}{\partial \mu_k^2}\right) = -\phi \sum_{k=1}^n \{w_k x_{ki} x_{kj}\}.$$

com,

$$w_k = \phi^2 [\Psi'(\mu_k \phi) - \Psi'((1 - \mu_k) \phi)] \frac{1}{[g'(\mu_k)]^2}.$$

Em forma matricial, temos,

$$E\left(\frac{\partial l^2(\beta, \phi)}{\partial \beta \partial \beta^k}\right) = -\phi X^T W X,$$

com $W = \text{diag}\{w_1, \dots, w_n\}$.

As derivadas de 2ª ordem de $l(\beta, \phi)$ com respeito a β_i e ϕ são dadas por:

$$\frac{\partial l^2(\beta, \phi)}{\partial \beta_i \partial \phi} = \sum_{k=1}^n \left[(y_k^* - \mu_k^*) - \phi \frac{\partial \mu_k^*}{\partial \phi} \right] \frac{1}{g'(\mu_k)} x_{ki}.$$

Como $E[y_k^*] = \mu_k^*$ e

$$\frac{\partial \mu_k^*}{\partial \phi} = [\Psi'(\mu_k \phi) \mu_k - \Psi'((1 - \mu_k) \phi) (1 - \mu_k)],$$

temos,

$$E \left(\frac{\partial l^2(\beta, \phi)}{\partial \beta_i \partial \phi} \right) = - \sum_{k=1}^n \left\{ c_k \frac{1}{g'(\mu_k)} x_{ki} \right\},$$

em que,

$$c_k = \phi \{ \Psi'(\mu_k \phi) \mu_k - \Psi'((1 - \mu_k) \phi) (1 - \mu_k) \}.$$

Em notação matricial

$$E \left(\frac{\partial l^2(\beta, \phi)}{\partial \beta_i \partial \phi} \right) = -X^T V c,$$

com $c = (c_1, \dots, c_n)^T$.

Finalmente, temos

$$\frac{\partial l^2(\beta, \phi)}{\partial^2 \phi} = - \sum_{k=1}^n \{ \Psi'(\mu_k \phi) \mu_k^2 + \Psi'((1 - \mu_k) \phi) (1 - \mu_k)^2 - \Psi'(\phi) \}.$$

Seja,

$$d_k = \{ \Psi'(\mu_k \phi) \mu_k^2 + \Psi'((1 - \mu_k) \phi) (1 - \mu_k)^2 - \Psi'(\phi) \},$$

logo,

$$E \left(\frac{\partial l^2(\beta, \phi)}{\partial^2 \phi} \right) = - \sum_{k=1}^n d_k,$$

que em notação matricial pode ser escrito como,

$$E \left(\frac{\partial l^2(\beta, \phi)}{\partial^2 \phi} \right) = -tr(D),$$

onde, $D = diag(d_1, \dots, d_n)$ e $tr(D)$ indica o traço da matriz D .

Finalmente, a matriz de informação de Fisher para (β, ϕ) é dada por

$$K = K(\beta, \phi) = \begin{pmatrix} K_{\beta\beta} & K_{\beta\phi} \\ K_{\phi\beta} & K_{\phi\phi} \end{pmatrix},$$

onde,

$$K_{\beta\beta} = -E \left(\frac{\partial l^2(\beta, \phi)}{\partial \beta \partial \beta^k} \right) = \phi X^T W X;$$

$$K_{\beta\phi} = K_{\phi\beta}^T = -E \left(\frac{\partial^2 l^2(\beta, \phi)}{\partial \beta_k \partial \phi} \right) = X^T V c;$$

$$K_{\phi\phi} = -E \left(\frac{\partial^2 l^2(\beta, \phi)}{\partial^2 \phi} \right) = \text{tr}(D).$$

3.6 Estimação dos Parâmetros via Função Escore

Os estimadores de máxima verossimilhança para β e ϕ são obtidos das equações $U_\beta(\beta, \phi) = 0$ e $U_\phi(\beta, \phi) = 0$. Estas equações não apresentam uma solução analítica em forma fechada. Assim, é necessário obter os estimadores através de maximização numérica do algoritmo de Newton – Raphson ou Escore de Fisher. Outra alternativa é utilizar o algoritmo de quase – Newton, conhecido como BFGS.

3.6.1 Processo Iterativo de Newton – Raphson

Seja $\theta = (\beta, \phi)^T$ o vetor de parâmetros. $U(\theta) = (U_\beta(\beta, \phi)^T; U_\phi(\beta, \phi))^T$, o vetor de funções escore de dimensão $(k+1) \times 1$. Para obtenção do estimador de máxima verossimilhança do vetor θ expandimos a função escore $U(\theta)$ em torno de um valor inicial $\theta^{(0)}$, tal que:

$$U(\theta) \simeq U(\theta^{(0)}) + U'(\theta^{(0)}) (\theta - \theta^{(0)}),$$

em que $U'(\theta)$ denota a derivada de 1ª ordem de $U(\theta)$ com respeito a θ^T . Fazendo $U(\theta) = 0$ e repetindo o processo acima, chegamos ao processo iterativo:

$$\theta^{(m+1)} = \theta^{(m)} + \left\{ -U'(\theta^{(m)}) \right\}^{-1} U(\theta^{(m)}), \quad m = 0, 1, \dots \quad (3.12)$$

O aspecto mais trabalhoso desse esquema iterativo é a inversão da matriz $U'(\theta)$.

3.6.2 Processo Iterativo de Escore de Fisher

A expressão (3.12) apresenta uma forma alternativa equivalente assintoticamente, uma vez que, pela lei dos grandes números, $U'(\theta)$ converge para a matriz K quando $n \rightarrow \infty$. Assim, substituindo a informação observada em (3.12) pela esperada, obtemos a seguinte aproximação:

$$\theta^{(m+1)} = \theta^{(m)} + \{-K^{(m)}\}^{-1} U(\theta^{(m)}), \quad m = 0, 1, \dots,$$

esse procedimento iterativo é denominado Escore de Fisher. Da mesma forma que o método de Newton – Raphson, o aspecto mais trabalhoso é a inversão da matriz K .

3.6.3 Método BFGS

Este método utiliza o mesmo princípio do método de Newton – Raphson, diferenciando-se pelo fato de utilizar uma seqüência de matrizes simétricas e positivas definidas $B^{(m)}$ no lugar da matriz $U'(\theta^{(m)})^{-1}$. Comumente, toma-se como matriz inicial, $B^{(0)}$, a matriz identidade de mesma ordem. A forma recursiva para obter as demais matrizes é dada por

$$B^{(m+1)} = B^{(m)} - \frac{B^{(m)}s^{(m)}(s^{(m)})^T B^{(m)}}{(s^{(m)})^T B^{(m)}s^{(m)}} + \frac{y^{(m)}(y^{(m)})^T}{(y^{(m)})^T y^{(m)}}, \quad m = 0, 1, \dots,$$

em que $s^{(m)} = \theta^{(m+1)} - \theta^{(m)}$ e $y^{(m)} = U(\theta^{(m+1)}) - U(\theta^{(m)})$. De forma análoga aos métodos anteriores, o máximo é obtido pela recorrência

$$\theta^{(m+1)} = \theta^{(m)} - \alpha^{(m)} B^{(m)} U(\theta^{(m)}), \quad m = 0, 1, \dots$$

Os três processos iterativos são sensíveis à estimativa inicial $\theta^{(0)}$. Ferrari e Cribari – Neto(2004) sugerem utilizar como uma estimativa inicial de β a estimativa de mínimos quadrados ordinários desse vetor de parâmetros, obtida de uma regressão linear da resposta transformada em $g(y_1), \dots, g(y_n)$ em X , isto é, $(X^T X)^{-1} X^T z$, em que $z = (g(y_1), \dots, g(y_n))^T$. Em relação ao parâmetro de precisão, os mesmos autores sugerem um valor inicial para ϕ baseado no fato de que $Var(Y_k) = \mu_k(1 - \mu_k) / (1 + \phi)$. De forma alternativa, podemos escrever:

$$\phi = \left[\mu_k \frac{(1 - \mu_k)}{\text{Var}(Y_k)} \right] - 1.$$

3.7 Testes de Hipóteses

Nesta seção apresentamos os testes de hipóteses utilizados na análise de um modelo de regressão beta, que são idênticos aos utilizados em modelos lineares generalizados. Os métodos de inferência para o modelo de regressão beta baseiam-se, fundamentalmente, na teoria da máxima verossimilhança. De acordo com esta teoria, existem três estatísticas para testar hipóteses relativas aos parâmetros β' s. São elas a razão de verossimilhanças, Escore e Wald.

3.7.1 Teste da Razão de Verossimilhança

É possível realizar testes assintóticos para fazer inferência sobre os parâmetros desconhecidos. Considere, por exemplo, o teste da hipótese nula $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$ versus a hipótese alternativa $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}^{(0)}$, em que $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^T$ e $\boldsymbol{\beta}^{(0)} = (\beta_1^{(0)}, \dots, \beta_m^{(0)})^T$ para $m < k$ e $\boldsymbol{\beta}^{(0)}$ dado. A estatística da razão de verossimilhanças é

$$\varpi_1 = 2 \left\{ l \left(\hat{\boldsymbol{\beta}}, \hat{\phi} \right) - l \left(\tilde{\boldsymbol{\beta}}, \tilde{\phi} \right) \right\},$$

sendo, $l \left(\hat{\boldsymbol{\beta}}, \hat{\phi} \right)$ o logaritmo natural da função de verossimilhança e $\left(\tilde{\boldsymbol{\beta}}, \tilde{\phi} \right)^T$ o estimador de máxima verossimilhança restrito de $(\boldsymbol{\beta}, \phi)$ obtido pela imposição da hipótese nula. Sob condições gerais de regularidade e sob H_0 , $\varpi_1 \xrightarrow{D} \chi_m^2$. Dessa forma o teste pode ser avaliado usando valores críticos aproximados da distribuição χ_m^2 , onde m é a dimensão do espaço paramétrico sob a hipótese H_0 .

3.7.2 Teste Escore

Para descrever o teste escore de $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}^{(0)}$ versus $H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}^{(0)}$, considere $U_{1\beta}$ um vetor coluna m dimensional contendo os primeiros m elementos da função escore de $\boldsymbol{\beta}$ e $K_{11}^{\beta\beta}$ a matriz $m \times m$ formada das m primeiras linhas e das m primeiras colunas da

matriz de K^{-1} . Pode-se mostrar que $U_{1\beta} = \phi X_1^k (y^* - \mu^*)$, em que X é particionada como $[X_1, X_2]$ seguindo a partição de β . A estatística escore de Rao pode ser escrita com

$$\varpi_2 = \tilde{U}_{1\beta}^T K_{11}^{\beta\beta} U_{1\beta},$$

em que o til indica que as quantidades estão sendo avaliadas no estimador de máxima verossimilhança restrito. Sob condições gerais de regularidade e sob H_0 , $\varpi_2 \xrightarrow{D} \chi_m^2$, onde m é a dimensão do espaço paramétrico sob a hipótese H_0 .

3.7.3 Teste de Wald

Da mesma forma, podemos utilizar o teste de Wald para realizar inferências assintóticas acerca do vetor de parâmetros β . A estatística do teste $H_0 : \beta = \beta^{(0)}$ versus $H_1 : \beta \neq \beta^{(0)}$ é:

$$\varpi_3 = \left(\hat{\beta} - \beta^{(0)} \right)^T \left(\hat{K}_{11}^{\beta\beta} \right)^{-1} \left(\hat{\beta} - \beta^{(0)} \right),$$

em que $\hat{K}_{11}^{\beta\beta}$ é igual a $K_{11}^{\beta\beta}$ avaliado no estimador de máxima verossimilhança sem restrição, e $\hat{\beta}_k$ é o estimador de máxima verossimilhança de β_k . Sob condições gerais de regularidade e sob H_0 , $\varpi_3 \xrightarrow{D} \chi_m^2$, onde m é a dimensão do espaço paramétrico sob a hipótese H_0 . Em particular, para testar a significância do k -ésimo parâmetro regressor β_k , $k= 1, 2, \dots, n$, podemos utilizar a raiz quadrada sinalizada da estatística de Wald, isto é,

$$\frac{\hat{\beta}_k}{sd \left(\hat{\beta}_k \right)},$$

no qual, $sd \left(\hat{\beta}_k \right)$ é o erro padrão assintótico do estimador de máxima verossimilhança de β_k , obtido da inversa da matriz de informação de Fisher avaliada nos estimadores de máxima verossimilhança.

Adicionalmente aos testes, podemos construir intervalos de confiança para algumas quantidades de interesse. Um intervalo de confiança de aproximadamente $(1 - \alpha) 100\%$

para β_k , $k=1, 2, \dots, n$ tem limites dados por

$$\hat{\beta}_k \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) sd \left(\hat{\beta}_k \right).$$

Similarmente, um intervalo de confiança assintótico $(1 - \alpha) 100\%$ para ϕ tem limites

$$\hat{\phi} \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) sd \left(\hat{\phi} \right).$$

Finalmente, um intervalo de confiança de aproximadamente $(1 - \alpha) 100\%$ para μ , a média da variável resposta, para um dado vetor de valores de covariadas, pode ser calculado como,

$$\left[g^{-1} \left(\hat{\eta} \pm \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) sd \left(\hat{\eta} \right) \right) \right],$$

ou seja,

$$\left[g^{-1} \left(\hat{\eta} - \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) sd \left(\hat{\eta} \right) \right); g^{-1} \left(\hat{\eta} + \Phi^{-1} \left(1 - \frac{\alpha}{2} \right) sd \left(\hat{\eta} \right) \right) \right],$$

em que $\hat{\eta} = x^T \hat{\beta}$ e $sd \left(\hat{\eta} \right) = \sqrt{x^T cov \left(\hat{\beta} \right) x}$. Aqui, $cov \left(\hat{\beta} \right)$ é obtida da inversa da matriz de informação de Fisher avaliada nas estimativas de máxima verossimilhança excluindo a linha e a coluna da matriz correspondente ao parâmetro de dispersão. Os intervalos de confiança acima são válidos apenas para funções de ligação estritamente crescentes.

3.8 Técnicas de Diagnóstico

Uma etapa importante na análise de um ajuste de regressão é a verificação de possíveis afastamentos das suposições feitas para o modelo, especialmente para a parte aleatória (y_k) e pelo componente sistemático (η_k), bem como a existência de observações extremas que podem causar desvios nos resultados do ajuste.

Inicialmente, serão apresentadas três medidas globais de qualidade do ajuste do modelo e ferramentas gráficas para a detecção de desvios do modelo postulado e observações influentes. Posteriormente, serão apresentados como medidas de análise de influência, o

leverage generalizado proposta por Wei, Hu e Fung (1998) e a distância de Cook (Cook, 1977).

Em princípio, uma medida global da qualidade do ajuste pode ser obtida através do cálculo do “pseudo” R^2 (R_p^2) definido como o quadrado do coeficiente de correlação amostral entre $g(y)$ e $\hat{\eta}$. Note que $0 \leq R_p^2 \leq 1$ e quanto mais próximo de 1 for o seu valor, melhor será o ajuste.

Mittböck e Schemper (1996) revisaram algumas medidas da qualidade do ajuste em modelos de regressão logísticos. Utilizaremos duas dessas medidas aqui. A primeira, ainda baseada no pseudo R^2 , é o quadrado do coeficiente de correlação amostral entre y e $\hat{\mu} \left(g^{-1} \left(\hat{\eta} \right) \right)$, que denominaremos por R_*^2 . Essa medida apresenta resultados satisfatórios nos estudos realizados por Mittböck e Schemper (1996). A segunda medida é baseada na razão de verossimilhança sendo definida como

$$R_{LR}^2 = 1 - \left[\frac{L_0}{\hat{L}} \right]^{2/n},$$

sendo, L_0 e \hat{L} denotam, respectivamente, as verossimilhanças maximizadas no modelo nulo (sem covariáveis) e no modelo de interesse.

3.8.1 Resíduo Componente do Desvio

Ferrari e Cribari–Neto (2004) propõem uma medida global de qualidade de ajuste, baseada no fato de que a discrepância de um ajuste pode ser medida como duas vezes a diferença entre o máximo do logaritmo da verossimilhança do modelo saturado e o do modelo de interesse (modelo postulado). Esta medida foi originalmente proposta por Nelder e Wedderburn (1972) no contexto dos modelos lineares generalizados. Seja

$$D(y; \mu, \phi) = \sum_{k=1}^n 2 \left\{ l_k \left(\tilde{\mu}_k, \hat{\phi} \right) - l \left(\hat{\mu}_k, \phi \right) \right\}^{1/2},$$

para, $\tilde{\mu}_k$ a solução de $\frac{\partial l_k}{\partial \mu_k} = 0$, isto é, $\phi(y_k^* - \mu_k^*) = 0$. Quando ϕ é grande, $\mu_k^* \approx \log \left(\frac{\mu_k^*}{1 - \mu_k^*} \right)$, e então segue que $\tilde{\mu}_k \approx y_k$. Para ϕ conhecido, podemos definir uma média

de discrepância como $D(y; \tilde{\mu}, \phi)$, em $\tilde{\mu}$ é o estimador de máxima verossimilhança de μ sob o modelo sendo pesquisado. Quando ϕ é desconhecido, uma aproximação para essa quantidade é $D(y; \hat{\mu}, \hat{\phi})$, denominada usualmente de desvio do modelo sob pesquisa. Note que

$$D(y; \hat{\mu}, \hat{\phi}) = \sum_{k=1}^n (r_k^d)^2,$$

sendo, $r_k^d = \text{ sinal}(y_k - \hat{\mu}_k) \left\{ 2 \left\{ l_k(\hat{\mu}_k, \hat{\phi}) - l(\hat{\mu}_k, \hat{\phi}) \right\}^{1/2} \right\}$.

Note que a t -ésima observação contribui com a quantidade $(r_k^d)^2$ para o desvio e uma observação com um valor absoluto grande de r_k^d pode ser vista como discrepante.

O resíduo ordinário padronizado é definido como,

$$r_k^d = \frac{y_k - \hat{\mu}_k}{\sqrt{\hat{Var}(Y_k)}},$$

em que, $\hat{\mu}_k = g^{-1}\left(x_k^T \hat{\beta}\right)$ e $\hat{Var}(Y_k) = \frac{\left\{ \hat{\mu}_k (1 - \hat{\mu}_k) \right\}}{(1 + \hat{\phi})}$. Um gráfico desses resíduos con-

tra a ordem das observações (k) não deveria mostrar tendência. Através desse gráfico é possível verificar se existem pontos suspeitos de serem atípicos. Um outro gráfico importante é o de r_k contra $\hat{\eta}_k$ que é utilizado para verificar se a função de ligação está mal especificada (caso o gráfico apresente alguma tendência).

3.8.2 Resíduos Ponderados e Resíduos Padronizados Ponderados

Espinheira, Ferrari e Cribari-Neto (2006) propuseram, ainda, dois novos tipos de resíduos para a regressão beta: os resíduos ponderados e os resíduos padronizados ponderados.

Resíduos ponderados

Os resíduos ponderados são baseados no algoritmo do método iterativo Score de Fisher para a estimação de β quando ϕ é fixo. Onde o m -ésimo passo da iteração é dado por

$$\beta^{(m+1)} = \beta^{(m)} - (X^T W^{(m)} X)^{-1} X^T V^{(m)} \left(y^{*(m)} - \mu^{*(m)} \right), \quad (3.13)$$

onde, os k -ésimos elementos de y^* e μ^* são dados, respectivamente, por

$$y_k^* = \log \left(\frac{y_k}{1 - y_k} \right)$$

e

$$\mu_k^* = - [\Psi(\mu_k \phi) - \Psi((1 - \mu_k) \phi)].$$

As matrizes V e W são dadas como $V = \text{diag}\{1/g'(\mu_1), \dots, 1/g'(\mu_n)\}$ e $W = \text{diag}\{w_1, \dots, w_n\}$, para, $w_k = \phi v_k \left\{ \frac{1}{[g'(\mu_k)]^2} \right\}$, com $v_k = \Psi'(\mu_k \phi) - \Psi'((1 - \mu_k) \phi)$.

É possível escrever o processo iterativo em (3.13) em termos de mínimos quadrados ponderados:

$$\beta^{(m+1)} = (X^k W^{(m)} X)^{-1} X^k W^{(m)} z^{(m)},$$

para,

$$z^{(m)} = \eta^{(m)} - W^{-1(m)} V^{(m)} \left(y^* - \mu^{*(m)} \right).$$

Se o algoritmo converge, tem-se

$$z = \hat{\eta} - \hat{W} V \left(y^* - \hat{\mu}^* \right).$$

Note que $\hat{\beta}$ pode ser visto como o estimador de mínimos quadrados de β obtido na regressão de z em X com matriz de pesos W . Os resíduos ordinários são:

$$r^* = \hat{W}^{1/2} \left(z - \hat{\eta} \right) = \hat{W}^{1/2} \hat{V} \left(y^* - \hat{\mu}^* \right).$$

Dessa forma, Espinheira, Ferrari e Cribari-Neto (2006) definiram os resíduos ponderados, como:

$$r^* = \frac{y_k^* - \hat{\mu}_k^*}{\sqrt{\phi v_k}}.$$

Resíduos padronizados ponderados

Os resíduos padronizados ponderados, que utiliza a variância de z , é dado por

$$r_k^* = \phi^{1/2} r^* = \frac{y_k^* - \hat{\mu}_k^*}{\sqrt{v_k}},$$

que pode ser denotado por r_k^p . Desde que,

$$\text{cov}(z) \approx \phi^{-1} \hat{W}^{-1},$$

segue que

$$\text{cov}(r^*) \approx \phi^{-1} (I_n - H),$$

onde I_n é a matriz identidade $n \times n$ e $H = \hat{W}^{1/2} X \left(X^T \hat{W} X \right)^{-1} X^T \hat{W}^{1/2}$. Espinheira, Ferrari e Cribari-Neto (2006) definiu resíduo padronizado ponderado como:

$$r_k^p = \frac{r_k^*}{\sqrt{\phi^{-1} (1 - h_{kk})}} = \frac{r_k^p}{\sqrt{(1 - h_{kk})}} = \frac{y_k^* - \hat{\mu}_k^*}{\sqrt{v_k (1 - h_{kk})}},$$

sendo que h_{kk} denota o k -ésimo elemento da diagonal de H . Espinheira, Ferrari e Cribari-Neto (2006) notaram que ϕ deve ser substituído por $\hat{\phi}$, que é o estimador de máxima verossimilhança de ϕ .

3.8.3 Gráficos de Probabilidade Meio – normal com Envelopes

Como a distribuição dos resíduos não é conhecida, gráficos de probabilidade half – normal com envelopes simulados são ferramentas de diagnóstico muito úteis. A idéia é acrescentar ao gráfico meio – normal usual um envelope simulado que pode ser usado para decidir se as respostas observadas são consistentes com modelo ajustado. Apresentamos a seguir os passos para a construção do gráfico com envelope simulado:

1. Ajuste o modelo e gere uma amostra simulada de n observações independentes utilizando o modelo ajustado como se este fosse o modelo verdadeiro;

2. Ajuste o modelo para a amostra gerada e calcule os valores absolutos ordenados da medida de diagnóstico de interesse;
3. Repita os passos (1) e (2) k vezes;
4. Considere n grupos de k estatísticas de ordem; para cada grupo, calcule suas respectivas médias, valores mínimos e máximos; e
5. Construa o gráfico desses valores e da medida de diagnóstico ordenada da amostra original contra os escores meio – normais $\Phi^{-1} \left((k + n - \frac{1}{8}) / (2n + \frac{1}{2}) \right)$ em que $\Phi(\cdot)$ é a função de distribuição acumulada da distribuição normal padrão.

Os valores mínimos e máximos das k estatísticas de ordem produzem o envelope. A sugestão de Atkinson (1985) é usar $k = 19$. Desse modo, a probabilidade do maior resíduo de um envelope particular exceder o limite superior fica sendo aproximadamente igual a $1/20 = 0,05$. As correspondentes observações dos resíduos absolutos que se encontram fora dos limites fornecidos pelo envelope simulado, merecem um estudo adicional. Tendências não aleatórias dos resíduos dentro do envelope gerado podem indicar especificação incorreta da distribuição dos dados, da parte sistemática do modelo ou da função de ligação.

3.8.4 *Leverage* Generalizado

Leverage é um componente chave na análise de influência em modelos de regressão. Usualmente, é medido pelos elementos h_{ii} da matriz H que é conhecida como matriz de projeção ou matriz chapéu e é usado para avaliar a importância individual de cada observação no próprio valor ajustado. Na regressão linear múltipla, por exemplo, é razoável utilizar h_{ii} como uma medida da influência da i -ésima observação sobre o próprio valor ajustado. Supondo que todos exerçam a mesma influência sobre os valores ajustados, pode-se esperar que h_{ii} esteja próximo de $\frac{\text{tr}(H)}{n} = \frac{k}{n}$. Uma sugestão é examinar aqueles pontos tais que $h_{ii} \geq \frac{2k}{n}$, que são conhecidos como pontos de *leverage* grandes. Wei, Hu e Fung (1998) generalizaram a definição de *leverage* para modelos gerais cuja variável resposta seja contínua. Nessa generalização incluem-se outros métodos de estimação, além de máxima verossimilhança, e outros enfoques tal como Bayesiano.

O *leverage* generalizado proposto por Wei, Hu e Fung (1998) é definido como

$$GL(\tilde{\theta}) = \frac{\partial \tilde{y}}{\partial y^k} = \left(\frac{\partial \tilde{y}}{\partial y_n} \right)_{n \times n},$$

sendo β um vetor s -dimensional tal que $E(y) = \mu(\theta)$ e $\tilde{\theta}$ é um estimador de θ , com $\tilde{y} = \mu(\tilde{\theta})$. Por esta definição, podemos observar que o *leverage* generalizado para a k -ésima observação é a razão instantânea de mudança do k -ésimo valor predito em relação ao k -ésimo valor da resposta.

Apresentamos a seguir o cálculo do *leverage* generalizado no modelo de regressão beta sob dois enfoques. Primeiramente sob a suposição de que o parâmetro de precisão ϕ é conhecido e, num segundo momento, admitindo que ϕ é desconhecido.

Considerando ϕ conhecido, Ferrari e Cribari-Neto (2004) obtiveram uma expressão de forma fechada para $GL(\beta)$ no modelo de regressão beta proposto anteriormente. O *leverage* generalizado é dado por

$$GL(\beta) = D_\beta \left(-\frac{\partial^2 l}{\partial \beta \partial \beta^T} \right)^{-1} \frac{\partial^2 l}{\partial \beta \partial \beta^T},$$

onde,

$$D_\beta = \frac{\partial \mu}{\partial \beta^T} = \frac{\partial \mu}{\partial \eta^T} \frac{\partial \eta}{\partial \beta^T} = VX,$$

e assim, temos que a expressão acima pode ser escrita como

$$GL(\beta) = VX (X^T Q X)^{-1} V X M.$$

Ferrari e Cribari-Neto (2004) notaram que ao substituir a informação observada, $-\frac{\partial^2 l}{\partial \beta \partial \beta^T}$, pela informação esperada, $E\left(-\frac{\partial^2 l}{\partial \beta \partial \beta^T}\right)$, a expressão para $GL(\beta)$ fica como a expressão anterior, mas com a matriz Q substituída por W , sendo dada por $GL^*(\beta)$. Pode-se notar que os elementos da diagonal $GL^*(\beta)$ são os mesmos de

$$M^{1/2} VX (X^T W X)^{-1} X^T V M^{1/2},$$

e que $M^{1/2}$ é uma matriz diagonal cujo k -ésimo elemento diagonal é dado por $\left\{ g^{-1}(\mu_k) V(\mu_k)^{1/2} \right\}^{-1}$. Outro fato importante aqui é notar que existe uma conexão

estreita entre os elementos da diagonal de $GL^*(\beta)$ e aqueles da “matriz chapéu” usual de modelos lineares generalizados,

$$H = W^{1/2}X (X^T W X)^{-1} X^T W^{1/2},$$

quando ϕ é grande.

3.8.5 Influência

Um outro tópico importante na análise de diagnóstico é a detecção de observações influentes, isto é, pontos que exercem um peso desproporcional nas estimativas dos parâmetros do modelo. Uma medida de influência de cada observação nas estimativas dos parâmetros de regressão é a distância de Cook dada por

$$K^{-1} \left(\hat{\beta} - \beta_j \right)^T X^T W X \left(\hat{\beta} - \beta_j \right),$$

onde $\hat{\beta}_j$ é a estimativa do parâmetro sem a t -ésima observação. Essa quantidade mede a distância quadrática entre $\hat{\beta}$ e $\hat{\beta}_j$. Para evitar ajustar o modelo $n + 1$ vezes utilizamos a aproximação usual para a distância de Cook dada por

$$C_k = \frac{h_{kk} r_k^2}{K (-h_{kk})^2},$$

que combina *leverage* e resíduos. Portanto C_k será grande quando o k -ésimo ponto fornecer r_k grande ou quando h_{kk} for próximo de 1. Uma prática comum é construir um gráfico de C_k contra a ordem das observações, k , para detectar pontos suspeitos de serem influentes.

Capítulo 4

Avaliação de Resultados Assintóticos em Amostras Finitas

4.1 Introdução

Neste capítulo avaliamos numericamente a qualidade das aproximações utilizadas sobre os parâmetros do modelo de regressão beta em amostras finitas. O interesse é avaliar o viés do estimador de máxima verossimilhança, a proximidade de sua distribuição à distribuição normal de referência, a distorção de tamanhos dos testes da razão de verossimilhanças, score e Wald e também verificar a qualidade da aproximação das distribuições das estatísticas dos testes por uma distribuição qui-quadrada adequada. Todos os resultados foram obtidos por simulação de Monte Carlo.

4.2 Detalhes Metodológicos

As simulações computacionais deste trabalho foram realizadas usando a versão 2.3 da linguagem de programação R.

Os resultados numéricos apresentados a seguir baseiam-se no modelo de regressão beta com estrutura

$$g(\mu_k) = \beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3} + \beta_4 x_{k4}, \quad k = 1, 2, \dots, n \quad (4.1)$$

com g sendo a função de ligação logito.

Os passos para a simulação de Monte Carlo são descritos a seguir. Definimos valores verdadeiros para os parâmetros de regressão

$$(\beta_0 = 0,591, \beta_1 = -0,138, \beta_2 = 0,06, \beta_3 = 0, \beta_4 = -0,0376) \quad (4.2)$$

que permanecem fixos para todas as simulações. Para o parâmetro de precisão atribuímos $\phi = 7,61$. A matriz regressora é gerada das distribuições $U(0; 1)$, $N(0; 1)$, t_4 e $Exp(2)$. Os valores de X permanecem constantes em todo o experimento. Consideramos os tamanhos amostrais $n = 20$, $n = 50$, $n = 100$ e $n = 500$. O número de replicações de Monte Carlo foi fixado em 1000.

Para cada réplica de Monte Carlo geramos uma amostra aleatória da variável resposta y_1, \dots, y_n com $y_k \sim beta(p_k, q_k)$ onde

$$p_k = \mu_k \phi$$

e

$$q_k = (1 - \mu_k) \phi$$

e

$$\mu_k = \frac{\exp(\beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3} + \beta_4 x_{k4})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \beta_2 x_{k2} + \beta_3 x_{k3} + \beta_4 x_{k4})}. \quad (4.3)$$

4.3 Comportamento dos Estimadores

Nesta seção apresentamos os resultados de simulação referentes ao comportamento dos estimadores de máxima verossimilhança dos parâmetros do modelo de regressão beta (4.1). A Tabela 4.1 apresenta algumas quantidades relativas às estimativas dos parâmetros quando estes têm valores verdadeiros dados por (4.2) e $\phi = 7,61$ e a matriz de regressores X gerada das distribuições $U(0, 1)$, $N(0, 1)$, k_4 e $Exp(2)$. Nas colunas da Tabela 4.1 são apresentadas as seguintes quantidades: valores reais dos parâmetros, médias das estimativas, variância e média dos limites inferiores e superiores dos intervalos

de confiança, construídos com um coeficiente de confiança de 95%.

Em relação ao comportamento assintótico, espera-se que cada estimador dos parâmetros do modelo tenha uma distribuição normal, ou seja, que as estimativas dos coeficientes de assimetria e curtose estejam em torno de zero. Se o coeficiente de assimetria for positivo, a distribuição é assimétrica à direita e, caso contrário, será assimetria à esquerda. Os parâmetros de regressão apresentam as estimativas desses coeficientes próximos de zero, mesmo para um tamanho amostral pequeno ($n = 20$). Para o parâmetro de dispersão ϕ não apresenta um resultado satisfatório para amostras menor que 50.

Um fato interessante é que os intervalos de confiança médios (obtidos pelas médias dos limites inferiores e superiores dos intervalos de confiança para cada amostra) para todos os parâmetros de regressão são, em média, simétricos em torno dos verdadeiros valores dos parâmetros, independente do tamanho da amostra. O contrário ocorre com o intervalo de confiança para o parâmetro de precisão, este sendo à esquerda, como observado na análise do comportamento assintótico da distribuição desse parâmetro. Além disso, com o aumento do tamanho da amostra, temos intervalos de amplitude média menor para todos os parâmetros, como era de se esperar. Por exemplo, o valor da amplitude média para o parâmetro ϕ quando $n = 20$ é de 21,7473 unidades, enquanto que para $n = 50$ este valor reduz-se para 7,2694 unidades. Verifica-se novamente que o valor do coeficiente de assimetria para o estimador do parâmetro ϕ é sempre negativo, caracterizando uma distribuição assimétrica à esquerda.

Por fim, percebemos que nesse modelo de regressão beta, em que a resposta média é modelada através de uma estrutura de regressão, adotando a função logito como função de ligação, os estimadores de máxima verossimilhança para os parâmetros de regressão apresentam boas propriedades em amostras finitas, pois mostraram-se quase não-viciados e com uma distribuição próxima da distribuição normal de referência.

Tabela 4.1: Médias, variância e intervalos de confiança médios para os parâmetros simulados para o modelo de Regressão Beta.

| n | Parâmetros | Valores Reais | Média das Est. | Var das Est. | LI | LS |
|-----|------------|---------------|----------------|--------------|---------|---------|
| 20 | β_0 | 0,591 | 0,6183 | 0,2331 | -0,7902 | 0,5945 |
| | β_1 | -0,138 | -0,1360 | 0,3852 | -0,2194 | 1,6230 |
| | β_2 | 0,06 | 0,0528 | 0,0309 | -0,3910 | 0,3525 |
| | β_3 | 0 | -0,0007 | 0,0221 | -0,3127 | 0,1151 |
| | β_4 | -0,0376 | -0,0668 | 0,2257 | -0,0083 | 1,4449 |
| | ϕ | 7,61 | 11,5658 | 21,6312 | 5,6770 | 27,4243 |
| 50 | β_0 | 0,591 | 0,6058 | 0,4885 | 0,1301 | 0,9250 |
| | β_1 | -0,138 | -0,1370 | 0,1201 | -0,4725 | 0,7550 |
| | β_2 | 0,06 | 0,0574 | 0,0140 | -0,0917 | 0,3461 |
| | β_3 | 0 | 0,0002 | 0,0045 | -0,0333 | 0,2207 |
| | β_4 | -0,0376 | -0,0440 | 0,0411 | -0,3644 | 0,3734 |
| | ϕ | 7,61 | 8,9131 | 3,6551 | 5,7999 | 13,0691 |
| 100 | β_0 | 0,591 | 0,5974 | 0,0216 | 0,2639 | 0,8228 |
| | β_1 | -0,138 | -0,1372 | 0,0586 | -0,4124 | 0,4738 |
| | β_2 | 0,06 | 0,0605 | 0,0072 | -0,1153 | 0,2104 |
| | β_3 | 0 | 0,0005 | 0,0018 | -0,0123 | 0,1454 |
| | β_4 | -0,0376 | -0,0400 | 0,0196 | -0,1634 | 0,3832 |
| | ϕ | 7,61 | 8,2085 | 1,3132 | 6,1695 | 10,6696 |
| 500 | β_0 | 0,591 | 0,5971 | 0,0043 | 0,3944 | 0,6566 |
| | β_1 | -0,138 | -0,1455 | 0,0120 | 0,1587 | 0,2643 |
| | β_2 | 0,06 | 0,0608 | 0,0009 | -0,0109 | 0,1106 |
| | β_3 | 0 | 0,0007 | 0,0004 | -0,0217 | 0,0630 |
| | β_4 | -0,0376 | -0,0386 | 0,0035 | -0,1365 | 0,0946 |
| | ϕ | 7,61 | 7,7159 | 0,2353 | 6,9975 | 8,8659 |

4.4 Tamanho dos Testes

Nesta seção apresentamos alguns resultados de simulação para avaliar os tamanhos dos testes. Comparamos os desempenhos dos três testes, isto é, o teste da razão de verossimilhanças, o teste escore e o teste Wald. Os desempenhos são avaliados em função da proximidade das probabilidades de rejeição da hipótese nula, sendo esta verdadeira (probabilidade do erro tipo I), aos respectivos níveis nominais dos testes. Adicionalmente, realizamos uma análise da qualidade da aproximação das três estatísticas dos testes por uma distribuição qui-quadrado adequada.

Tabela 4.2

Tamanhos observados dos testes:

| Teste $H_0 : \beta_3 = \beta_4 = 0$ vs $H_1 : \beta_3 \neq 0$ ou $\beta_4 \neq 0$; | | | | | | | | | |
|---|------|--------|------|------|--------|------|-------|--------|------|
| | n=20 | | | n=50 | | | n=100 | | |
| Nominal | RV | Escore | Wald | RV | Escore | Wald | RV | Escore | Wald |
| 10% | 16,8 | 13,0 | 20,5 | 15,2 | 12,7 | 17,5 | 13,1 | 11,9 | 14,4 |
| 5% | 9,9 | 6,2 | 13,2 | 8,5 | 6,4 | 10,6 | 7,1 | 5,9 | 8,2 |
| 1% | 2,9 | 0,8 | 5,2 | 2,3 | 1,0 | 3,7 | 1,6 | 1,0 | 2,4 |
| 2,5% | 1,6 | 0,3 | 3,7 | 1,2 | 0,4 | 2,4 | 0,8 | 0,4 | 1,3 |

Para todas as simulações, a hipótese nula considerada foi $H_0 : \beta_3 = \beta_4 = 0$ e os níveis nominais adotados foram 10%, 5%, 1% e 0,5%. O número de replicações de Monte Carlo foi novamente fixado em 1000.

Na Tabela 4.2 apresentamos os resultados de simulação supondo que os valores verdadeiros dos parâmetros são dados em (4.2) e $\phi = 7,61$ e a matriz de regressores X gerada das distribuições $U(0,1)$, $N(0,1)$, t_4 e $Exp(2)$, ou seja, cada covariável foi gerada de uma uma distribuição diferente. Para que não tivéssemos problemas na geração de cada covariável devido aos tamanhos diferentes de amostras, fixamos uma semente na geração e só modificamos o tamanho das amostras geradas. Para cada tamanho amostral e cada nível nominal considerado, calculamos as taxas de rejeição de cada teste, isto é, estimamos, via simulação, $P(\omega_1 > x_\alpha)$, $P(\omega_2 > x_\alpha)$ e $P(\omega_3 > x_\alpha)$, em que ω_1 , ω_2 e ω_3 são como vistos anteriormente e x_α representa o quantil de ordem $(1 - \alpha)$ da distribuição χ_2^2 .

Observa-se que os tamanhos empíricos dos testes aproximam-se dos níveis de significância nominais com o aumento do tamanho amostral, conforme esperado. Destaca-se ainda na Tabela 4.2, o fato de que os três testes são liberais, pois seus tamanhos empíricos estão sempre acima do nível nominal adotado. Verifica-se que o tamanho do teste escore é o que mais se aproxima do nível nominal. Por exemplo, para o nível nominal de 5% e tamanho amostral $n = 20$, seu valor é 6,2%, enquanto que os valores correspondentes para os testes de RV e Wald são 9,9% e 13,2%, respectivamente. O tamanho do teste de Wald apresentou-se mais distante dos níveis nominais. Para $n = 20$ e um nível nominal de 10%, encontramos um valor de 20,5% para o tamanho do teste.

Com o objetivo de avaliar a qualidade da aproximação das distribuições das três estatísticas por uma distribuição χ_2^2 , construímos gráficos de quantis (“q-q plot”) para as estatísticas da razão de verossimilhanças, escore e Wald, para cada tamanho amostral analisado na Tabela 4.2. Os gráficos foram construídos com os valores ordenados de cada estatística contra quantis correspondentes da distribuição χ_2^2 . A Figura 4.1 ilustra os resultados obtidos. Se a distribuição da estatística analisada apresenta uma boa aproximação pela distribuição χ_2^2 , espera-se que esse gráfico seja próximo de uma reta. Para facilitar essa avaliação, construímos uma reta tracejada considerada ideal, onde os pontos deveriam estar caso houvesse uma perfeita concordância entre os quantis amostrais da distribuição da estatística do teste analisado e os quantis correspondentes da distribuição de referência (χ_2^2). A análise da Figura 4.1 confirma os resultados encontrados para os tamanhos dos testes na Tabela 4.2. Verifica-se que os pontos do gráfico referentes à estatística escore se apresentam, de forma geral, mais próximo da reta tracejada do que os dos gráficos referentes ‘as estatísticas de razão de verossimilhanças e de Wald, que estão sempre acima dessa reta. Esse fato explica os tamanhos grandes dos testes encontrados para essas duas estatísticas.

Verifica-se também na Figura 4.1, que a estatística de Wald tem quantis amostrais bem maiores que os correspondentes quantis da distribuição χ_2^2 . Este fato ocorre para todos os tamanhos amostrais considerados nas análises, sendo mais agravante para $n = 20$. Também, podemos verificar que o efeito do tamanho da amostra (n) está intimamente relacionado com a qualidade da aproximação das estatísticas do teste pela distribuição de referência (χ_2^2). Com o aumento de n , verifica-se uma melhor aproximação das três

estatísticas pela distribuição χ_2^2 .

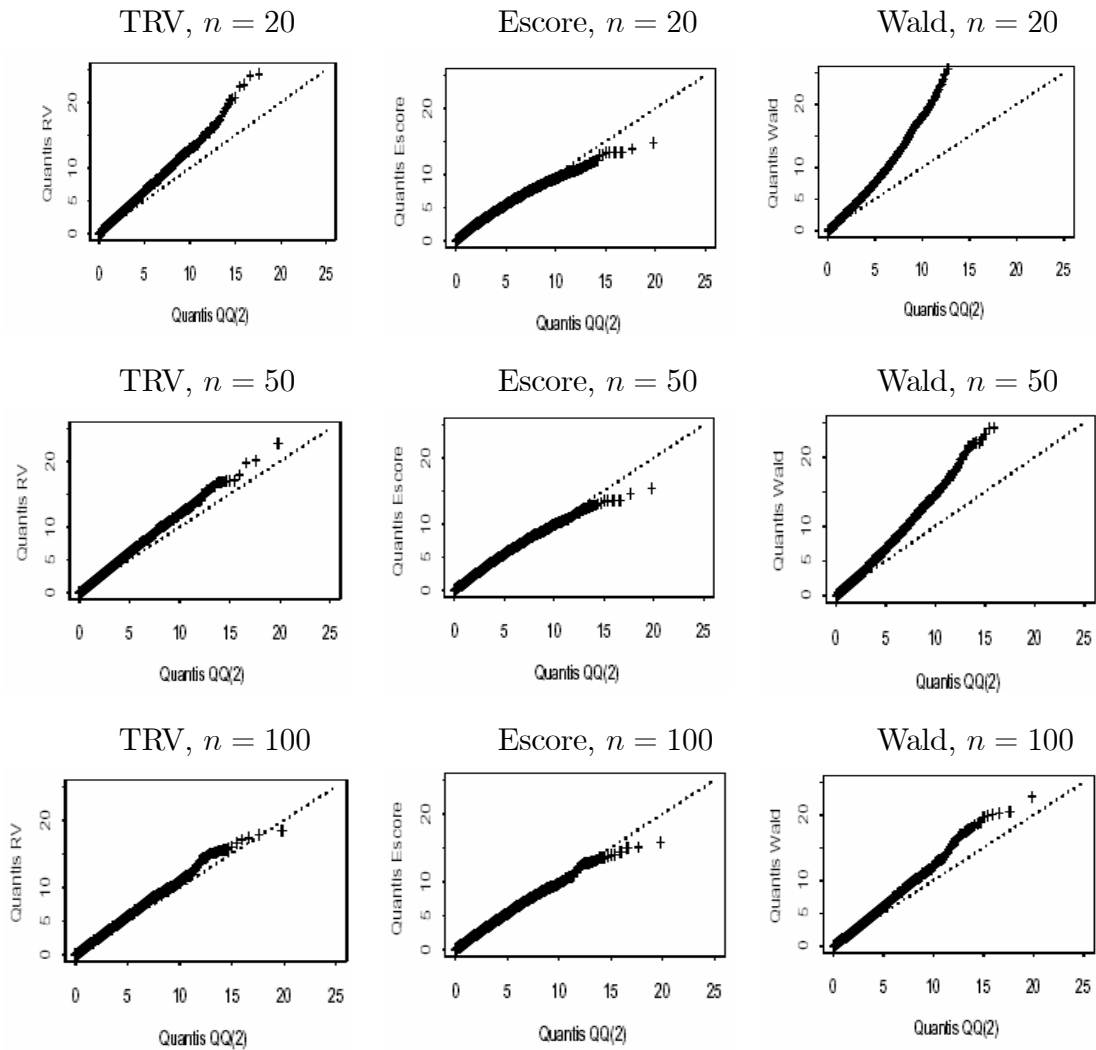


Figura 4.1: Gráficos (Q-Q-Plot) para verificar a aproximação dos testes a distribuição χ_2^2 , para a comparação dos tamanhos dos testes.

Capítulo 5

Regressão Beta: Procedimentos Práticos em Estimação

Neste capítulo discutimos um suplemento ao artigo de Smithson e Verkuilen (2005) que focaliza na estimação, via máxima verossimilhança, dos parâmetros de uma regressão beta e discutimos a implementação desse método de estimação em vários pacotes estatísticos.

5.1 Procedimentos Gerais

A maximização da função de verossimilhança pode ser obtida usando o Método de Newton-Raphson ou o Método de quasi-Newton (BFGS). Ferrari e Cribari-Neto (2004) usam o Método Escore de Fisher. Buckley e Kúscová (2002) utilizam a estimação via MCMC em WinBugs, que fornece uma densidade *a posteriori* bayesiana. Buckley (2002b) utiliza o *software* Stata e Paolino (2001) utiliza o *software* Gauss, ambos dos quais calculam as estimativas através de máxima verossimilhança. Nesta aplicação, estimamos os parâmetros da regressão beta usando o R, SPlus e SAS. Outros pacotes que disponibilizam tais procedimento de estimação são o SPSS, o Mathematica e o WinBUGS. O WinBUGS será usado na parte envolvendo o enfoque bayesiano.

Diferentes pacotes requerem mais ou menos informações do usuário. O SAS, por exemplo, requer apenas função de verossimilhança e calcula analiticamente as derivadas

pelo método de Newton, ou pode-se entrar com a função e as derivadas. Enquanto o Mathematica utiliza o método de Newton-Raphson, que requer as expressões para as derivadas. A maior diferença entre Newton-Raphson e quase-Newton (BFGS) é o número de iterações por repetições (maior para Newton-Raphson) e o número de iterações necessárias (maior para quase-Newton). Os domínios de convergências para algoritmos diferentes e, conseqüentemente, bons valores iniciais, serão diferentes. A região de confiança para o método de Newton implementado em SAS parece ser, particularmente, mais estável para problemas mais complexo e recomenda-se seu uso quando a convergência é um problema. Em geral, a velocidade de execução para a estimação de máxima verossimilhança é proporcional a $N + p^2$, onde N é o tamanho da amostra e p é o número de parâmetros.

Se o método de Newton ou quasi-Newton é usado, o erro padrão assintótico é usualmente estimado da matriz Hessian final.

São necessários bons valores iniciais para assegurar a convergência quando várias covariáveis estão incluídas no modelo. Ferrari e Cribari-Neto (2004) usam como valores iniciais os estimadores OLS da regressão da variável dependente transformada pela função de ligação. Por exemplo, se a função de ligação é a logito então os valores iniciais para o β , poderiam ser obtidos via:

$$\ln \left(\frac{y_k}{1 - y_k} \right) = x_k^T \beta_{OLS} + \varepsilon_k.$$

Em geral, o estimador OLS estima a faixa do modelo de ligação que pode ser tolerada, entretanto, é claro que os erros padrões serão ineficientes e, mais importante, é que a estrutura de variância será limitada a um intercepto. Infelizmente, para o nosso conhecimento, parece não haver nenhuma proposta semelhante para os valores iniciais para os coeficientes do sub-modelo de dispersão.

Uma outra forma de encontrar valores iniciais é começar com o modelo nulo, i.e., somente com o intercepto dos submodelos com os coeficientes β_0 e γ_0 , e utilizando o método dos momentos para prover valores iniciais para β_0 e γ_0 . Então a verossimilhança resultante calculada de β_0 e γ_0 é usada conjuntamente para calcular os valores iniciais próximos de 0 para β_1 e/ou γ_1 ($\beta_1 = \gamma_1 = 0.1$) e a máxima verossimilhança calculada

para β_0 , γ_0 , β_1 e/ou γ_1 é obtido. Estes valores são utilizados como valores iniciais para o próximo modelo (mais complexo). E assim por diante.

Outro problema, envolvendo estimação de parâmetros, é quando as covariáveis têm valores absolutos particularmente grandes, o que pode resultar em uma perda de precisão nas estimativas, ou criar problemas para o algoritmo de estimação. A saída, nestes casos, é re-escalonar as covariáveis a amplitudes menores.

5.2 Implementação em Pacotes Estatísticos

Nesta seção descrevemos, brevemente, as implementações de um modelo de regressão beta em R/SPlus e SAS. Todas as três implementações exigem do usuário introduzir termos e valores iniciais ao modelo, mas, não faz nenhuma demanda especial. Sempre que possível, recomenda-se ajustar o modelo de regressão beta em mais de um pacote, utilizando diferentes valores iniciais e diferentes métodos de otimização para conferir a convergência. Métodos com várias covariáveis no sub-modelo de dispersão, porém, são mais difíceis de estimar.

A implementação atual em R e SPlus fornecem o valor de máxima verossimilhança atingido pelo modelo, coeficientes, gradiente para os coeficientes, erro padrão assintótico estimado e o valor-p. Valores preditos e resíduos podem ser produzidos ou salvos no data-file. Que usam as duas funções “betareg” e “grad” e um pequeno conjunto de comandos. A versão SPlus usa a rotina nlminb quasi-newton para MLE e o vcov.nlminb no pacote MASS para calcular o Hessian e as estimativas assintóticas. A versão R usa a rotina optim para MLE com o método BFGS quasi-Newton. Uma versão de optim para SPlus também está na biblioteca MASS.

Em SAS o proc nlmixed e o proc glmix são utilizadas para ajustar modelos de regressão beta.

O procedimento glmix não permite um modelo com heterocedasticidade, mas gera estimativas razoáveis para a componente da variância do modelo com resposta beta.

As saídas do nlmixed são extensas. Informações de cada interação são detalhadas para o cálculo de cada parâmetro estimado além da verossimilhança, AIC, BIC e intervalos de

confiança assintóticos. Ainda é possível utilizar comandos adicionais para gerar valores preditos, resíduos e salvar saídas em formato RTF e Latex.

Para obter as estatísticas *bootstrap* e *Jackknife*, via SAS, podemos fazer *download* do código *jackboot* disponível nas páginas do SAS.

Capítulo 6

Modelo de Regressão Beta sob o Enfoque Bayesiano

6.1 Abordagem Bayesiana do Modelo

Neste capítulo, propomos uma análise do modelo de regressão beta sob o enfoque bayesiano.

Quando há necessidade do pesquisador de incorporar alguma informação sobre os parâmetros de interesse, pode-se utilizar técnicas Bayesianas. Na abordagem Bayesiana, os parâmetros são variáveis aleatórias e não mais constantes desconhecidas. Assim, a incerteza presente em um modelo dado θ é representada através de uma distribuição de probabilidade $p(\theta)$ sobre os possíveis valores do parâmetro θ desconhecido. O Teorema de Bayes

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

permite, então, incorporar a informação contida em um conjunto de dados x , produzindo uma descrição conjunta sobre a incerteza dos parâmetros de θ através da distribuição final $p(\theta|x)$.

6.2 Informação *a priori*

Na abordagem Bayesiana, o conhecimento a priori do pesquisador a respeito da interpretação dos parâmetros pode ser expresso utilizando prioris informativas. Neste trabalho, o

conhecimento a priori sobre os parâmetros de interesse foi expresso utilizando, primeiramente, prioris não informativas, isso significa que a distribuição a priori dos parâmetros deve descrever ausência de conhecimento. Em seguida, utilizamos prioris informativas, como descritas no Capítulo 7.

6.3 Função de Verossimilhança

Como discutido anteriormente, a variável resposta segue distribuição beta, $Y \sim Beta(p, q)$, em que os parâmetros p e q estão sob uma estrutura de regressão.

Supondo disponível uma amostra de dados de proporção, Y_1, \dots, Y_n , tal que, para cada k ,

$$Y_k \sim Beta(p_k, q_k),$$

onde, p_k e q_k são dados pelas expressões (3.6) e (3.7), como nos modelos estimados por máxima verossimilhança.

A função de verossimilhança é dada por

$$L(y, x | \mu_1, \dots, \mu_n, \phi) = \prod_{k=1}^n f(y_k; \mu_k, \phi),$$

sendo $f(y_k; \mu_k, \phi)$ dada em (3.9); assim chegamos a

$$\begin{aligned} & L(\mathbf{y}, \mathbf{x} | \mu_1, \dots, \mu_n, \phi) \\ &= [\Gamma(\phi)]^n \left[\prod_{k=1}^n [\Gamma(\mu_k \phi)] \right]^{-1} \left[\prod_{k=1}^n [\Gamma(\phi(1 - \mu_k))] \right]^{-1} \\ & \quad \times \prod_{k=1}^n \left[y_k^{(\mu_k \phi - 1)} \right] \prod_{k=1}^n \left[(1 - y_k)^{[\phi(1 - \mu_k) - 1]} \right], \end{aligned}$$

para ϕ conhecido, $\mu_k = \exp(x_k^T \beta) / [1 + \exp(x_k^T \beta)]$, x^T é o vetor de covariáveis e β é o vetor de parâmetros desconhecidos, que temos interesse em estimar. Ou seja, a função de

verossimilhança é dada por

$$\begin{aligned}
 L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \phi) &= [\Gamma(\phi)]^n \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(x_k^T \boldsymbol{\beta})}{1 + \exp(x_k^T \boldsymbol{\beta})} \phi \right) \right]^{-1} \\
 &\times \prod_{k=1}^n \left[\Gamma \left(\phi \left(1 - \frac{\exp(x_k^T \boldsymbol{\beta})}{1 + \exp(x_k^T \boldsymbol{\beta})} \right) \right) \right]^{-1} \\
 &\times \prod_{k=1}^n y_k \frac{\exp(x_k^T \boldsymbol{\beta})}{1 + \exp(x_k^T \boldsymbol{\beta})}^{\phi-1} \\
 &\times \prod_{k=1}^n (1 - y_k) \phi \left(1 - \frac{\exp(x_k^T \boldsymbol{\beta})}{1 + \exp(x_k^T \boldsymbol{\beta})} \right)^{-1},
 \end{aligned}$$

ou

$$\begin{aligned}
 L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \phi) &= [\Gamma(\phi)]^n \prod_{k=1}^n \left[\Gamma \left(\frac{\phi \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
 &\times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
 &\times \prod_{k=1}^n y_k \left[\frac{(\phi - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
 &\times \prod_{k=1}^n (1 - y_k) \left[\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]^{-1}.
 \end{aligned}$$

Se ϕ não for constante e $\phi_k = \exp(Z_k^T \boldsymbol{\gamma})$, onde Z_k^T é o vetor de covariáveis, relativo a k -ésima observação e $\boldsymbol{\gamma}$ é o vetor de parâmetros desconhecidos que temos interesse em estimar juntamente com $\boldsymbol{\beta}$ teremos que

$$\begin{aligned}
& L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \phi) \tag{6.1} \\
&= \prod_{k=1}^n [\Gamma(\exp(Z_k^T \gamma))] \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \exp(Z_k^T \gamma) \right) \right] \\
&\quad \times \prod_{k=1}^n \left[\Gamma \left(\exp(Z_k^T \gamma) \left(1 - \frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \right) \right) \right] \\
&\quad \times \prod_{k=1}^n y_k^{\left(\frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \exp(Z_k^T \gamma) - 1 \right)} \\
&\quad \times \prod_{k=1}^n (1 - y_k)^{\left[\exp(Z_k^T \gamma) \left(1 - \frac{\exp(x_k^T \beta)}{1 + \exp(x_k^T \beta)} \right) - 1 \right]},
\end{aligned}$$

ou

$$\begin{aligned}
& L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\gamma}) \tag{6.2} \\
&= \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr}))] \\
&\quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
&\quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
&\quad \times \prod_{k=1}^n y_k^{\left[\frac{(\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr}) - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]} \\
&\quad \times \prod_{k=1}^n (1 - y_k)^{\left[\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} - 1 \right]}.
\end{aligned}$$

Como sabemos, a abordagem Bayesiana requer a determinação de prioris para os parâmetros de interesse. Iniciamos o estudo com prioris vagas, como, por exemplo, distribuições normais com médias conhecidas e grandes variâncias.

6.3.1 Estimação Bayesiana com Prioris vagas

a) Supondo ϕ desconhecido, mas constante:

i) **Prioris normais:** Considere para os parâmetros de interesse, β' s prioris $N(0; 100)$, ou seja, com média igual a 0 e variância igual a 100, vagas. Assim, temos a *posteriori*

$$\begin{aligned}
 & p(\beta_0, \beta_1, \dots, \beta_p | \phi, y, \mathbf{X}) \tag{6.3} \\
 = & L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \phi) p(\beta_0) \dots p(\beta_p) \propto [\Gamma(\phi)]^n \\
 & \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
 & \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
 & \times \prod_{k=1}^n y_k \left[\frac{(\phi - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
 & \times \prod_{k=1}^n (1 - y_k) \left[\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]^{-1} \\
 & \times \exp \left(\frac{-\beta_0^2 - \beta_1^2 - \dots - \beta_p^2}{20000} \right).
 \end{aligned}$$

Usando a regra de Bayes para combinar estas distribuições *a priori* com a função de verossimilhança dada em (6.1), a seguinte distribuição *a posteriori* conjunta para os β' s

é obtida

$$\begin{aligned}
& p(\beta_0, \beta_1, \dots, \beta_p | \phi, y, \mathbf{X}) \\
& \propto L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \phi) p(\beta_0) \dots p(\beta_p) \propto [\Gamma(\phi)]^n \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{(\phi - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_0^2 - \beta_1^2 - \dots - \beta_p^2}{20000} \right).
\end{aligned}$$

Como essa distribuição a *posteriori* é intratável analiticamente, será proposto o procedimento de *Metropolis-Hastings* para retirar amostras da distribuição a *posteriori*. O nosso interesse direto é fazer inferência em cada um dos parâmetros, portanto devemos encontrar a distribuição a *posteriori* marginal de cada um dos mesmos. A distribuição a *posteriori* pode ser obtida integrando a *posteriori* conjunta em relação a todos os outros parâmetros. Por exemplo, a *posteriori* marginal para β_0 seria dada por:

$$p(\beta_0 | \beta_1, \dots, \beta_p, \phi, \mathbf{y}, \mathbf{X}) \propto \int \int \dots \int p(\beta_0, \beta_1, \dots, \beta_p | \phi, \mathbf{y}, \mathbf{X}) d\beta_1 \dots d\beta_p d\phi.$$

Para isso, é importante obter as distribuições a *posteriori* condicionais de cada parâmetro, a seguir temos um exemplo de uma distribuição a *posteriori* marginal.

Dado $\beta_1, \beta_2, \dots, \beta_p$, a distribuição a *posteriori* condicional de β_0 é

$$\begin{aligned}
& p(\beta_0 | \beta_1, \beta_2, \dots, \beta_p, \phi, y, X) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\phi \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{(\phi - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_0^2}{20000} \right).
\end{aligned}$$

Gerar uma amostra de $\beta_0 | \beta_1, \beta_2, \dots, \beta_p, \phi, y, X$ não é fácil, uma vez que a distribuição de $\beta_0 | \beta_1, \beta_2, \dots, \beta_p, \phi, y, X$ não pertence a uma família de distribuição conhecida. O mesmo ocorre com as demais condicionais. Desta forma, é utilizado um procedimento Monte Carlo em Cadeias de Markov-MCMC (Carlin e Gelfand, 1991), mais precisamente *Metropolis Hastings*, para gerar amostras aproximadas destas distribuições a *posteriori*.

As estimativas são dadas pela média das cadeias geradas a partir da distribuição a *posteriori* de cada parâmetro do modelo, obtidas via Método MCMC.

b) Supondo que ϕ não seja o mesmo para todos os Y'_k s e que exista uma relação funcional de ϕ_k com as covariáveis, através de uma estrutura de regressão,

$$\phi_k = \exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr}),$$

onde, $z_{kj} = x_{ki}, j = 1, 2, \dots, r$ e $i = 1, 2, \dots, p, r < p$, ou seja, o conjunto de covariáveis $\mathbf{Z} = (z_1, \dots, z_r)$ é um sub-conjunto do conjunto de covariáveis $\mathbf{X} = (x_1, \dots, x_p)$, que explica a dispersão do modelo. Logo, poderíamos trabalhar com:

i) **Prioris normais:** Considere para os parâmetros de interesse, β' s e γ' s, prioris $N(0; 100)$, ou seja, com média igual a 0 e variância igual a 100. Teremos a *posteriori* dada por

$$\begin{aligned}
& p(\beta_0, \beta_1, \dots, \beta_p, \gamma_0, \dots, \gamma_r | \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
&= L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \phi) p(\beta_0) \dots p(\beta_p) p(\gamma_0) \dots p(\gamma_r) \\
&\propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_r z_{kr}))] \\
&\quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
&\quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
&\quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
&\quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]^{-1} \\
&\quad \times \exp \left(\frac{-\beta_0^2 - \beta_1^2 - \dots - \beta_p^2 - \gamma_0^2 - \gamma_1^2 - \dots - \gamma_r^2}{20000} \right).
\end{aligned}$$

Como essa distribuição a *posteriori* é intratável analiticamente, novamente será proposto o procedimento de *Metropolis-Hastings* para retirar amostras da distribuição a *posteriori*. Para isso, é importante obter as distribuições a *posteriori* condicionais de cada parâmetro.

Dado $\beta_0, \beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_r$, a distribuição a *posteriori* condicional de γ_0 é

$$\begin{aligned}
& p(\gamma_0 | \beta_0, \beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_r, \mathbf{y}, \mathbf{X}) \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_r z_{kr}))]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\gamma_0^2}{20000} \right)
\end{aligned}$$

Como discutimos anteriormente, gerar uma amostra de $\gamma_0 | \beta_0, \beta_1, \beta_2, \dots, \beta_7, \gamma_1, \dots, \gamma_2, y, X$, também não é fácil, uma vez que a distribuição de $\gamma_0 | \beta_0, \beta_1, \beta_2, \dots, \beta_7, \gamma_1, \dots, \gamma_2, y, X$ não pertence a uma família de distribuição conhecida. O mesmo ocorre com as demais condicionais. Novamente devemos utilizar um procedimento MCMC para gerar amostras aproximadas destas distribuições a *posteriori*.

6.3.2 Estimação Bayesiana com Prioris Informativas

Uma vantagem da metodologia bayesiana, como foi dito anteriormente, é a possibilidade do pesquisador de, efetivamente, incorporar informações à análise. Esta incorporação de informação pode ser feita através de prioris informativas. Nesta seção apresentamos a análise bayesiana do modelo de regressão beta considerando prioris informativas para os parâmetros de interesse.

Uma forma de determinar a densidade a priori, juntamente com os hiperparâmetros, na situação onde vários históricos de dados estão disponíveis, é estimando, em cada histórico, os parâmetros de interesse, criando assim, uma seqüência de estimadores. Desse modo teríamos uma seqüência de estimadores. Em seguida buscaríamos uma distribuição que

se adere a esta seqüência. Desta forma, teríamos então, constituído uma densidade a priori informativa para os parâmetros do modelo sob pesquisa.

Supondo que ϕ não seja o mesmo para todos os Y'_k s e que exista uma relação funcional de ϕ_k com as covariáveis, através de uma estrutura de regressão,

$$\phi_k = \exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr}),$$

poderíamos trabalhar com:

i) **Prioris normais:** Suponha densidades normais para os parâmetros, ou seja, considere para os parâmetros de interesse, β' s, prioris $N(a_i; b_i^2)$ e para os parâmetros γ' s prioris $N(c_i; d_i^2)$, onde $a_i; b_i^2, c_i; d_i^2$ são encontradas via máxima verossimilhança, usando o histórico como descrito anteriormente. Ou seja, temos

$$\begin{aligned} \beta_0 &\sim N(a_0; b_0^2), \beta_1 \sim N(a_1; b_1^2), \dots, \beta_p \sim N(a_p; b_p^2), \\ \gamma_0 &\sim N(c_0; d_0^2), \gamma_1 \sim N(c_1; d_1^2), \dots, \gamma_r \sim N(c_r; d_r^2), \end{aligned}$$

tem-se que as f.d.p.'s desses parâmetros são

$$\begin{aligned} p(\beta_0) &= \frac{1}{\sqrt{2\pi b_0^2}} \exp\left(\frac{-\beta_0^2 + 2a_0\beta_0 - a_0^2}{2b_0^2}\right), \\ p(\beta_1) &= \frac{1}{\sqrt{2\pi b_1^2}} \exp\left(\frac{-\beta_1^2 + 2a_1\beta_1 - a_1^2}{2b_1^2}\right), \\ &\vdots \\ p(\beta_p) &= \frac{1}{\sqrt{2\pi b_p^2}} \exp\left(\frac{-\beta_p^2 + 2a_p\beta_p - a_p^2}{2b_p^2}\right), \\ p(\gamma_0) &= \frac{1}{\sqrt{2\pi d_0^2}} \exp\left(\frac{-\gamma_0^2 + 2c_0\gamma_0 - c_0^2}{2d_0^2}\right), \\ p(\gamma_1) &= \frac{1}{\sqrt{2\pi d_1^2}} \exp\left(\frac{-\gamma_1^2 + 2c_1\gamma_1 - c_1^2}{2d_1^2}\right), \\ &\vdots \\ p(\gamma_r) &= \frac{1}{\sqrt{2\pi d_r^2}} \exp\left(\frac{-\gamma_r^2 + 2c_r\gamma_r - c_r^2}{2d_r^2}\right). \end{aligned}$$

Teremos a *posteriori* dada por

$$\begin{aligned}
& p(\beta_0, \beta_1, \dots, \beta_p, \gamma_0, \gamma_1, \dots, \gamma_r | \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr}))] \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{i=1}^n y_i \left[\frac{(\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr}) - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
& \quad \times \prod_{i=1}^n (1 - y_i) \left[\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} - 1 \right] \\
& \quad \times \exp \left(\frac{-\beta_0^2 + 2a_0 \beta_0}{2b_0^2} \right) \exp \left(\frac{-\beta_1^2 + 2a_1 \beta_1}{2b_1^2} \right) \dots \exp \left(\frac{-\beta_p^2 + 2a_p \beta_p}{2b_p^2} \right) \\
& \quad \times \exp \left(\frac{-\gamma_0^2 + 2c_0 \gamma_0}{2d_0^2} \right) \exp \left(\frac{-\gamma_1^2 + 2c_1 \gamma_1}{2d_1^2} \right) \dots \exp \left(\frac{-\gamma_r^2 + 2c_r \gamma_r}{2d_r^2} \right).
\end{aligned}$$

Como essa distribuição a *posteriori* é, também, intratável analiticamente, novamente será proposto o procedimento de *Gibbs Sampling* para retirar amostras da distribuição a *posteriori*. Para isso, é importante obter as distribuições a *posteriori* condicionais de cada parâmetro.

i) Distribuição a *posteriori* condicional de β_0 dado $\beta_1, \dots, \beta_p, \gamma_0, \gamma_1, \dots, \gamma_r$.

Dado $\beta_1, \dots, \beta_p, \gamma_0, \gamma_1, \dots, \gamma_r$, a distribuição a *posteriori* condicional de β_0 é

$$\begin{aligned}
& p(\beta_0 | \beta_1, \dots, \beta_p, \gamma_0, \gamma_1, \dots, \gamma_r, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_0^2 + a_0 \beta_0}{2b_0^2} \right).
\end{aligned}$$

ii) Analogamente, temos a distribuição a *posteriori* condicional de β_1 dado $\beta_0, \beta_2, \dots, \beta_p, \gamma_0, \gamma_1, \dots, \gamma_r$.

Dado $\beta_0, \beta_2, \dots, \beta_p, \gamma_0, \gamma_1, \dots, \gamma_r$, a distribuição a *posteriori* condicional de β_1 é

$$\begin{aligned}
& p(\beta_1 | \beta_0, \beta_2, \dots, \beta_p, \gamma_0, \gamma_1, \dots, \gamma_r, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_r z_{kr})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_p x_{kp})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_1^2 + a_1 \beta_1}{2b_1^2} \right).
\end{aligned}$$

Todas as demais condicionais são análogas as essas.

Como discutimos anteriormente, gerar uma amostra de

$\beta_1 | \beta_0, \beta_2, \dots, \beta_7, \gamma_0, \gamma_1, \dots, \gamma_2$, também não é fácil, uma vez que a distribuição de $\beta_1 | \beta_0, \beta_2, \dots, \beta_7, \gamma_0, \gamma_1, \dots, \gamma_2$ não pertence a uma família de distribuição conhecida. O mesmo ocorre com as demais condicionais. Novamente devemos utilizar um procedimento MCMC para gerar amostras aproximadas destas distribuições *a posteriori*, *mais precisamente o Metropolis-Hastings*.

As estimativas são dadas pela média das cadeias geradas a partir da distribuição *a posteriori* de cada parâmetro do modelo, obtidas via Método MCMC.

Capítulo 7

Aplicação: O Caso *Charter Schools*

7.1 Introdução

Neste capítulo apresentamos uma aplicação do modelo de regressão beta a um conjunto de dados reais sob o enfoque clássico e sob o enfoque bayesiano.

Todos os cálculos computacionais para os ajustes dos modelos sob o enfoque clássico são realizados usando a linguagem de programação R. Para os modelos sob o enfoque bayesiano, os cálculos são realizados via *software* WinBugs.

A estimação dos parâmetros dos modelos sob o enfoque clássico foi realizada usando o algoritmo quase-Newton (BFGS) e a escolha dos valores iniciais para os parâmetros desconhecidos segue a sugestão de Ferrari e Cribari-Neto (2004), apresentada na seção 2.2.1.

7.2 Estatística e Marketing: O Marketing na Educação

O interesse deste estudo é em uma área não muito rica na qualidade de análises e com várias barreiras intrínsecas, que é a área educacional, em particular, nas políticas administrativas. As Charter Schools se destacaram na década passada em todos os Estados Unidos devido a sua maneira não ortodoxa de administração, fazendo uso fortemente

do marketing, com vários analistas modelando a sua participação no mercado.

As *Charter Schools* (escolas terceirizadas) têm sido um sucesso devido a vários fatores, entre os mais importantes, citamos:

- A possibilidade de haver vários financiadores para uma mesma escola;
- Aplicações de vários fundos possíveis: poder público, particular com ou sem fins lucrativos e organizações mistas;
- A grande penetração, no mercado, desse tipo de escola.

Há muitos tipos de reformas escolares sendo implementadas nos Estados Unidos e muitas delas difundem suas atividades, como no caso das *Charter Schools*.

As *Charter Schools* se identificam como empresas, na busca do melhor retorno de divisas possível, na adequação de gastos, ou na melhor aplicação de recursos e na busca de mais e melhores "clientes" (alunos) e financiadores, para poderem se consolidar cada vez mais no mercado e dessa forma, voltam-se para o marketing (através de campanhas publicitárias) e para a competição por uma maior fatia do mercado.

Dessa forma, as *Charter Schools* agendam muitos dados financeiros, que têm estimulado muitos debates de como deve ser a linha a qual a educação deve seguir.

As *Charter Schools* surgiram por volta dos anos 90 e a sua participação no mercado educacional cresce exponencialmente desde então. Elas têm um importante papel no sistema de educação pública nacional americana. Em maio de 2003, havia cerca de 2700 *Charter Schools*, operando nos 51 estados e no Distrito de Columbia, educando acima de 684000 alunos. Como é bem conhecido, *Charter Schools* são escolas que geralmente empregam dinheiro público e/ou particular de maneira que tipicamente tem menos restrições, regulam e administram seu próprio comportamento.

Assim, na busca de cada vez mais liberdade de atuação financeira e administrativa, as *Charter Schools* ficam responsáveis pelo seu desempenho. Sendo assim, elas se tornaram extremamente atrativas, passando a ser um ponto fundamental das pesquisas na área educacional americana.

Desta forma a interação *marketing* e educação se tornou um assunto muito interessante, no que se diz respeito ao setor administrativo educacional. O interesse desta pesquisa está em modelar a participação de mercado (*Market Share*) das *Charter Schools*.

Como estamos nos referindo a proporções, e tentamos explicar tal variável em função

de outras variáveis, temos assim, a justificativa de empregar um modelo de regressão beta.

Para determinar os efeitos da variação em leis de terceirização na participação de mercado que atinge instituições terceirizadas, foram reunidos dados de várias fontes. Nossas covariáveis foram construídas utilizando dados de escolas públicas regulares, escolas alternativas, escolas terceirizadas e escolas privadas, coletados no Centro Estatístico Nacional para Educação (CER) e simplesmente dividimos o número de escolas terceirizadas nos estados e os respectivos tipos de *marketing* empregados em cada estado.

Para a variável resposta (y), participação de mercado (*Market Share*), utilizamos o *ranking* anual de leis de terceirizações do Centro de Reformas para a Educação, de acordo com sua força. Em 2000 o CER pediu um painel a especialistas no assunto, que classificasse as leis de terceirizações de cada estado em várias dimensões.

O painel criou uma escala de zero a cinco, onde a maior nota corresponde ao ambiente que fortemente sustenta a expansão das escolas terceirizadas. Desta forma, se a provisão estadual é abaixo de um particular critério for avaliado como zero, isto indica que neste estado o desenvolvimento das escolas terceirizadas é extremamente restrito, enquanto a nota cinco sugere que as *Charter Schools* têm uma alta chance de prosperar nesse dado estado.

Usamos o *ranking* da CER de 37 regulamentos de *Charter Schools* das existentes em abril de 2000, para construir variáveis independentes, cada uma delas classificadas de zero a cinco, onde o maior número correspondente à lei que é mais sustentável a produção das *Charter Schools*. *As variáveis independentes são:*

x_1 : Representantes financeiros facilitando o processo de concessão de terceirizações. Estados que limitam a autorização para mudanças de escolas recebem o ranking mais baixo.

x_2 : A variedade de *Charter Schools*, pretendentes à permissão. A nota mais alta é associada com o potencial dos diversos pretendentes em conciliar interesses.

x_3 : O número de debates que avaliam o processo de criação de uma Charter School com o julgamento para qualquer nova escola.

Foram incluídas algumas variáveis adicionais no modelo como variáveis de controle:

x_4 : A proporção de escolas públicas naquele estado.

x_5 : A quantidade de anos desde que a lei foi adotada. Esta variável é limitada de 0

para Oklahoma e Oregon a 8 para Minneasota.

x_6 : Dimensões da população, que representa a proporção da população daquele estado em relação a população nacional.

x_7 : Medidas Financeiras: Nível estadual do gasto por aluno. Nós esperamos que quanto mais dinheiro o estado gasta por criança inscrita na escola pública tradicional, haverá menos financiadores dispostos a abrigar escolas particulares.

7.3 Inferência Clássica

Como visto anteriormente, a variável resposta, y_k , corresponde ao *Market Share_k* e o interesse reside na modelagem da proporção de *Charter Schools* em cada estado. A análise original deste conjunto de dados foi realizada por Buckley (2003), consistindo no ajuste de um modelo de regressão beta com sete variáveis explicativas, sem utilizar transformação na resposta, mas sob a óptica bayesiana e utilizando o software Stata.

É conveniente aqui uma melhor nomenclatura dos modelos, que apresentamos abaixo. Trabalhamos com duas funções de ligação: a logito e a probito, com ambas ajustamos os modelos com todas as covariáveis disponíveis e depois com apenas as covariáveis significativas. Os efeitos das funções de ligações nos modelos ajustados foram comparados.

Modelo 1: Modelo de regressão beta com função de ligação logito e todas as covariáveis.

Modelo 2: Modelo de regressão beta com função de ligação logito e todas as covariáveis, reajustado sem a quarta observação, pois esta foi considerada ter sido coletada erroneamente.

Modelo 3: Modelo de regressão beta com função de ligação probito e todas as covariáveis.

Modelo 4: Modelo de regressão beta com função de ligação logito com somente as covariáveis significativas.

Modelo 5: Modelo de regressão beta com função de ligação probito com somente as covariáveis significativas.

Nesse primeiro momento a abordagem será sob o enfoque clássico. Analisaremos esse conjunto de dados fazendo uso do modelo de regressão beta. Para isso, assumiremos que a variável resposta y_k (*Market Share_k*) tem uma distribuição beta com média μ_k e usando as demais variáveis como explicativas (x_1, \dots, x_7) . Inicialmente, consideraremos o modelo.

$$g(\mu_k) = \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7}$$

em que g representa a função de ligação logito ou probito. Na verdade ajustamos modelos utilizando cada uma das duas funções de ligação.

7.4 Estimação dos Parâmetros

Nesta seção apresentamos os resultados da estimação dos parâmetros via máxima verossimilhança. Para esta estimação foi utilizado o pacote *betareg* da linguagem R e o software SAS, os resultados apresentados a seguir foram consoantes em ambos os softwares.

7.4.1 Modelo 1

Os resultados de estimação para o modelo utilizando a função de ligação logito são dados na Tabela 7.1. Verifica-se que os níveis significativos das variáveis explicativas, *A variedade de Charter Schools*, *Número de Debates*, *Proporções de Escolas Públicas*, estão abaixo de 10%, indicando que essas variáveis são estatisticamente importantes para explicar a variável resposta “o *Market Share* Médio”. No entanto, os níveis descritivos das variáveis explicativas: *Representantes Financeiros*, *Tempo*, *Dimensões da População* e *Medidas Financeiras* são de 0,6183; 0,6202; 0,1555 e 0,2572, respectivamente, o que indica, inicialmente, que essas variáveis podem ser retiradas do modelo, desde que sejam mantidas as outras quatro.

Tabela 7.1

Estimativas dos parâmetros do estudo de Charter Scholls.

| Parâmetro | Variáveis | Valores Estimados | valor-p | Intervalos de Confiança |
|-----------|-------------------------|-------------------|---------|-------------------------|
| β_0 | | -0,1180 | 0,7941 | $[-1, 03; 0, 79]$ |
| β_1 | Repr.. fin. | -0,0339 | 0,6183 | $[-0, 17; 0, 10]$ |
| β_2 | Var. de Charter Schools | -0,0836 | 0,0652 | $[-0, 17; 0, 008]$ |
| β_3 | Nº de debates | 0,1079 | 0,0125 | $[0, 02; 0, 19]$ |
| β_4 | prop. de escolas pub. | 0,6497 | 0,0263 | $[0, 06; 1, 23]$ |
| β_5 | quantidade de anos | -0,0169 | 0,6202 | $[-0, 08; 0, 05]$ |
| β_6 | dim. da pop. | -0,4000 | 0,1555 | $[-0, 96; 0, 16]$ |
| β_7 | medidas financeiras | -0,7400 | 0,2572 | $[-2, 07; 0, 57]$ |
| ϕ | | 17,2951 | | $[10, 56; 24, 02]$ |

Para o modelo 1 o pseudo R_p^2 foi de 0,2129, a razão de verossimilhanças foi 13,59099.

Com o leverage generalizado, temos

$$GL[4; 4] = 0,5367 > 0,3137 = 2k/n$$

e

$$GL[30; 30] = 0,3350 > 0,3137 = 2k/n,$$

onde k é o número de parâmetros e n o tamanho da amostra. O que nos diz que as observações de número 4 e 30 são influentes.

Para uma análise gráfica dos resíduos apresentamos os gráficos 7.1 e 7.2. A *deviance* residual deste modelo é igual a 48,9021, o que pode ser considerado um valor razoável, porém, uma pouco maior do que o esperado, que seria em torno de 48.50.

Os sete gráficos que compõe a Figura 7.1 envolvem:

i) Gráfico dos resíduos da deviance vs os índices das observações. Por este gráfico podemos notar que a observações 4 têm resíduos grandes.

ii) Gráfico dos resíduos padronizados vs os índices das observações. Por este gráfico podemos notar, também, que a observações 4 tem resíduo grande.

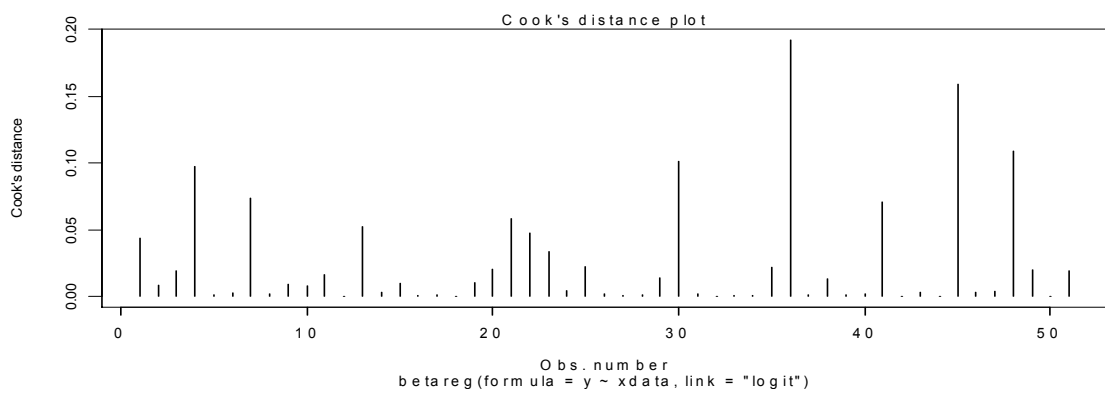
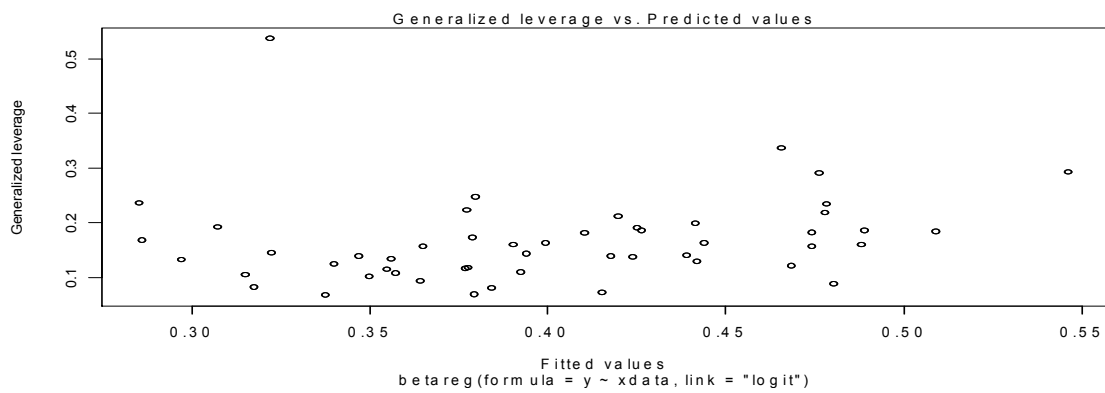
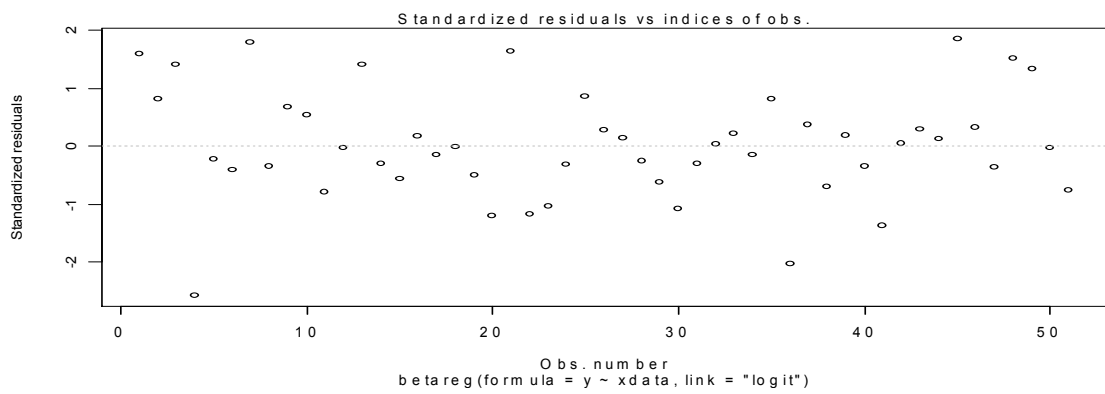
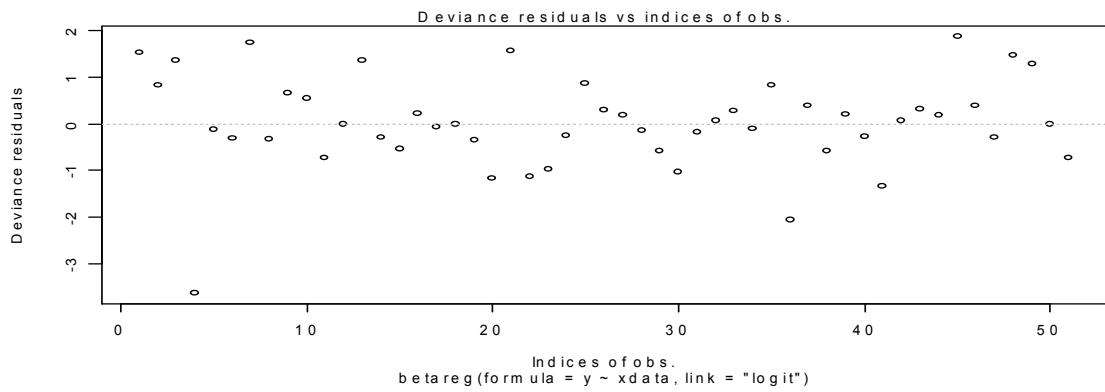
iii) Gráfico dos leverages generalizados vs os valores preditos. Por este gráfico podemos notar que a quarta observação é muito influente na obtenção dos valores preditos.

iv) Gráfico da distância de Cook vs os índices. Por este gráfico podemos notar que as observações de número 36 e 45 são muito influentes no processo de estimação.

v) Gráfico dos resíduos ordinários vs os valores preditos. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar pontos atípicos como a observação de número 4.

vi) Gráfico dos resíduos ordinários vs o preditor linear estimado. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar se a função de ligação é adequada ao ajuste do modelo. A única observação que se afasta é a observação de número 4. Desse modo podemos afirmar que a função de ligação é adequada.

vii) Gráfico de probabilidades meio – normal com envelopes. Esse gráfico nos ajuda a decidir se as respostas observadas são consistentes com o modelo ajustado. Nesse caso, notamos um certo desvio, o que evidencia uma inconsistência dos dados em relação ao ajuste.



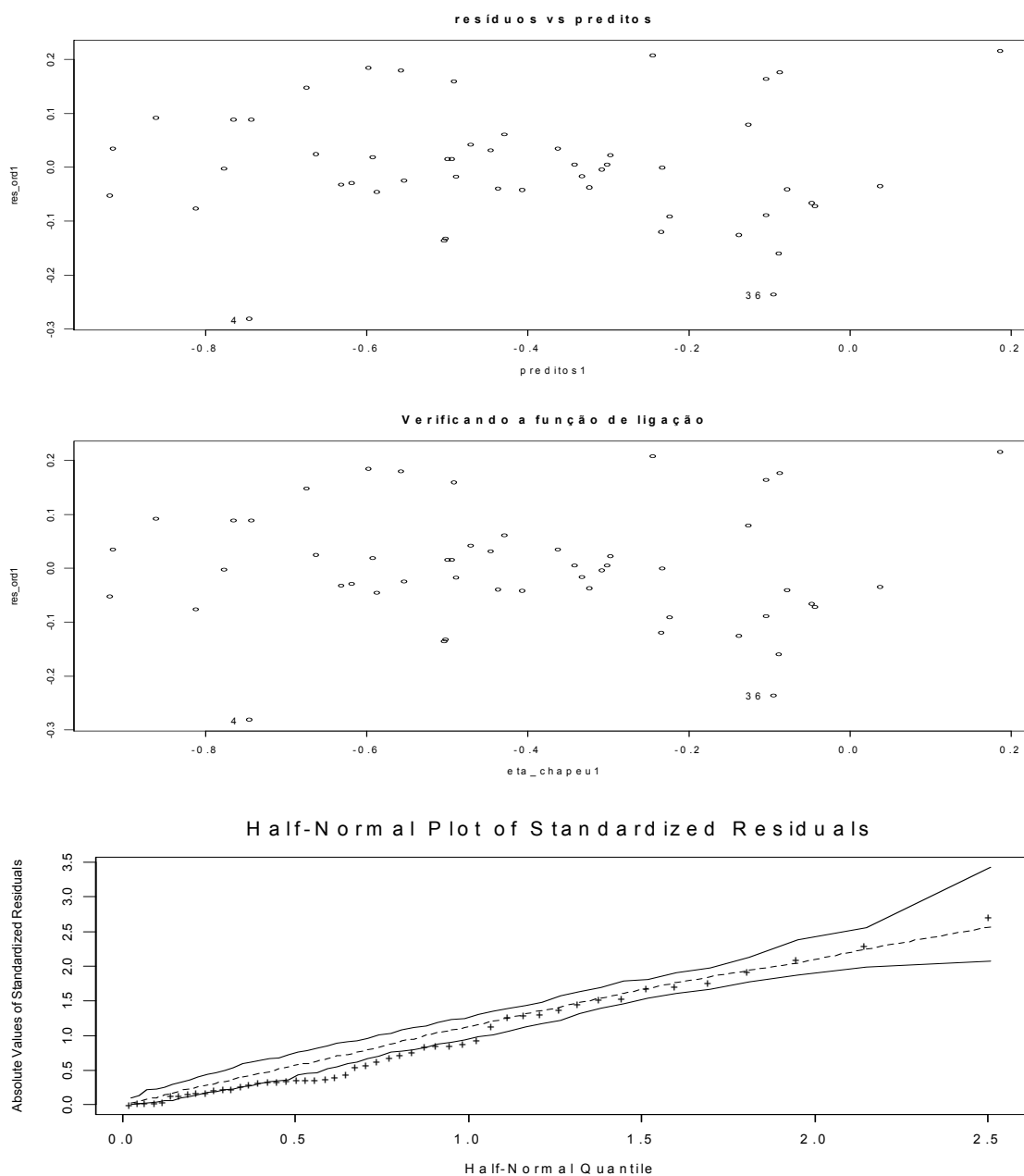


Figura 7.1: Gráficos de diagnósticos para o modelo 1 referente aos dados do Caso Charter Schools.

Vimos que a quarta observação tem resíduo grande e é um ponto influente, ou seja, é um ponto atípico e influente. Realizamos uma análise descritiva dos dados e verificamos que o valor da variável resposta para a quarta observação é bastante destoante das demais, como podemos ver na Figura 7.2. Não demos atenção especial para os pontos 7, 13, 45 e 48

devido a não serem pontos atípicos, ou seja não têm resíduos grandes nem são influentes.

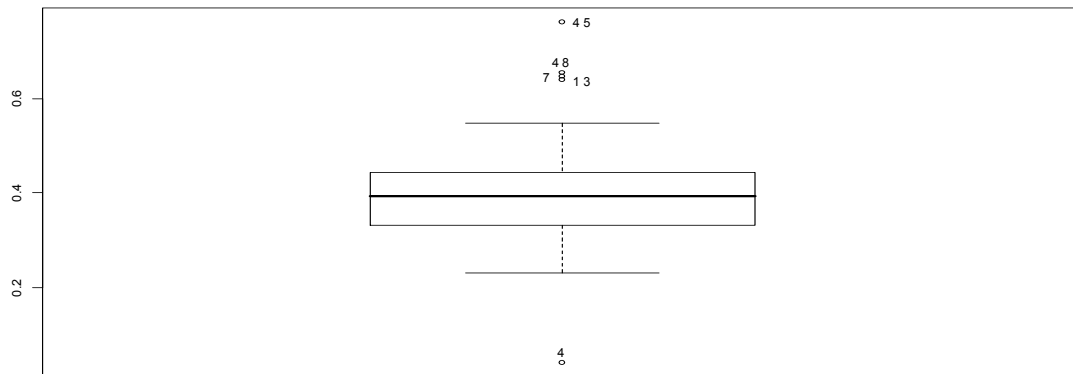


Figura 7.2: Gráfico Boxplot da variável resposta.

Nosso próximo passo é reajustar o modelo 1 sem a presença da quarta observação, considerando que a mesma tenha sido coletada erroneamente.

7.4.2 Modelo 2

O modelo 2 é o modelo apresentado na seção anterior ajustado sem a quarta observação. Os resultados da estimação para o modelo utilizando a função de ligação logito são dados na Tabela 7.2. Verifica-se que o nível significância das variáveis explicativas, *Variedade de Charter Schools*, *Números de Debates*, *Proporções de Escolas Públicas*, *Dimensões*, estão próximos de zero, indicando que essas variáveis são estatisticamente importantes para explicar a variável resposta *Média do Market Share*. No entanto, níveis descritivos das variáveis explicativas: *Representantes Financeiros*, *Tempo e Medidas Financeiras* são de 0,5053; 0,8266 e 0,2142, respectivamente, o que indica, inicialmente, que essas variáveis podem ser retiradas do modelo, desde que sejam mantidas as outras quatro. É importante ressaltar que o modelo ajustado sem a quarta observação se tornou muito mais plausível, note que com essa observação tínhamos apenas duas covariáveis significativas e, sem essa observação passamos a ter quatro covariáveis significativas, além de aumentar muito o valor do pseudo R^2 .

Tabela 7.2

Estimativas dos parâmetros do estudo de Charter Scholls.

| Parâmetro | | Valores Estimados | valor-p | Intervalos de Confiança |
|-----------|-------------------------|-------------------|---------|-------------------------|
| β_0 | | -0,08405 | 0,8256 | [-0, 8536; 0, 6855] |
| β_1 | Repr.. fin. | -0,03818 | 0,5053 | [-0, 1538; 0, 0774] |
| β_2 | Var. de Charter Schools | -0,06624 | 0,0854 | [-0, 1439; 0, 0114] |
| β_3 | Nº de debates | 0,08311 | 0,0233 | [0, 0091; 0, 1570] |
| β_4 | prop. de escolas pub. | 0,55897 | 0,0238 | [0, 0598; 1, 0580] |
| β_5 | qunatidade de anos | 0,00637 | 0,8266 | [-0, 0523; 0, 0650] |
| β_6 | dim. da pop. | -0,49821 | 0,0366 | [-0, 9792; -0, 0171] |
| β_7 | medidas financeiras | -0,68497 | 0,2142 | [-1, 7978; 0, 4279] |
| ϕ | | 24,53339 | | [14, 8194; 34, 2473] |

Para o modelo 2 o pseudo R_p^2 foi de 0,2308, a Razão de verossimilhanças foi 13,10169.

Com o leverage generalizado, temos

$$GL[35, 35] = 0,2995 < 0,32 = 2k/n$$

e

$$GL[44, 44] = 0,2986 < 0,32 = 2k/n,$$

onde k é o número de parâmetros e n o tamanho da amostra. O que nos diz que as observações de número 35 e 44 não são influentes. A *deviance* residual deste modelo é igual a 48,75.

Para uma análise gráfica dos resíduos apresentamos os gráficos da Figura 7.3. Onde, os sete gráficos que compõe a Figura 7.3 envolvem:

i)Gráfico dos resíduos da deviance vs os índices das observações. Por este gráfico podemos notar que as observações 35, 44 e 47 têm resíduos grandes.

ii)Gráfico dos resíduos padronizados vs os índices das observações. Por este gráfico podemos notar, também, que as observações 35, 44 e 47 têm resíduos grandes.

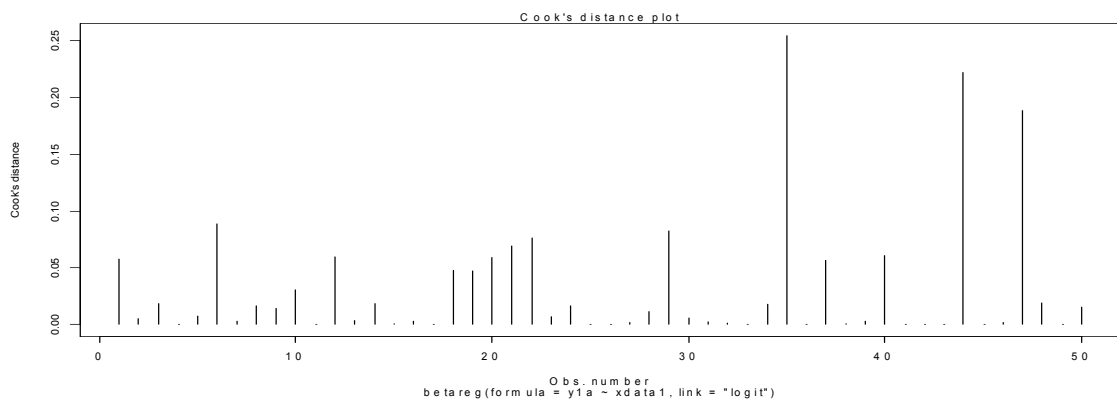
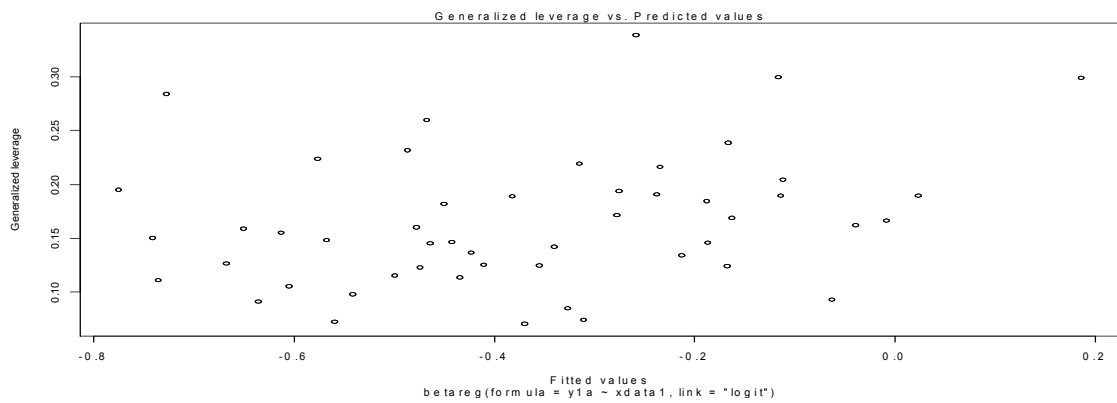
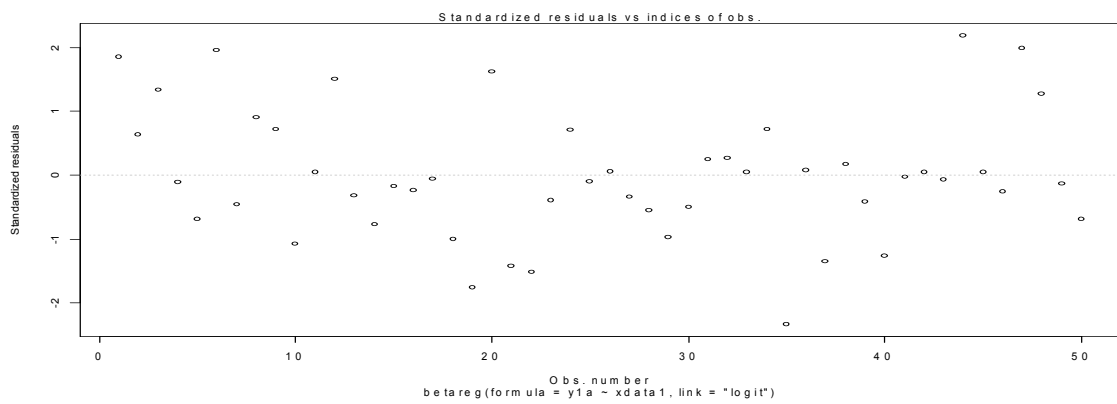
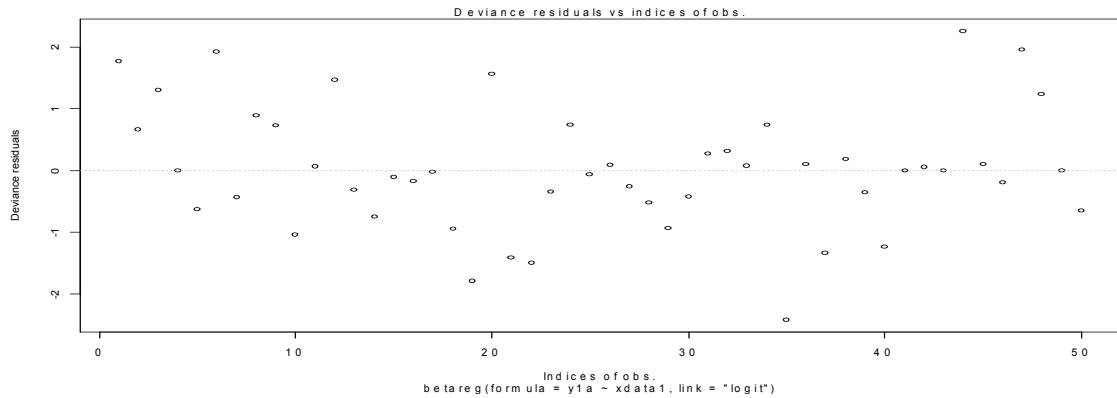
iii)Gráfico dos leverages generalizados vs os valores preditos. Por este gráfico podemos notar que aa observações 35, 44 e 47 são um pouco influentes na obtenção dos valores preditos.

iv) Gráfico da distância de Cook vs os índices. Por este gráfico podemos notar que as observações de número 35, 44 e 47 são influentes no processo de estimação, porém pelos valores do leverage generalizados, nenhuma delas são realmente influentes.

v) Gráfico dos resíduos ordinários vs os valores preditos. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar pontos atípicos como as observações de número 35, 44 e 47.

vi) Gráfico dos resíduos ordinários vs o preditor linear estimado. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar se a função de ligação é adequada ao ajuste do modelo. As únicas observações que se afastam são as observações de número 35, 44 e 47. Desse modo podemos afirmar que a função de ligação é adequada.

vii) Gráfico de probabilidades meio – normal com envelopes. Esse gráfico nos ajuda a decidir se as respostas observadas são consistentes com o modelo ajustado. Nesse caso, notamos um certo desvio, o que evidencia uma pequena inconsistência dos dados em relação ao ajuste.



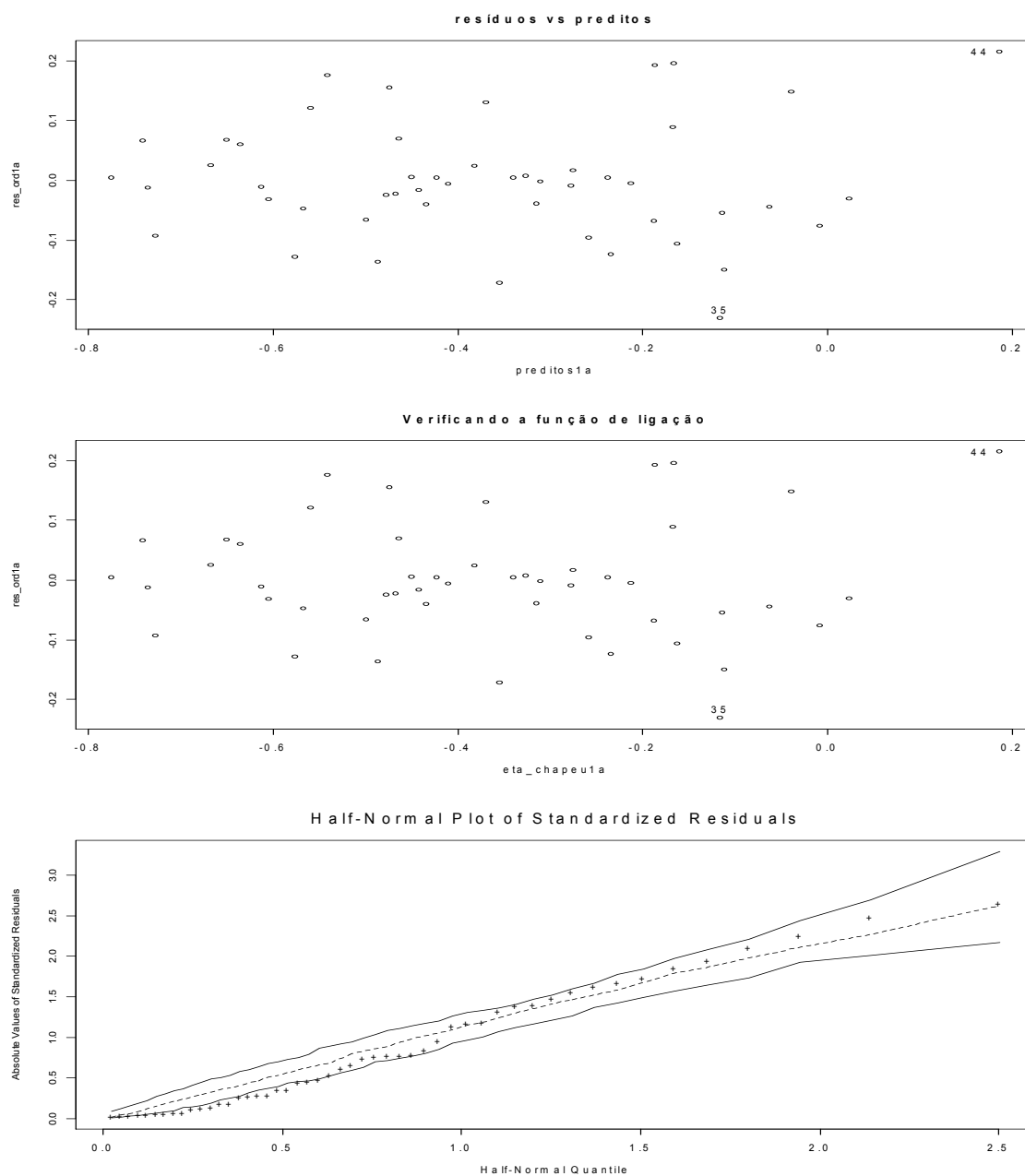


Figura 7.3: Gráficos de diagnósticos para o modelo 2 referente aos dados do Caso Charter Schools.

7.4.3 Modelo 3

Os resultados de estimação para o modelo utilizando a função de ligação logito são dados na Tabela 7.3. Verifica-se que os níveis significativos das variáveis explicativas *A variedade de Charter Schools*, *Número de debates* e *Proporções de Escolas Públicas* e *Dimensões da População* estão próximos de zero, indicando que essas variáveis são estatisticamente importantes para explicar a variável resposta *Média do Market Share*. No entanto, níveis descritivos marginais das variáveis explicativas: *Representantes Financeiros*, *Tempo* e *Medidas Financeiras* são de 0,4985; 0,8344 e 0,2173, respectivamente, o que indica, inicialmente, que essas variáveis podem ser retiradas do modelo, desde que sejam mantidas as outras quatro.

Tabela 7.3

Estimativas dos parâmetros do estudo de Charter Scholls.

| Parâmetro | Variáveis | Valores Estimados | valor-p | Intervalos de Confiança |
|-----------|-------------------------|-------------------|---------|-------------------------|
| β_0 | | -0,0515 | 0,8278 | [-0, 5299; 0, 4268] |
| β_1 | Repr.. fin. | -0,0240 | 0,4985 | [-0, 0958; 0, 0477] |
| β_2 | Var. de Charter Schools | -0,0411 | 0,0856 | [-0, 0894; 0, 0071] |
| β_3 | Nº de debates | 0,0515 | 0,0233 | [0, 0056; 0, 0973] |
| β_4 | prop. de escolas pub. | 0,3472 | 0,0238 | [0, 0372; 0, 6572] |
| β_5 | qunatidade de anos | 0,0037 | 0,8344 | [-0, 0326; 0, 0402] |
| β_6 | dim. da pop. | -0,3100 | 0,0360 | [-0, 6085; -0, 0115] |
| β_7 | medidas financeiras | -0,4229 | 0,2173 | [-1, 1147; 0, 2689] |
| ϕ | | 24,5171 | | [14, 8097; 34, 2244] |

Para o modelo 3 o pseudo R_p^2 foi de 0,2311, a Razão de verossimilhanças foi 13,07126.

Com o leverage generalizado, temos

$$GL[35, 35] = 0,2906 < 0,32 = 2k/n$$

e

$$GL[44, 44] = 0,2910 < 0,32 = 2k/n,$$

onde k é o número de parâmetros e n o tamanho da amostra. O que nos diz que as observações de número 35 e 44 não são influentes. A *deviance* residual deste modelo é igual a 48,7460.

Para uma análise gráfica dos resíduos apresentamos os gráficos da Figura 7.4. Onde, os sete gráficos que compõe a Figura 7.4 envolvem:

i) Gráfico dos resíduos da deviance vs os índices das observações. Por este gráfico podemos notar que as observações 35, 44 e 47 têm resíduos grandes.

ii) Gráfico dos resíduos padronizados vs os índices das observações. Por este gráfico podemos notar, também, que as observações 35, 44 e 47 têm resíduos grandes.

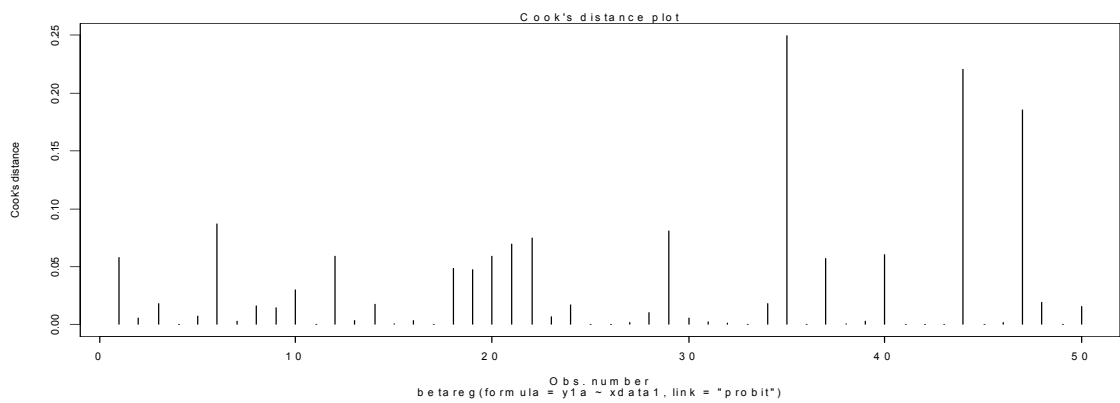
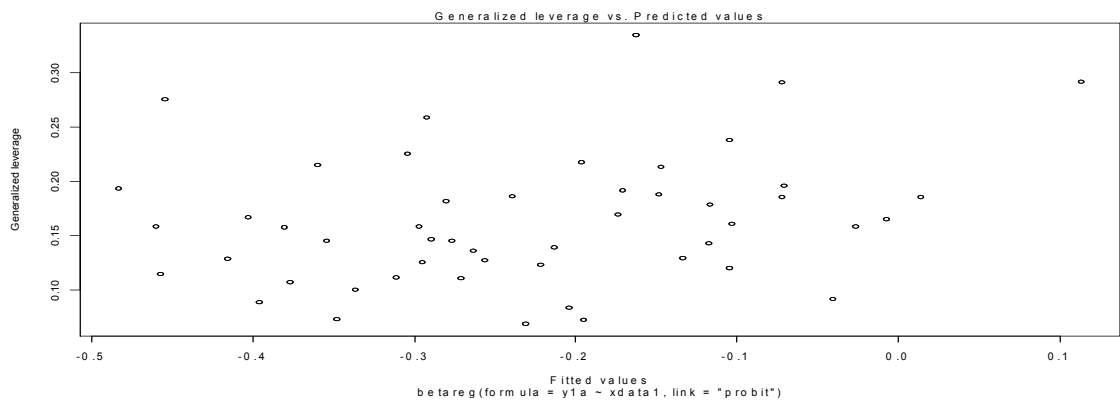
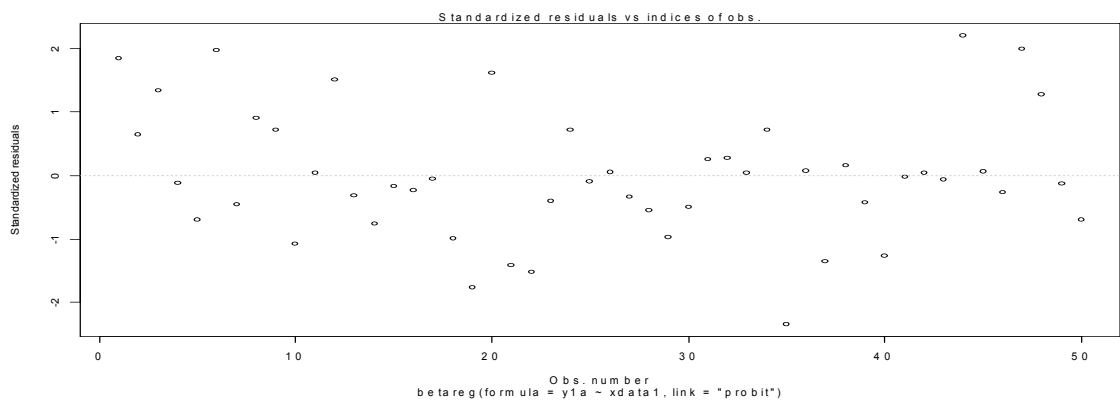
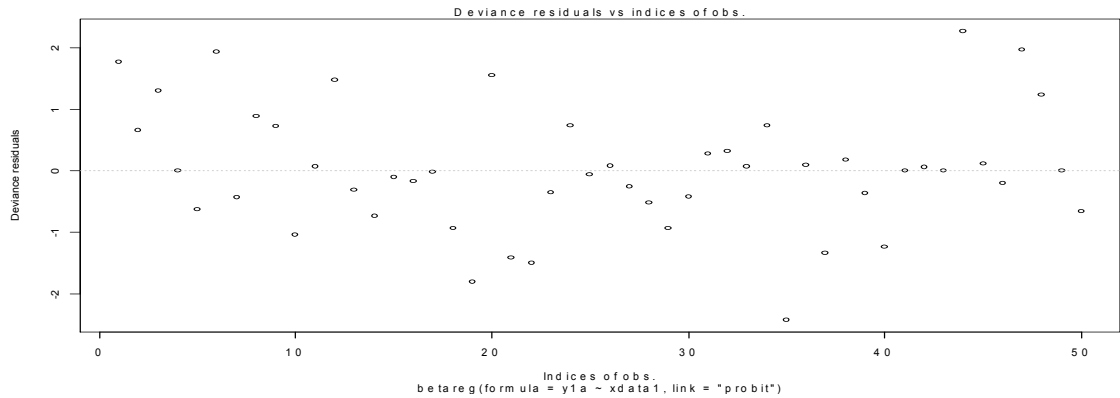
iii) Gráfico dos leverages generalizados vs os valores preditos. Por este gráfico podemos notar que nenhuma observação é muito influente na obtenção dos valores preditos.

iv) Gráfico da distância de Cook vs os índices. Por este gráfico podemos notar que as observações de números as observações 35, 44 e 47 são muito influentes no processo de estimação.

v) Gráfico dos resíduos ordinários vs os valores preditos. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar pontos atípicos como as observações de números 35, 44 e 47.

vi) Gráfico dos resíduos ordinários vs o preditor linear estimado. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar se a função de ligação é adequada ao ajuste do modelo. As únicas observações que se afastam são as observações de número 35, 44 e 47. Desse modo podemos afirmar que a função de ligação é adequada.

vii) Gráfico de probabilidades meio – normal com envelopes. Esse gráfico nos ajuda a decidir se as respostas observadas são consistentes com o modelo ajustado. Nesse caso, notamos um certo desvio, o que evidencia uma pequena inconsistência dos dados em relação ao ajuste.



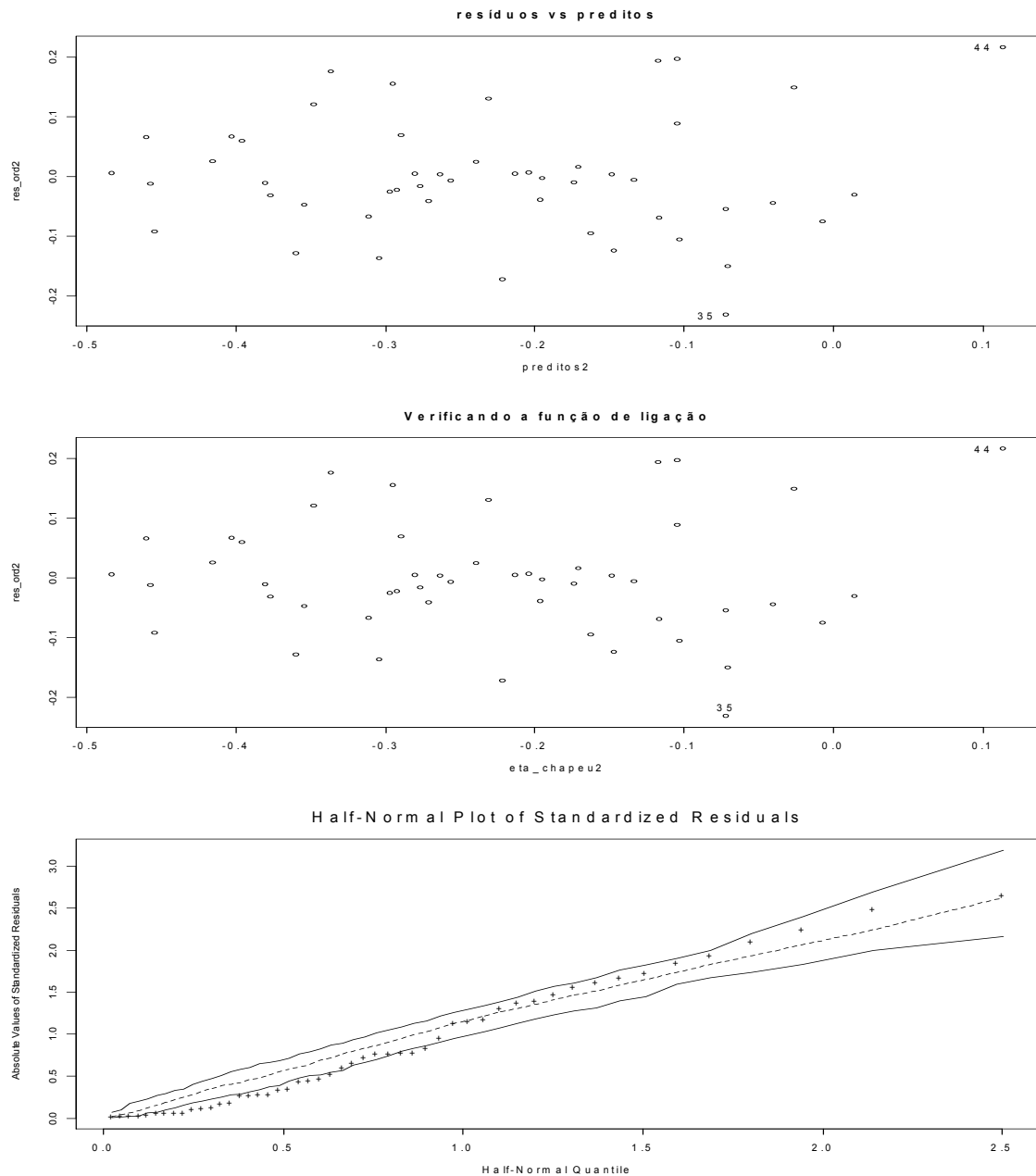


Figura 7.5: Gráficos de diagnósticos para o modelo 3 referente aos dados do Caso Charter Schools.

Testamos a hipótese $H_0 : \beta_3 = \beta_4 = 0$ vs $H_1 : \beta_3 = 0$ ou $\beta_4 = 0$ através do teste de Razão de Verossimilhanças, encontrando

$$\varpi_1 = 5,341824 > \chi_{2,095}^2 = 0,3781149$$

desse modo temos razões suficientes para rejeitar a hipótese nula.

7.4.4 Modelo 4

No próximo passo ajustamos o modelo apresentado na seção 7.4.2 com a função de ligação logito, mas sem as covariáveis: *Representantes Financeiros*, x_1 , *Tempo*, x_5 , e *Medidas Financeiras*, x_7 . As estimativas de máxima verossimilhança desse novo modelo são dados na Tabela 7.4.

Tabela 7.4

Estimativas dos parâmetros do estudo de Charter Scholls.

| Parâmetro | Variáveis | Valor Estimado | valor-p | Intervalos de Confiança |
|-----------|-------------------------|----------------|---------|-------------------------|
| β_0 | | -0,4180 | 0,0444 | [-0, 8369; 0, 0007] |
| β_2 | Var. de Charter Schools | -0,0647 | 0,0947 | [-0, 1426; 0, 0132] |
| β_3 | Nº de debates | 0,0812 | 0,0286 | [0, 0064; 0, 1560] |
| β_4 | prop. de escolas pub. | 0,4189 | 0,0599 | [-0, 0294; 0, 8674] |
| β_6 | dim. da pop. | -0,4451 | 0,0532 | [-0, 9089; 0, 0186] |
| ϕ | | 23,6080 | | [14, 2860; 32, 9300] |

Para o modelo 4 o pseudo R_p^2 foi de 0,2018, a Razão de verossimilhanças foi 11,1902.

Com o leverage generalizado, temos

$$GL[6, 6] = 0,0736 < 0,2 = 2k/n,$$

e

$$GL[35, 35] = 0,1636 < 0,2 = 2k/n,$$

e

$$GL[44, 44] = 0,2142 > 0,2 = 2k/n,$$

e

$$GL[47, 47] = 0,1325 < 0,2 = 2k/n,$$

onde k é o número de parâmetros e n o tamanho da amostra. O que nos diz que apenas a observação de número 44 é influente. A *deviance* residual deste modelo é igual a 48,6921.

Para uma análise gráfica dos resíduos apresentamos os gráficos da Figura 7.5. Onde, os sete gráficos que compõe a Figura 7.5 envolvem:

i) Gráfico dos resíduos da deviance vs os índices das observações. Por este gráfico podemos notar que a observação 35 tem resíduo grande.

ii) Gráfico dos resíduos padronizados vs os índices das observações. Por este gráfico podemos notar, também, que as observações 35, 44 e 47 têm resíduos grandes.

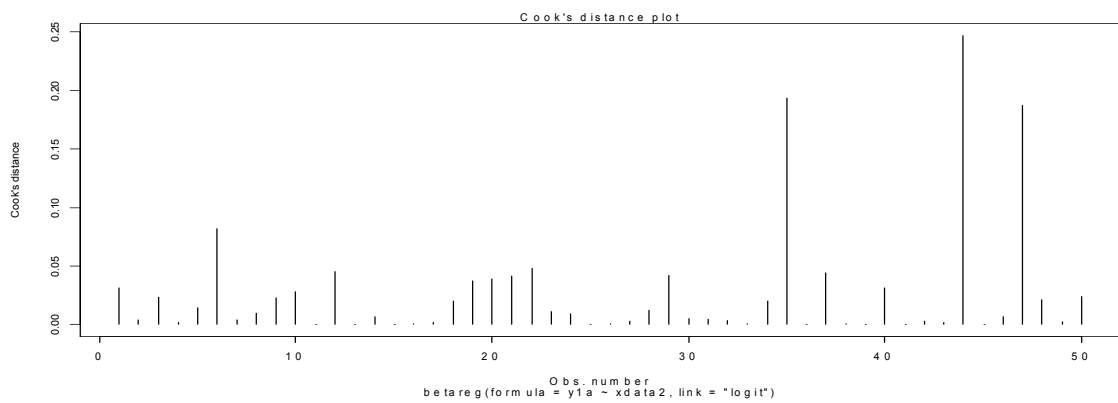
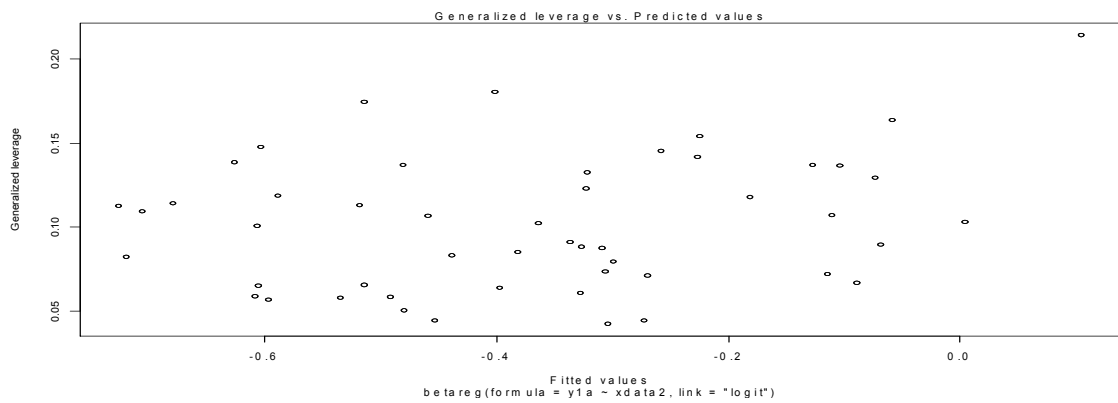
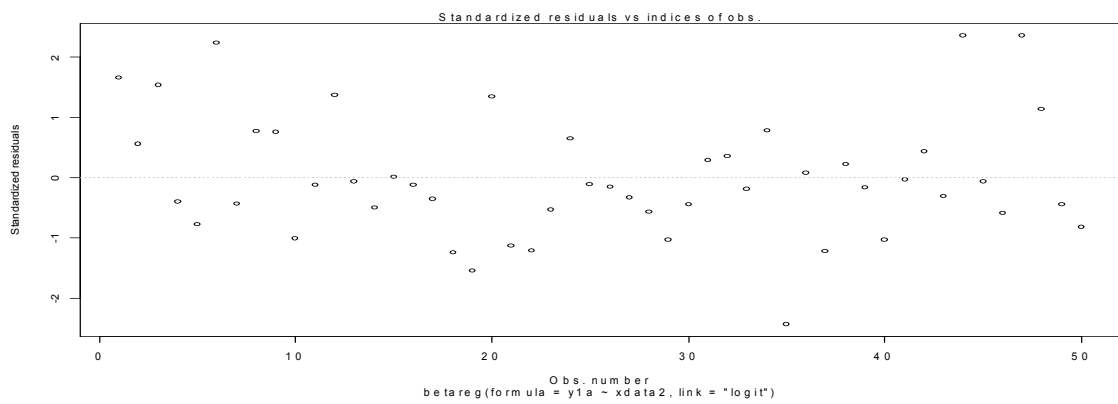
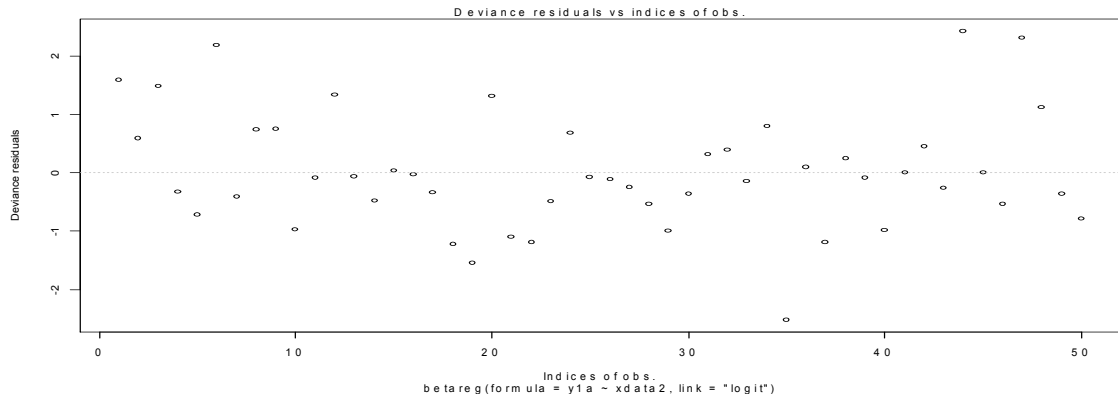
iii) Gráfico dos leverages generalizados vs os valores preditos. Por este gráfico podemos notar que não há observações muito influentes na obtenção dos valores preditos.

iv) Gráfico da distância de Cook vs os índices. Por este gráfico podemos notar que as observações de número 35, 44 e 47 são muito influentes no processo de estimação.

v) Gráfico dos resíduos ordinários vs os valores preditos. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar pontos atípicos como as observações de número 35, 44 e 47.

vi) Gráfico dos resíduos ordinários vs o preditor linear estimado. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar se a função de ligação é adequada ao ajuste do modelo. As únicas observações que se afastam são as observações de número 35, 44 e 47. Desse modo podemos afirmar que a função de ligação é adequada.

vii) Gráfico de probabilidades meio – normal com envelopes. Esse gráfico nos ajuda a decidir se as respostas observadas são consistentes com o modelo ajustado. Nesse caso, notamos um pequeno desvio, porém muito menor que o do modelo 2 o que evidencia uma pequena inconsistência dos dados em relação ao ajuste, mas que pode ser ignorada.



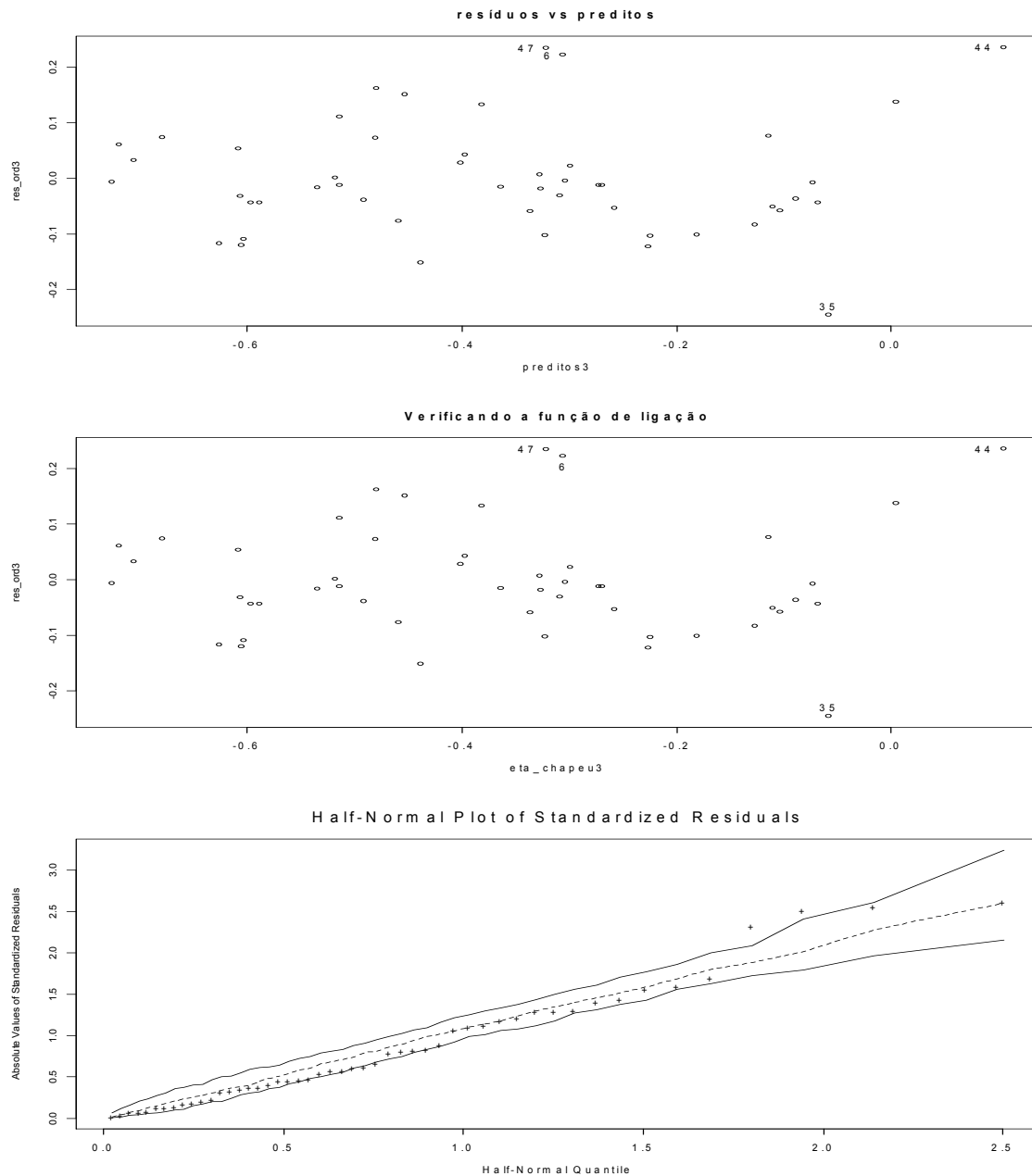


Figura 7.4: Gráficos de diagnósticos para o modelo 4 referente aos dados do Caso Charter Schools.

7.4.5 Modelo 5

Ajustamos o modelo com função de ligação probito, mas sem as covariáveis: *Representantes Financeiros*, x_1 , *Tempo*, x_5 , e *Medidas Financeiras*, x_7 . As estimativas de máxima

verossimilhança desse novo modelo são dados na Tabela 7.5.

Tabela 7.5

Estimativas dos parâmetros do estudo de Charter Scholls.

| Parâmetro | Variáveis | Valor Estimado | valor-p | Intervalos de Confiança |
|-----------|-------------------------|----------------|---------|-------------------------|
| β_0 | | -0,2601 | 0,0434 | $[-0, 5196; -0, 0007]$ |
| β_2 | Var. de Charter Schools | -0,0402 | 0,0940 | $[-0, 0887; 0, 0081]$ |
| β_3 | Nº de debates | 0,0504 | 0,0284 | $[0, 0040; 0, 0968]$ |
| β_4 | prop. de escolas pub. | 0,2602 | 0,0599 | $[-0, 0183; 0, 5388]$ |
| β_6 | dim. da pop. | -0,2767 | 0,0528 | $[-0, 5645; 0, 0110]$ |
| ϕ | | 23,6022 | | $[14, 2826; 32, 9219]$ |

Para o modelo 5 o pseudo R_p^2 foi de 0,2017, a Razão de verossimilhanças foi 11,7889. Com o leverage generalizado, temos

$$GL[35, 35] = 0,1617 < 0,2 = 2k/n,$$

e

$$GL[44, 44] = 0,2084 > 0,2 = 2k/n,$$

onde k é o número de parâmetros e n o tamanho da amostra. O que nos diz que a observação de número 44 é altamente influente. A *deviance* residual deste modelo é igual a 48,6908.

Para uma análise gráfica dos resíduos apresentamos os gráficos da Figura 7.6. Onde, os sete gráficos que compõe a Figura 7.6 envolvem:

i) Gráfico dos resíduos da deviance vs os índices das observações. Por este gráfico podemos notar que a observação 35 têm resíduo grande.

ii) Gráfico dos resíduos padronizados vs os índices das observações. Por este gráfico podemos notar, também, que a observações 35 têm resíduo grande.

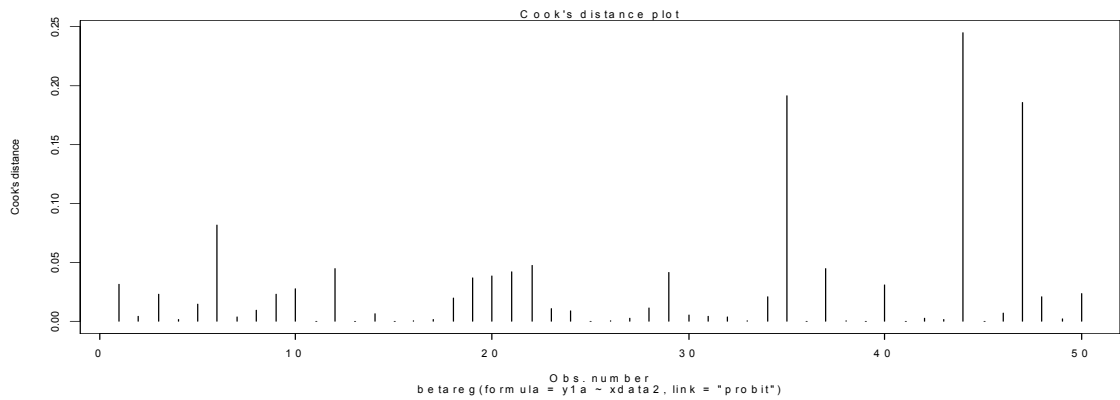
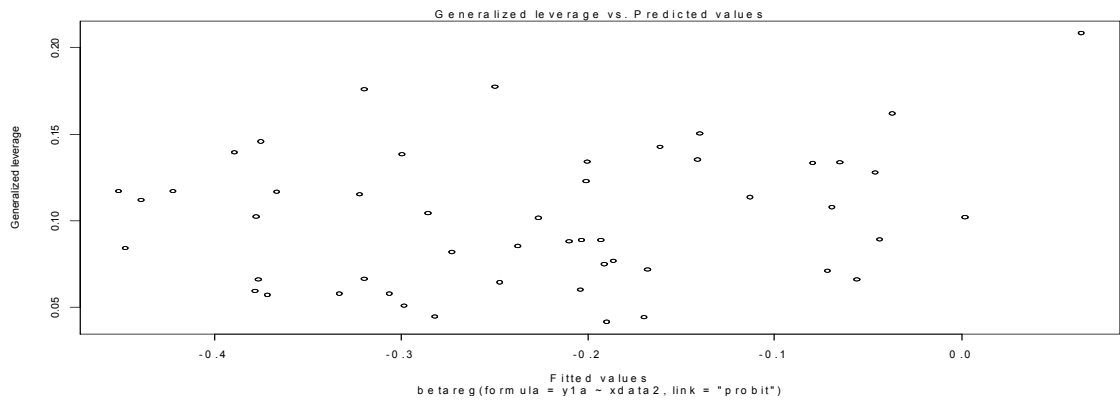
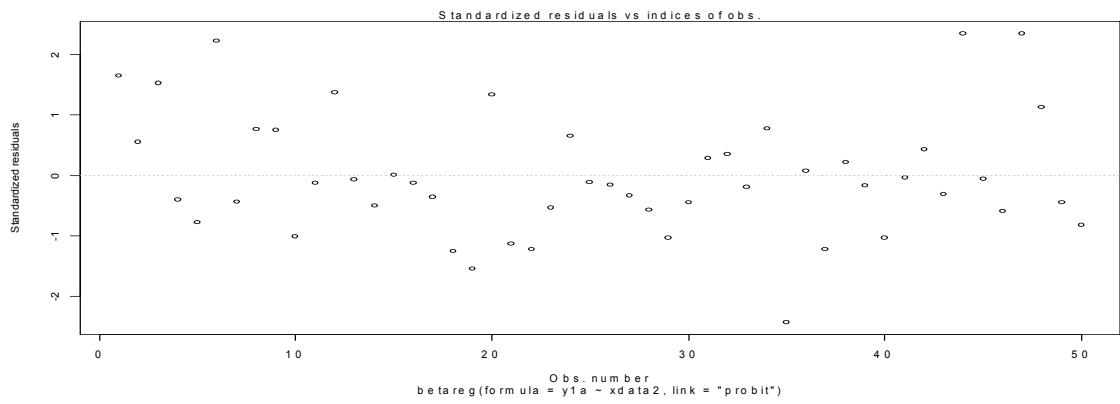
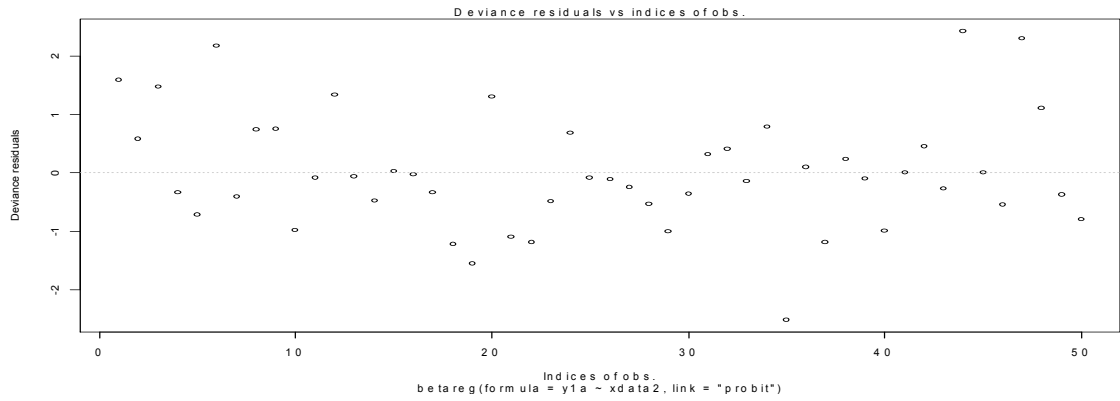
iii) Gráfico dos leverages generalizados vs os valores preditos. Por este gráfico podemos notar que a nenhuma observação é muito influente na obtenção dos valores preditos.

iv) Gráfico da distância de Cook vs os índices. Por este gráfico podemos notar que as observações de número 35, 44 e 47 são muito influentes no processo de estimação.

v) Gráfico dos resíduos ordinários vs os valores preditos. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar pontos atípicos como as observações de número 6, 35, 44 e 47.

vi) Gráfico dos resíduos ordinários vs o preditor linear estimado. O padrão desse gráfico deve ser uma nuvem ao redor do zero. Portanto, por este gráfico podemos notar se a função de ligação é adequada ao ajuste do modelo. As únicas observações que se afastam são as observações de número 44 e 47. Desse modo podemos afirmar que a função de ligação é adequada.

vii) Gráfico de probabilidades meio – normal com envelopes. Esse gráfico nos ajuda a decidir se as respostas observadas são consistentes com o modelo ajustado. Nesse caso, notamos um pequeno desvio, porém muito menor que o do modelo 2 o que evidencia uma inconsistência muito pequena dos dados em relação ao ajuste, mas que pode ser ignorada.



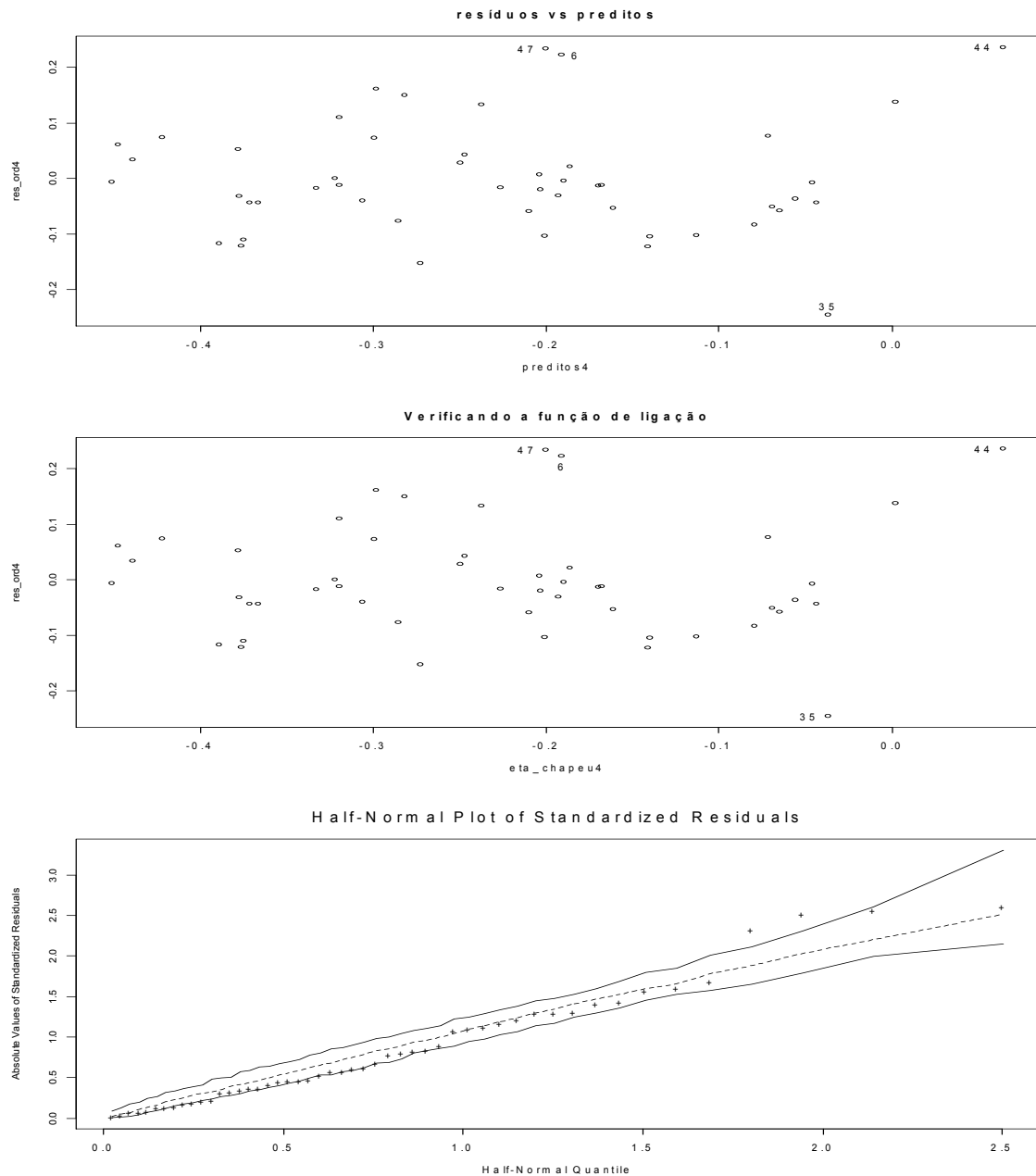


Figura 7.6: Gráficos de diagnósticos para o modelo 5 referente aos dados do Caso Charter Schools.

7.5 Comparação entre os cinco modelos

Para compararmos os cinco modelos ajustados, podemos analisar as estatísticas abaixo, apresentadas na Tabela 7.6.

Tabela 7.6

| Estatísticas para comparação entre os modelos ajustados sob o enfoque clássico. | | | | | |
|---|----------|----------|----------|----------|----------|
| Estatísticas | Modelo 1 | Modelo 2 | Modelo 3 | Modelo 4 | Modelo 5 |
| R_p^2 | 0,2129 | 0,2308 | 0,2311 | 0,2018 | 0,2017 |
| AIC | -62,0047 | -76,0755 | -76,0451 | -80,1640 | -80,1527 |
| <i>Residual Deviance</i> | 48,9021 | 48,75 | 48,7460 | 48,6921 | 48,6908 |
| SQRes | 0,5733 | 0,4735 | 0,4735 | 0,4927 | 0,4927 |
| Razão de Verossimilhanças | 13,59099 | 13,1016 | 13,0712 | 11,1902 | 11,1788 |

A Tabela 7.7 mostra alguns valores ajustados com o modelo de regressão beta via função de ligação logito (modelo 4) e probito (modelo 5), com seus respectivos resíduos. Fizemos também uma comparação através da soma dos quadrados dos resíduos. O que demonstrou que a função de ligação logito gera menor soma de quadrados dos resíduos, sendo assim ela é considerada mais apropriada para ajustar o modelo.

Tabela 7.7

| Comparação logito vs probito. | | | | | |
|-------------------------------|-------|---------------|---------------|---------------|---------------|
| Obs. | Y | Preditos mod4 | Preditos mod5 | Resíduos mod4 | Resíduos mod5 |
| 1 | 0,544 | 0,4787 | 0,2983 | 0,1614 | 0,1612 |
| 6 | 0,311 | 0,3055 | 0,1907 | 0,2222 | 0,2221 |
| 11 | 0,353 | 0,2717 | 0,1695 | 0,0126 | 0,0128 |
| 16 | 0,374 | 0,5134 | 0,3197 | 0,0121 | 0,0123 |
| 21 | 0,538 | 0,6022 | 0,3751 | 0,1097 | 0,1096 |
| 26 | 0,421 | 0,3635 | 0,2264 | 0,0159 | 0,0163 |
| 31 | 0,314 | 0,4002 | 0,2498 | 0,0282 | 0,0281 |
| 36 | 0,239 | 0,3267 | 0,2039 | 0,0070 | 0,0068 |
| 41 | 0,317 | 0,3039 | 0,1896 | 0,0040 | 0,0042 |
| 46 | 0,319 | 0,3361 | 0,2097 | 0,0591 | 0,0593 |

Comparando a soma de quadrados dos resíduos, temos $SQ Re_{s_{\log ito}} = 0,4926638$ e $SQ Re_{s_{probito}} = 0,4927584$, dessa forma temos que $SQ Re_{s_{\log ito}} < SQ Re_{s_{probito}}$. E ainda a deviance do modelo 4 é maior que a do modelo 5.

$$Deviance_{\text{mod}4} = 48,69218 > Deviance_{\text{mod}5} = 48,69089.$$

7.5.1 Comparações entre o modelo de regressão Beta ajustados com a função logito e com um modelo de regressão linear

Fizemos uma comparação entre alguns valores observados com os valores preditos de uma regressão beta e uma regressão linear normal, presentes na Tabela 7.8. Os resultados apontam uma desvantagem do uso do modelo de regressão normal linear, relativamente ao modelo de regressão beta, quando a variável é medida de forma contínua no intervalo unitário.

Tabela 7.8

Comparação entre a regressão beta e a regressão linear normal.

| Obs. | y | Valores preditos | | Obs. | y | Valores preditos | |
|------|---------|------------------|-------------|------|--------|------------------|-------------|
| | | Reg. Beta | Reg. Linear | | | Reg. Beta | Reg. Linear |
| 1 | 0,54400 | 0,4787 | 0,3831 | 31 | 0,3141 | 0,4002 | 0,3965 |
| 3 | 0,5300 | 0,4525 | 0,3894 | 36 | 0,2395 | 0,3267 | 0,4194 |
| 12 | 0,4198 | 0,0044 | 0,4997 | 37 | 0,4261 | 0,6250 | 0,3488 |
| 21 | 0,5389 | 0,6022 | 0,3524 | 39 | 0,4479 | 0,5339 | 0,3669 |
| 28 | 0,3208 | 0,1034 | 0,4688 | 40 | 0,3525 | 0,3218 | 0,4152 |
| 30 | 0,3400 | 0,5881 | 0,3546 | 47 | 0,3174 | 0,3213 | 0,4166 |

Comparando a soma de quadrados dos resíduos, temos $SQ Re_{s_{Beta-\log ito}} = 0.4926638$ e $SQ Re_{s_{linear}} = 0.492796$, dessa forma temos que $SQ Re_{s_{Beta-\log ito}} < SQ Re_{s_{linear}}$. Nesta aplicação específica, a diferença entre as somas de quadrados dos resíduos do modelo de regressão beta com função de ligação beta e o modelo regressão linear usual parece

ser insignificante, mas analisamos também outros conjuntos de dados que se tratava de proporções e vimos que essa diferença é bastante significativa.

7.6 Simulações

Utilizando os valores estimados para os parâmetros de regressão, simulamos 51 valores para a variável resposta e comparamos com os valores observados no caso *Charter Schools*.

Temos que $Y \sim \text{Beta}(p, q)$. Fazendo, $\mu = p / (p + q)$ e $\phi = p + q$.

Chegamos a

$$p = \mu\phi$$

e

$$q = \phi(1 - \mu).$$

Desse modo

$$E(Y) = \frac{p}{p + q} = \frac{\mu\phi}{\mu\phi + \phi(1 - \mu)} = \frac{\mu\phi}{\phi} = \mu$$

e

$$\text{Var}(Y) = \frac{pq}{(p + q)^2(p + q + 1)} = \frac{\mu\phi\phi(1 - \mu)}{\phi^2(\phi + 1)} = \frac{\mu(1 - \mu)}{(\phi + 1)}.$$

Assim, podemos obter ϕ , por,

$$\phi = \frac{\mu(1 - \mu)}{\text{Var}(Y)} - 1.$$

Para as simulações, procedemos da seguinte maneira:

1º: Dados os 51 valores de Y e os valores das covariáveis (fornecidos no caso *Charter Schools*), estimamos o vetor de parâmetros β através do pacote *betareg* do R e estimamos $\hat{\mu}_k$, como:

$$\hat{\mu}_k = \frac{\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{k1} + \hat{\beta}_2 x_{k2} + \hat{\beta}_3 x_{k3} + \hat{\beta}_4 x_{k4} + \hat{\beta}_5 x_{k5} + \hat{\beta}_6 x_{k6} + \hat{\beta}_7 x_{k7}\right)}{1 + \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{k1} + \hat{\beta}_2 x_{k2} + \hat{\beta}_3 x_{k3} + \hat{\beta}_4 x_{k4} + \hat{\beta}_5 x_{k5} + \hat{\beta}_6 x_{k6} + \hat{\beta}_7 x_{k7}\right)}.$$

2º: Estimamos $\hat{\phi}$;

3º: Calculamos

$$\hat{p} = \hat{\mu}\hat{\phi}.$$

4º: Calculamos

$$\hat{q} = \hat{\phi} (1 - \hat{\mu}).$$

5º: Com \hat{p} e \hat{q} , fazemos uma simulação de Y, tamanhos 51;

i) Gerar 51 valores de $\hat{y} \sim \text{Beta}(\hat{p}, \hat{q})$;

6º: Comparar graficamente o y observado com o \hat{y} gerado (figura 7.7).

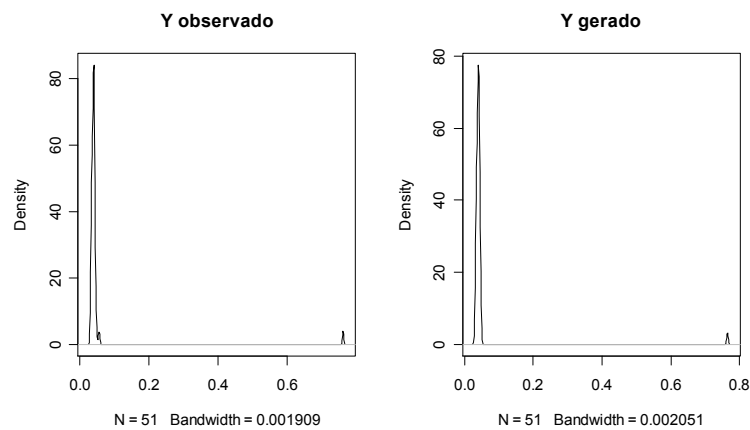


Figura 7.7. Gráficos para comparação entre os valores observados e os valores gerados.

Notamos pelo gráfico acima que os valores simulados se aproximam muito dos valores observados, ou seja, os valores estimados para os parâmetros estão bem próximos dos valores reais.

7.7 Inferência Bayesiana

Aqui adotamos a abordagem bayesiana para analisar o conjunto de dados em questão.

Suponha que nossa variável resposta tem distribuição beta:

$$\text{Market Share}_k \sim \text{beta}(p_k, q_k),$$

onde:

$$p_k = \frac{(E(\text{Market Share}_k))^2 (1 - E(\text{Market Share}_k))}{\text{Var}(\text{Market Share}_k)} - E(\text{Market Share}_k)$$

e

$$q_k = \frac{(E(\text{Market Share}_k))(1 - E(\text{Market Share}_k))}{\text{Var}(\text{Market Share}_k)} - (1 - E(\text{Market Share}_k)).$$

Fazendo uma reparametrização, ficamos com

$$p_k = \frac{(a_k)^2 (1 - a_k)}{b_k} - a_k$$

e

$$q_k = \frac{(a_k)(1 - a_k)}{b_k} - (1 - a_k),$$

onde, $a_k = E(\text{Market Share}_k)$ e $b_k = \text{Var}(\text{Market Share}_k)$.

A especificação, a seguir, nos permite expressar a variância e o valor esperado ou média da variável dependente usando funções que permitem ao modelo linear, combinações lineares das covariadas. No nosso caso a média é dada por:

$$E(\text{Market Share}_k) = \frac{\exp(\mathbf{x}_k^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_k^T \boldsymbol{\beta})},$$

ou ainda,

$$\mu_k = \frac{\exp(\mathbf{x}_k^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_k^T \boldsymbol{\beta})}.$$

É claro que,

$$\mathbf{x}_k^T \boldsymbol{\beta} = \log\left(\frac{\mu_k}{1 + \mu_k}\right),$$

onde o vetor das covariáveis \mathbf{x}_k^T , inclui todas as variáveis acima (mais a constante) e o vetor dos coeficientes $\boldsymbol{\beta}$, é o alvo da estimação. Nenhuma interação ou combinação não linear está incluso.

A variância da participação de mercado é dada por:

$$Var(Y_k) = \frac{\mu_k(1 - \mu_k)}{\phi_Y + 1}.$$

Onde ϕ_k , a dispersão, é definida em termos de uma diferente combinação linear das covariáveis:

$$\phi_k = \exp(\mathbf{Z}_k^T \boldsymbol{\gamma}).$$

Aqui incluímos somente uma constante e as covariáveis: dimensões da população, x_6 e, as medidas das financeiras x_7 no vetor \mathbf{Z}_k^T e estamos interessados em estimar o vetor $\boldsymbol{\gamma}$. Na verdade, pretendemos refazer todos os cálculos realizados por Buckley (2003), mas agora com outras prioris e formalizar a análise de um modelo de regressão beta sob o enfoque bayesiano.

Discutiremos qual tipo de modelo é mais adequado ao conjunto de dados citado. Esta discussão foi baseada nos critérios de seleção de modelos **DIC** (Deviance Information Criterion) e **BIC** (Bayesian Information Criterion).

Para a análise sob o enfoque bayesiano é conveniente aqui uma melhor nomenclatura dos modelos, que apresentamos abaixo.

Modelo 6: Modelo de regressão beta com prioris vagas normais para os parâmetros de regressão $\beta's$ e sendo o parâmetro de dispersão conhecido e constante. ;

Modelo 7: Modelo de regressão beta com prioris vagas normais para os parâmetros de regressão $\beta's$ e $\gamma's$, supondo o parâmetro de dispersão desconhecido e não-constante;

Modelo 8: Modelo de regressão beta com prioris informativas normais para os parâmetros de regressão $\beta's$ e $\gamma's$, supondo o parâmetro de dispersão desconhecido e não-constante;

7.8 Estimação dos Parâmetros via Métodos Bayesianos

Para os cálculos das estimativas Bayesianas foi utilizado o WinBugs v. 1.4. As *posteriors* são simuladas usando o algoritmo *Metropolis-Hastings*.

Para isso, é gerada uma cadeia de 52000 valores, com os primeiros 2000 valores descartados para eliminar o efeito dos valores iniciais (*burn-in*). Com os 50000 valores restantes, espaçados de 5 em 5, uma amostra de 10000 valores é utilizada para caracterizar a distribuição *a posteriori* do parâmetro de interesse.

7.8.1 Modelo 6

Estamos considerando um modelo de regressão beta com prioris vagas normais para os parâmetros de regressão $\beta_0, \beta_1, \dots, \beta_7$ e supondo o hiperparâmetro de dispersão conhecido e constante. Assim, assumimos que os parâmetros $\beta_0, \beta_1, \dots, \beta_7$ têm distribuições *a priori* normais com média 0 e variância 100 e consideramos esses parâmetros independentes. Ou seja,

$$\begin{aligned}\beta_0 &\sim N(0; 100), \beta_1 \sim N(0; 100), \beta_2 \sim N(0; 100), \beta_3 \sim N(0; 100), \\ \beta_4 &\sim N(0; 100), \beta_5 \sim N(0; 100), \beta_6 \sim N(0; 100), \beta_7 \sim N(0; 100),\end{aligned}$$

tem-se que as f.d.p.'s desses parâmetros são

$$\begin{aligned}p(\beta_0) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\beta_0^2}{20000}\right), \\ p(\beta_1) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\beta_1^2}{20000}\right), \\ &\vdots \\ p(\beta_7) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\beta_7^2}{20000}\right).\end{aligned}$$

Usando a regra de Bayes para combinar estas distribuições *a priori* com a função de verossimilhança, dada em (6.1), a seguinte distribuição *a posteriori* conjunta para os β 's é obtida

$$\begin{aligned}
& p(\beta_0, \beta_1, \dots, \beta_7 | \phi, y, \mathbf{X}, \mathbf{Z}) \\
& \propto L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \phi) p(\beta_0) \dots p(\beta_7) \propto [\Gamma(\phi)]^n \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{(\phi - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_0^2 - \beta_1^2 - \dots - \beta_7^2}{20000} \right).
\end{aligned}$$

Como essa distribuição a *posteriori* é intratável analiticamente, será proposto o procedimento de *Gibbs Sampling* para retirar amostras da distribuição a *posteriori*. Para isso, é importante obter as distribuições a *posteriori* condicionais de cada parâmetro.

i) Distribuição a *posteriori* condicional de β_0 dado $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$.

Dado $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$, a distribuição a *posteriori* condicional de β_0 é

$$\begin{aligned}
& p(\beta_0 | \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \phi, y, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\phi \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{(\phi - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_0^2}{20000} \right).
\end{aligned}$$

ii) Distribuição a *posteriori* condicional de β_1 dado $\beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$.

Dado $\beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$, a distribuição a *posteriori* condicional de β_1 é

$$\begin{aligned} & p(\beta_1 | \beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \phi, y, \mathbf{X}, \mathbf{Z}) \\ \propto & \prod_{k=1}^n \left[\Gamma \left(\frac{\phi \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\ & \times \prod_{k=1}^n \left[\Gamma \left(\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\ & \times \prod_{k=1}^n y_k \left[\frac{(\phi - 1) \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7}) - 1}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\ & \times \prod_{k=1}^n (1 - y_k) \left[\frac{\phi}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\ & \times \exp \left(\frac{-\beta_1^2}{20000} \right). \end{aligned}$$

de forma análoga para $\beta_2, \beta_3, \dots, \beta_7$.

Gerar uma amostra de $\beta_0 | \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \phi, y, X$ não é fácil, pois como a distribuição de $\beta_0 | \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \phi, y, X$ não pertence a uma família de distribuição conhecida, assim como as demais condicionais, com isso, é utilizado um procedimento Monte Carlo em Cadeias de Markov-MCMC (Carlin e Gelfand, 1991), mais precisamente *Metropolis Hastings*, para gerar uma amostra aproximada desta distribuição a *posteriori*.

As estimativas são dadas pela média das cadeias geradas a partir da distribuição a *posteriori* de cada parâmetro do modelo, obtidas via Método de Monte Carlo em Cadeias de Markov-MCMC (Carlin e Gelfand, 1991).

Como sabemos o processo de estimação para o modelo de regressão beta no Software WinBugs é extremamente suscetível aos valores iniciais e, os valores verdadeiros dos parâmetros estão entre zero e um, então propomos valores iniciais para a geração das cadeias nesse mesmo intervalo. Portanto, serão geradas duas cadeias com os seguintes valores iniciais:

Cadeia 1: $(\beta_0 = 0, \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0, \beta_5 = 0, \beta_6 = 0, \beta_7 = 0)$,

Cadeia 2: $(\beta_0 = 0,5; \beta_1 = 0,5; \beta_2 = 0,5; \beta_3 = 0,5; \beta_4 = 0,5; \beta_5 = 0,5; \beta_6 = 0,5; \beta_7 = 0,5)$.

Abaixo, na Tabela 7.9., temos o sumário a *posteriori* para os parâmetros do Modelo 6 e as estatísticas DIC e BIC.

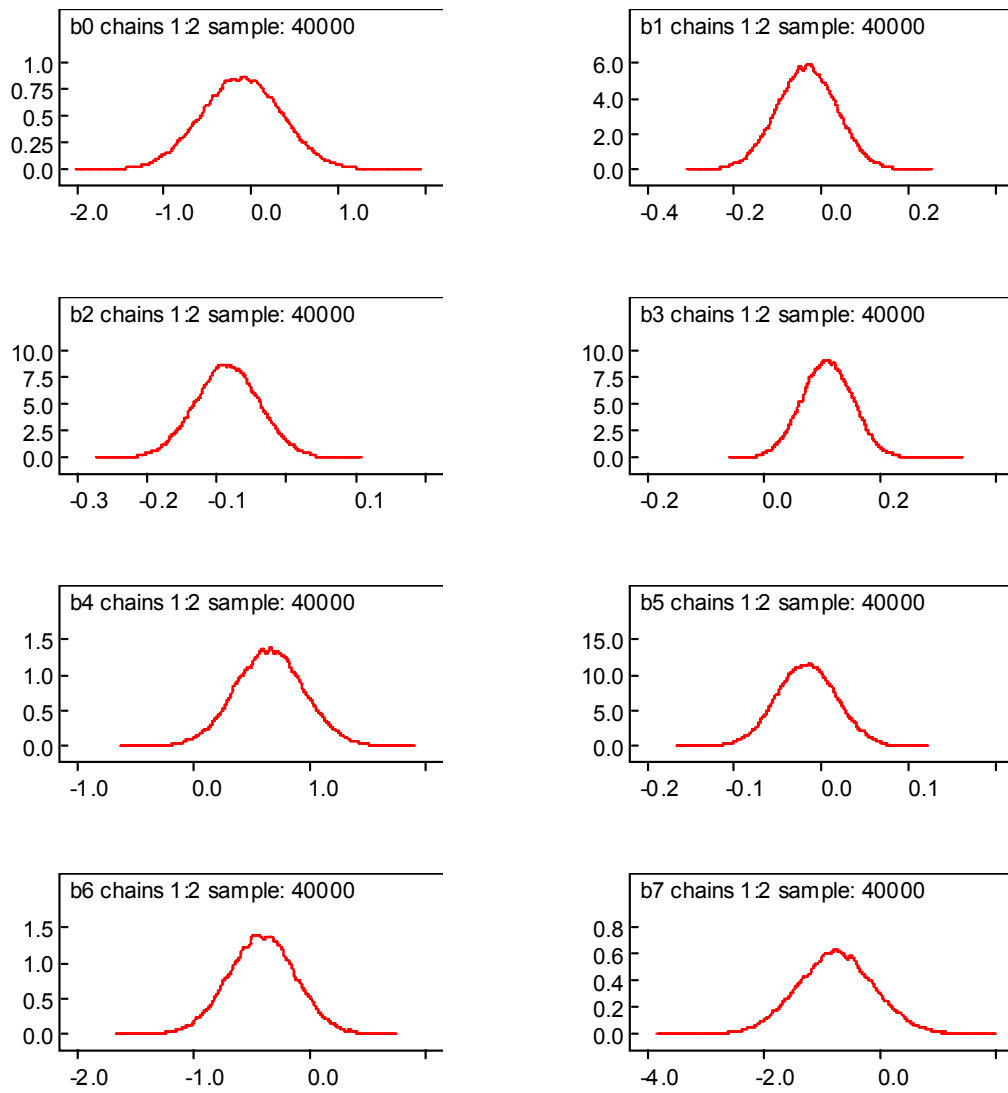
Tabela 7.9

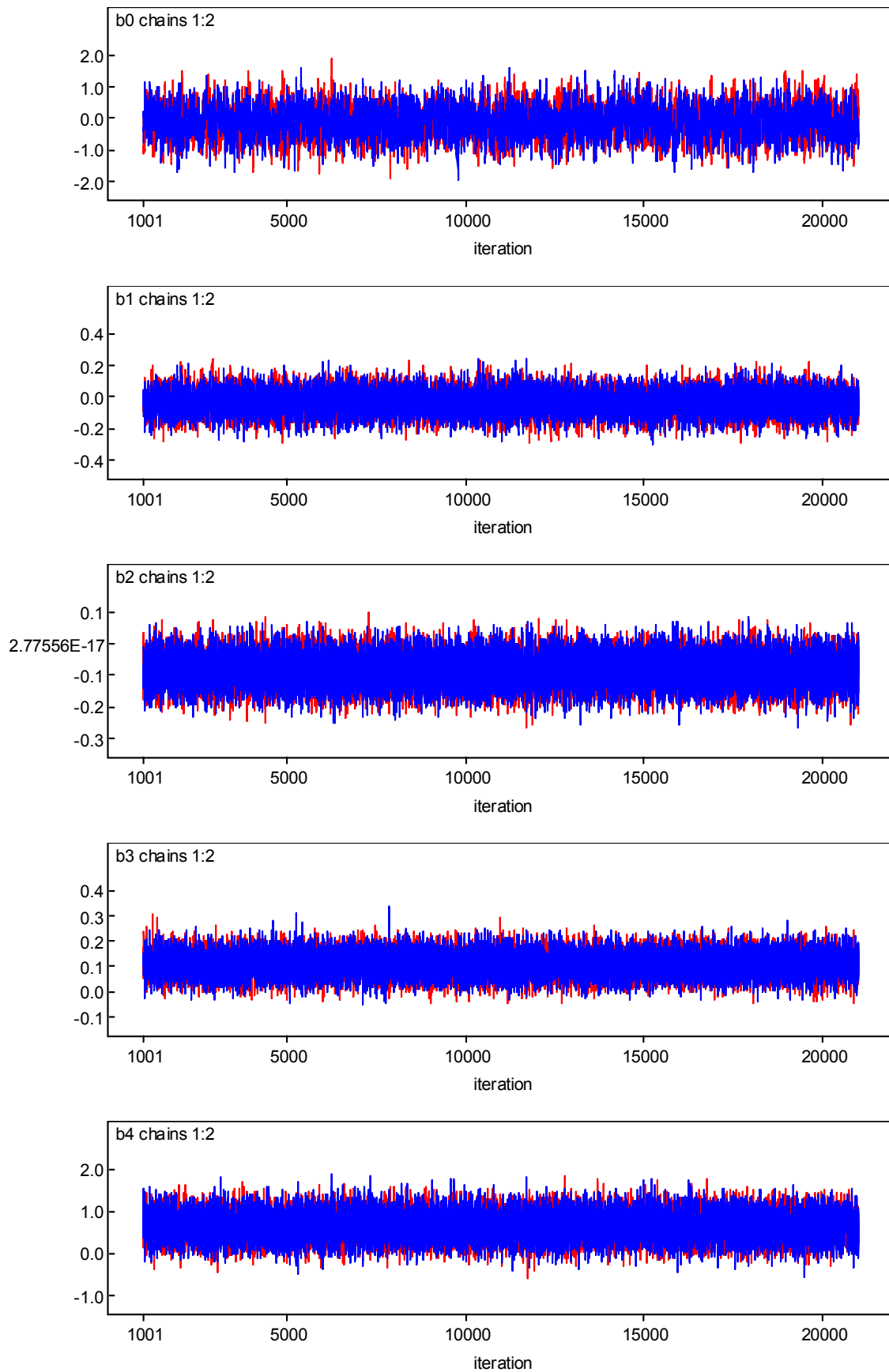
Sumário a *posteriori* para os parâmetros do **Modelo 6**.

| Parâmetros | Média | D.P. | 2,5% | mediana | 97,5% |
|------------|---------|--------|---------|---------|--------|
| β_0 | -0,1141 | 0,4596 | -1,015 | -0,1126 | 0,7782 |
| β_1 | -0,0344 | 0,0691 | -0,1694 | -0,0341 | 0,0997 |
| β_2 | -0,0844 | 0,0459 | -0,1745 | -0,0845 | 0,0060 |
| β_3 | 0,1094 | 0,0436 | 0,0239 | 0,1093 | 0,1943 |
| β_4 | 0,6578 | 0,2961 | 0,0765 | 0,6573 | 1,243 |
| β_5 | -0,0176 | 0,0344 | -0,0855 | -0,0176 | 0,0500 |
| β_6 | -0,4085 | 0,2861 | -0,9725 | -0,4079 | 0,1517 |
| β_7 | -0,7570 | 0,6604 | -2,051 | -0,7569 | 0,5434 |

DIC = -61,883 e **BIC** = -30,21453.

Gráficos das distribuições a *posteriori* marginais para os parâmetros do modelo 6 são apresentados na Figura 7.8. Na Figura 7.9 temos uma indicação de convergência para todos os parâmetros através do gráfico de convergência das cadeias. Através da Figura 7.10 as estatísticas de Gelman-Rubin nos mostram que à medida que aumentamos o número de iterações, temos um indicativo de convergência para todos os parâmetros do modelo 6, pois o índice R se aproxima de 1.

Figura 7.8: Densidade *a posteriori* marginais



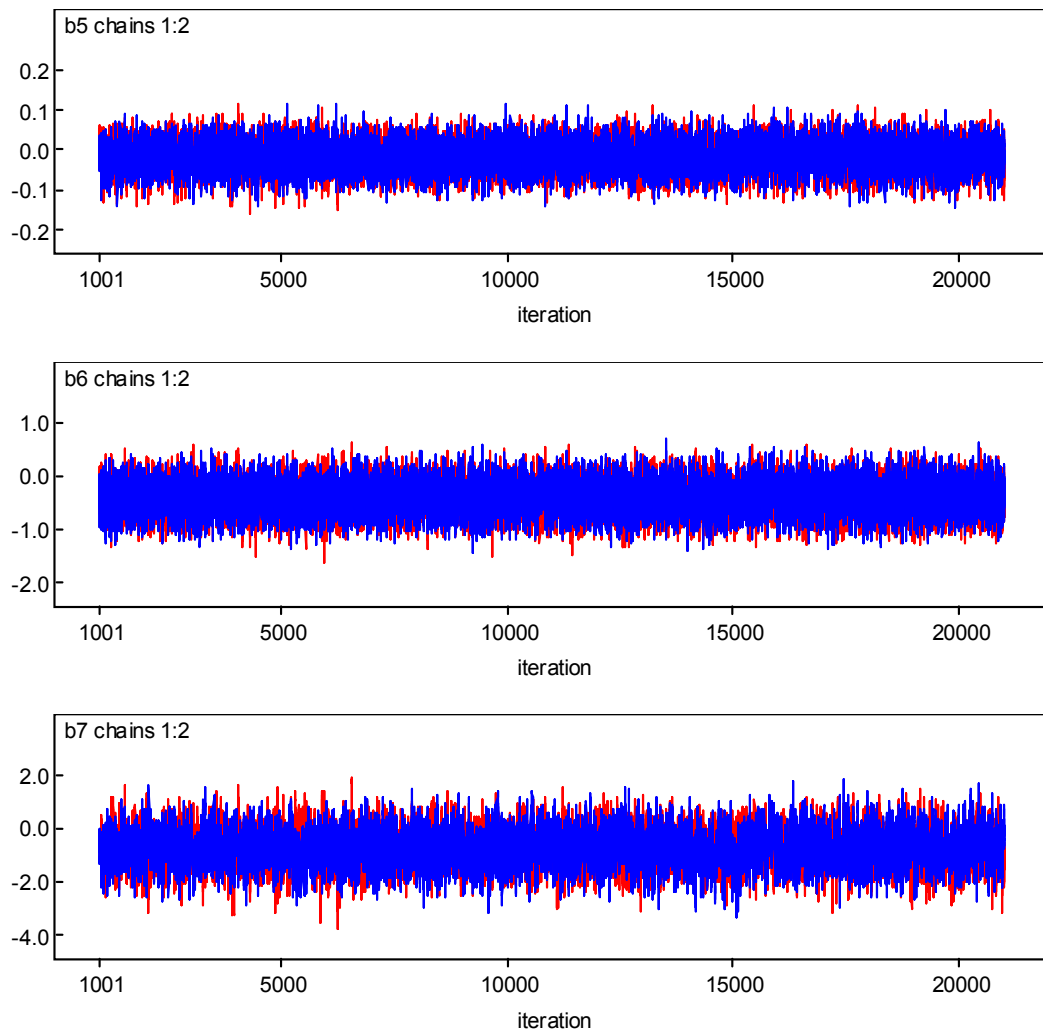


Figura 7.9: Gráficos de convergência das cadeias do modelo 6.

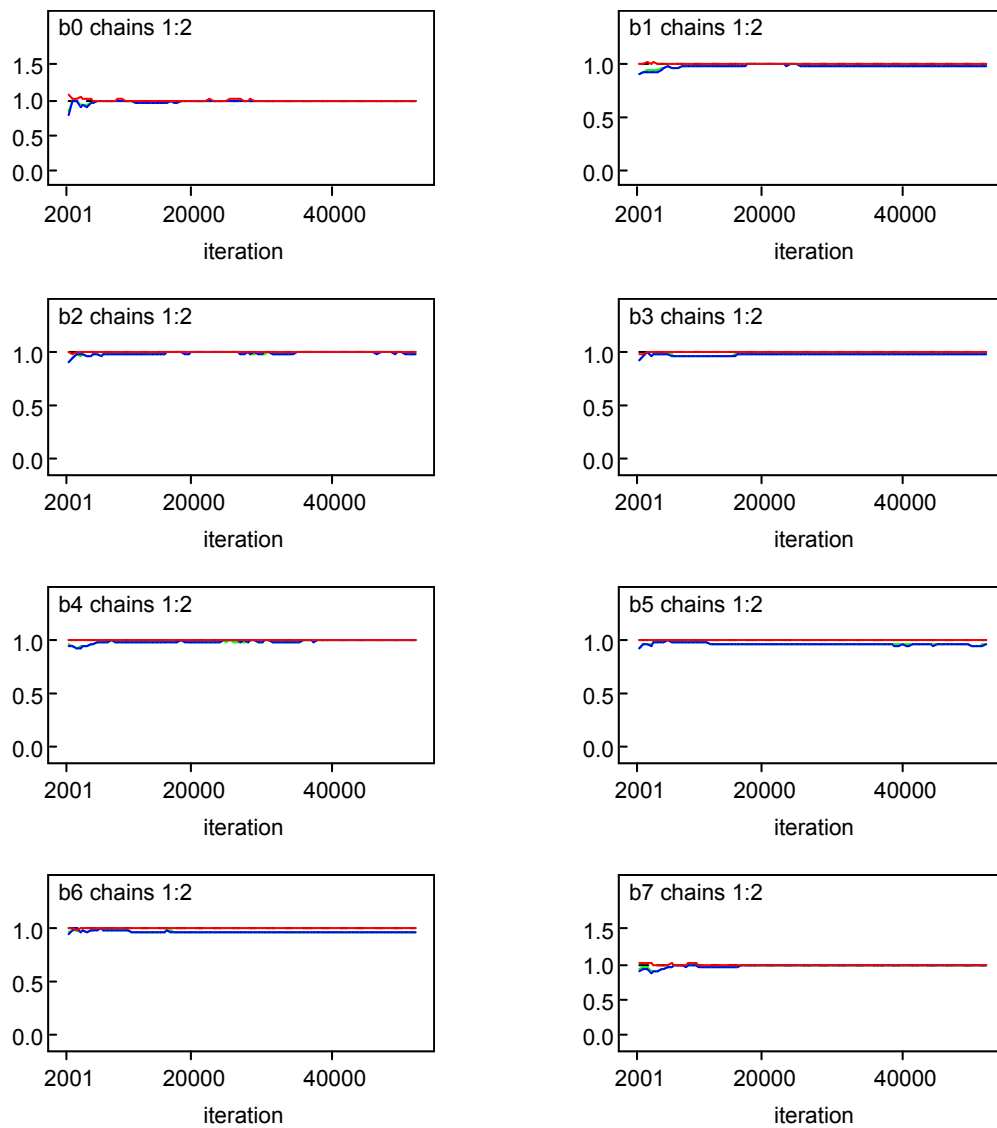


Figura 7.10: Gráficos dos índices R de Gelman e Rubin.

7.8.2 Modelo 7

Consideramos, aqui, um modelo de regressão beta, onde o hiperparâmetro de dispersão não é constante. Para estimar os parâmetros, pelo método bayesiano, especificamos prioris vagas para os parâmetros. Assumimos densidades normais com média 0 e variância 100 para os parâmetros β' s e para os parâmetros γ' s, considerando esses parâmetros independentes, ou seja, serão consideradas as seguintes distribuições a priori:

$$\begin{aligned}\beta_0 &\sim N(0, 100), \beta_1 \sim N(0, 100), \beta_2 \sim N(0, 100), \beta_3 \sim N(0, 100), \\ \beta_4 &\sim N(0, 100), \beta_5 \sim N(0, 100), \beta_6 \sim N(0, 100), \beta_7 \sim N(0, 100), \\ \gamma_0 &\sim N(0, 100), \gamma_1 \sim N(0, 100), \gamma_2 \sim N(0, 100).\end{aligned}$$

tem-se que as f.d.p.'s desses parâmetros são

$$\begin{aligned}p(\beta_0) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\beta_0^2}{20000}\right), \\ p(\beta_1) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\beta_1^2}{20000}\right), \\ &\vdots \\ p(\beta_7) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\beta_7^2}{20000}\right), \\ p(\gamma_0) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\gamma_0^2}{20000}\right), \\ p(\gamma_1) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\gamma_1^2}{20000}\right), \\ p(\gamma_2) &= \frac{1}{100\sqrt{2\pi}} \exp\left(\frac{-\gamma_2^2}{20000}\right).\end{aligned}$$

Usando a regra de Bayes para combinar estas distribuições a *priori* com a função de verossimilhança dada em (6.2), a seguinte distribuição a *posteriori* conjunta para os β' s

e os γ 's é obtida

$$\begin{aligned}
& p(\beta_0, \beta_1, \dots, \beta_7, \gamma_0, \dots, \gamma_2 | \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto L(\mathbf{y}, \mathbf{x} | \boldsymbol{\beta}, \boldsymbol{\gamma}) p(\beta_0) \dots p(\beta_7) p(\gamma_0) \dots p(\gamma_2) \propto \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_2 z_{k2}))] \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_0^2 - \beta_1^2 - \dots - \beta_7^2 - \gamma_0^2 - \gamma_1^2 - \gamma_2^2}{20000} \right).
\end{aligned}$$

Como essa distribuição a *posteriori* é intratável analiticamente, novamente será proposto o procedimento de *Gibbs Sampling* para retirar amostras da distribuição a *posteriori*. Para isso, é importante obter as distribuições a *posteriori* condicionais de cada parâmetro.

i) Distribuição a *posteriori* condicional de β_0 dado $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_0 é

$$\begin{aligned}
& p(\beta_0 | \beta_1, \dots, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_0^2}{20000} \right).
\end{aligned}$$

ii) Distribuição a *posteriori* condicional de β_1 dado $\beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_1 é

$$\begin{aligned}
& p(\beta_1 | \beta_0, \beta_2, \dots, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_1^2}{20000} \right)
\end{aligned}$$

As outras condicionais para os demais β' s são análogas a essas. Já as condicionais para os parâmetros γ' s são:

i) Distribuição a *posteriori* condicional de γ_0 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de γ_0 é

$$\begin{aligned}
& p(\gamma_0 | \beta_0, \beta_1, \dots, \beta_7, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_2 z_{k2}))]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\gamma_0^2}{20000} \right)
\end{aligned}$$

ii) Distribuição a *posteriori* condicional de γ_1 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_2$, a distribuição a *posteriori* condicional de γ_1 é

$$\begin{aligned}
& p(\gamma_1 | \beta_0, \beta_1, \dots, \beta_7, \gamma_0, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_2 z_{k2}))]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\gamma_1^2}{20000} \right)
\end{aligned}$$

iii) Distribuição a *posteriori* condicional de γ_2 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1$, a distribuição a *posteriori* condicional de γ_2 é

$$\begin{aligned}
& p(\gamma_2 | \beta_0, \beta_1, \dots, \beta_7, \gamma_0, \gamma_1, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_2 z_{k2}))]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\gamma_2^2}{20000} \right)
\end{aligned}$$

Gerar uma amostra de $\beta_0 | \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, y, X$ não é fácil, pois como a distribuição de $\beta_0 | \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, y, X$ não pertence a uma família

de distribuição conhecida, assim como as demais condicionais, com isso, é utilizado um procedimento Monte Carlo em Cadeias de Markov-MCMC (Carlin e Gelfand, 1991), mais precisamente *Metropolis Hastings*, para gerar uma amostra aproximada desta distribuição *a posteriori*.

As estimativas são dadas pela média das cadeias geradas a partir da distribuição *a posteriori* de cada parâmetro do modelo, obtidas via Método de Monte Carlo em Cadeias de Markov-MCMC (Carlin e Gelfand, 1991).

Como sabemos o processo de estimação para o modelo de regressão beta no Software WinBugs é extremamente suscetível aos valores iniciais e, os valores verdadeiros dos parâmetros estão entre zero e um, então propomos valores iniciais para a geração das cadeias nesse mesmo intervalo. Portanto, serão geradas duas cadeias com os seguintes valores iniciais:

Cadeia 1: ($\beta_0 = 0, \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0, \beta_5 = 0, \beta_6 = 0, \beta_7 = 0, \gamma_0 = 0, \gamma_1 = 0, \gamma_2 = 0$),

Cadeia 2: ($\beta_0 = 0,5; \beta_1 = 0,5; \beta_2 = 0,5; \beta_3 = 0,5; \beta_4 = 0,5; \beta_5 = 0,5; \beta_6 = 0,5; \beta_7 = 0,5; \gamma_0 = 0,5; \gamma_1 = 0,5; \gamma_2 = 0,5$).

Abaixo, na Tabela 7.10., temos o sumário *a posteriori* para os parâmetros do Modelo 7 e as estatísticas DIC e BIC.

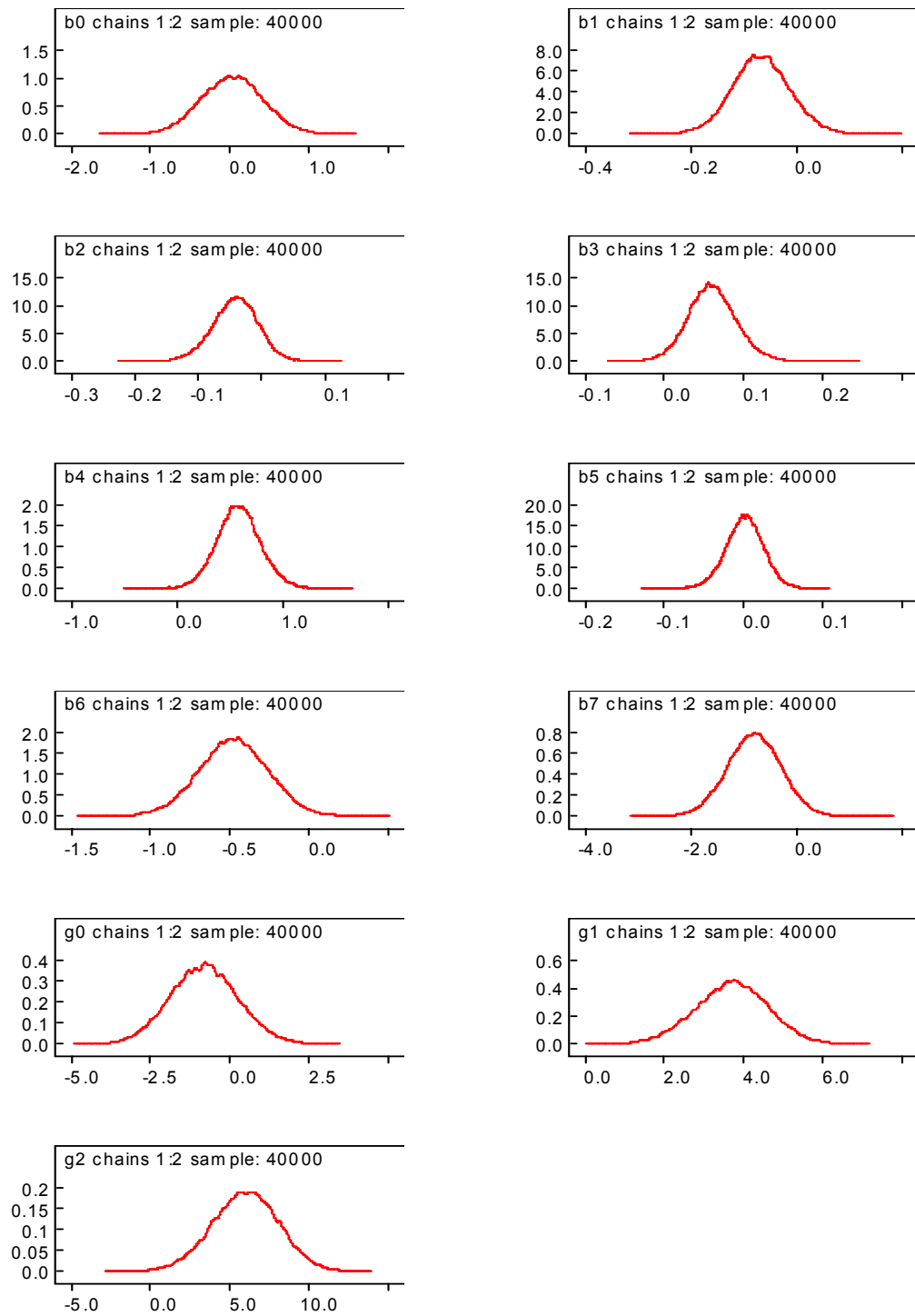
Tabela 7.10

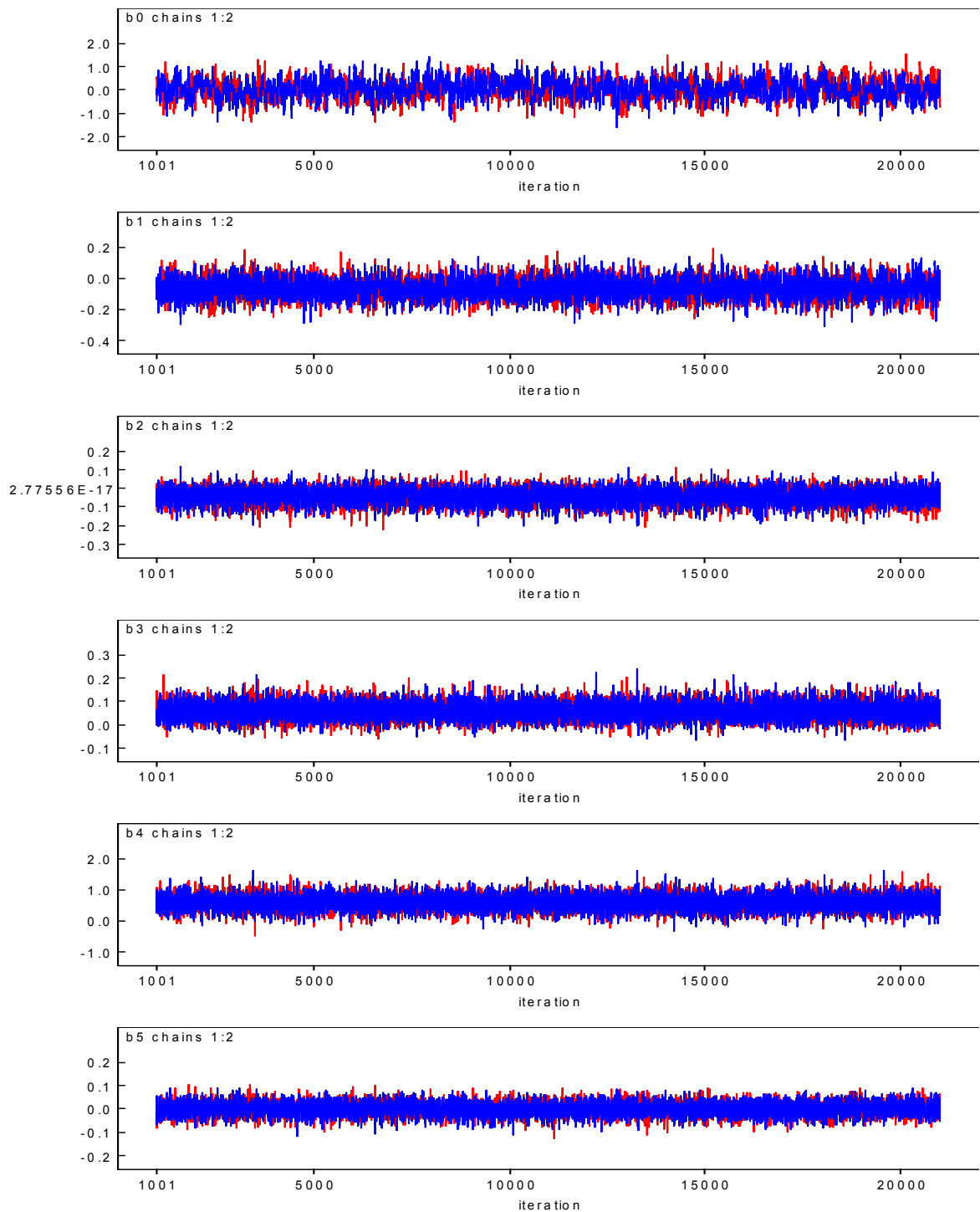
Sumário a *posteriori* para os parâmetros do **Modelo 7.**

| Parâmetros | Média | D.P. | 2,5% | mediana | 97,5% |
|------------|-----------|--------|---------|---------|---------|
| β_0 | 0,0257 | 0,3828 | -0,7168 | 0,0274 | 0,7719 |
| β_1 | -0,0668 | 0,0546 | -0,1728 | -0,0677 | 0,0419 |
| β_2 | -0,0421 | 0,0359 | -0,1166 | -0,041 | 0,0255 |
| β_3 | 0,0602 | 0,0306 | 0,0018 | 0,0593 | 0,1236 |
| β_4 | 0,582 | 0,2099 | 0,1743 | 0,5795 | 1,006 |
| β_5 | 8,223e-04 | 0,0239 | -0,0479 | 0,0011 | 0,0470 |
| β_6 | -0,4777 | 0,2198 | -0,9141 | -0,4777 | -0,0506 |
| β_7 | -0,8048 | 0,5082 | -1,802 | -0,8022 | 0,1822 |
| γ_0 | -0,8162 | 1,066 | -2,884 | -0,8168 | 1,295 |
| γ_1 | 3,692 | 0,9095 | 1,881 | 3,712 | 5,435 |
| γ_2 | 5,96 | 2,086 | 1,775 | 5,996 | 9,945 |

DIC = -78,407 e **BIC** = -38,03436.

Gráficos das distribuições a *posteriori* marginais para os parâmetros do modelo 7. estão na Figura 7.11. Na Figura 7.12 temos uma indicação de convergência para todos os parâmetros através do gráfico de convergência das cadeias do modelo 7. Através da Figura 7.13 as estatísticas de Gelman-Rubin nos mostram que à medida que aumentamos o número de iterações, temos um indicativo de convergência para todos os parâmetros do modelo 7.

Figura 7.11: Densidades *a posteriori* marginais.



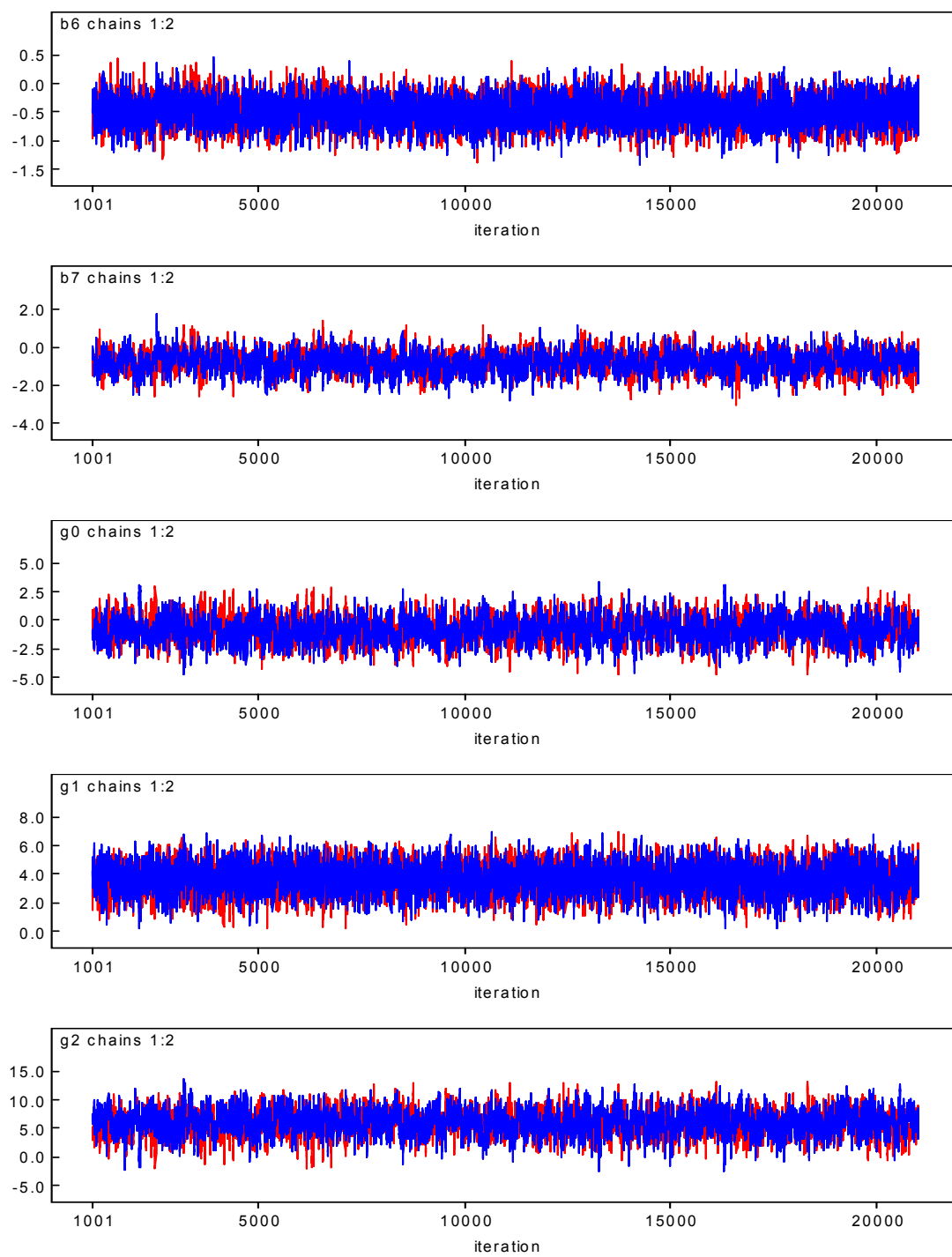


Figura 7.12: Gráficos de convergência das cadeias para o modelo 7.

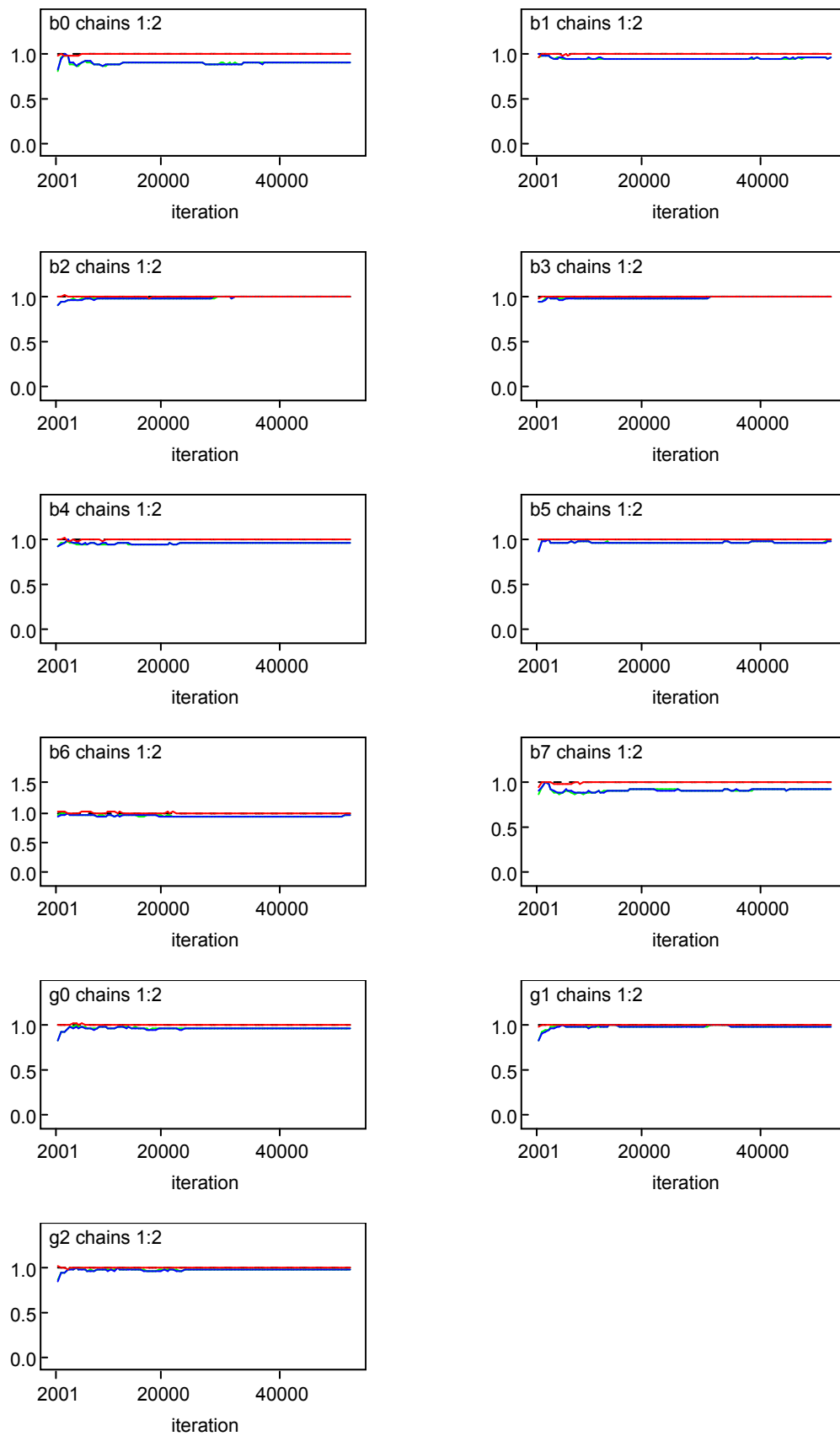


Figura 7.13: Gráficos dos índices R de Gelman e Rubin.

7.8.3 Modelo 8

Para simular a condição onde são utilizadas prioris informativas, utilizamos os dados históricos presentes no artigo de Keiser e Soss (1998) para definição das prioris. Em cada histórico os parâmetros são estimados utilizando o método de máxima verossimilhança, e, a partir desta seqüência de estimadores para cada um dos parâmetros, teríamos, então, a média amostral e a variância amostral para cada parâmetro. Assim, supomos densidades normais para construir densidades a priori para os parâmetros. Estas prioris são, então, utilizadas na obtenção da distribuição a *posteriori* do modelo bayesiano para os dados de 1991. Assim, foram consideradas as seguintes distribuições a priori:

$$\begin{aligned}\beta_0 &\sim N(2,77; 0,21), \beta_1 \sim N(0,107; 4,0E2), \beta_2 \sim N(-0,17; 73), \\ \beta_3 &\sim N(-0,08; 2,0E3), \beta_4 \sim N(-0,20; 7), \beta_5 \sim N(0,03; 2), \beta_6 \sim N(-0,14; 15), \\ \beta_7 &\sim N(-0,12; 8), \gamma_0 \sim N(-2; 1), \gamma_1 \sim N(0,54; 29), \gamma_2 \sim N(0,18; 1),\end{aligned}$$

tem-se que as f.d.p.'s desses parâmetros são

$$\begin{aligned}p(\beta_0) &= \frac{1}{0,21\sqrt{2\pi}} \exp\left(\frac{-\beta_0^2 + 5,54\beta_0 - 7,6729}{0,0882}\right), \\ p(\beta_1) &= \frac{1}{0,625\sqrt{2\pi}} \exp\left(\frac{-\beta_1^2 + 0,214\beta_1 - 0,011449}{0,78125}\right), \\ p(\beta_2) &= \frac{1}{73\sqrt{2\pi}} \exp\left(\frac{-\beta_2^2 - 0,34\beta_2 - 0,289}{10658}\right), \\ p(\beta_3) &= \frac{1}{0,125\sqrt{2\pi}} \exp\left(\frac{-\beta_3^2 - 0,16\beta_3 - 0,0064}{0,03125}\right), \\ p(\beta_4) &= \frac{1}{7\sqrt{2\pi}} \exp\left(\frac{-\beta_4^2 - 0,4\beta_4 - 0,04}{98}\right), \\ p(\beta_5) &= \frac{1}{2\sqrt{2\pi}} \exp\left(\frac{-\beta_5^2 + 0,06\beta_5 - 0,0009}{8}\right), \\ p(\beta_6) &= \frac{1}{15\sqrt{2\pi}} \exp\left(\frac{-\beta_6^2 - 0,28\beta_6 - 0,0196}{450}\right),\end{aligned}$$

$$\begin{aligned}
p(\beta_7) &= \frac{1}{8\sqrt{2\pi}} \exp\left(\frac{-\beta_7^2 - 0,24\beta_7 - 0,0144}{128}\right), \\
p(\gamma_0) &= \frac{1}{1\sqrt{2\pi}} \exp\left(\frac{-\gamma_0^2 - 4\gamma_0 - 4}{2}\right), \\
p(\gamma_1) &= \frac{1}{29\sqrt{2\pi}} \exp\left(\frac{-\gamma_1^2 - 1,08\gamma_1 - 0,2916}{1682}\right), \\
p(\gamma_2) &= \frac{1}{1\sqrt{2\pi}} \exp\left(\frac{-\gamma_2^2 + 0,36\gamma_2 - 0,0324}{2}\right).
\end{aligned}$$

Usando a regra de Bayes para combinar estas distribuições a *priori* com a função de verossimilhança dada em (6.2), a seguinte distribuição a *posteriori* conjunta para os β 's e os γ 's é obtida

$$\begin{aligned}
& p(\beta_0, \beta_1, \dots, \beta_7, \gamma_0, \gamma_1, \gamma_2 | \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
\propto & \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \gamma_1 z_{k1} + \gamma_2 z_{k2}))] \\
& \times \prod_{k=1}^n \left[\Gamma\left(\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \gamma_2 z_{k2}) \exp(\beta_0 + \sum_{i=1}^7 \beta_i x_{ki})}{1 + \exp(\beta_0 + \sum_{i=1}^7 \beta_i x_{ki})}\right) \right]^{-1} \\
& \times \prod_{k=1}^n \left[\Gamma\left(\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \sum_{i=1}^7 \beta_i x_{ki})}\right) \right]^{-1} \\
& \times \prod_{i=1}^n y_i \left[\frac{(\exp(\gamma_0 + \gamma_1 z_{k1} + \gamma_2 z_{k2}) - 1) \exp(\beta_0 + \sum_{i=1}^7 \beta_i x_{ki}) - 1}{1 + \exp(\beta_0 + \sum_{i=1}^7 \beta_i x_{ki})} \right] \\
& \times \prod_{i=1}^n (1 - y_i) \left[\frac{\exp(\gamma_0 + \gamma_1 z_{k1} + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \sum_{i=1}^7 \beta_i x_{ki})} - 1 \right] \\
& \times \exp\left(\frac{-\beta_0^2 + 5,54\beta_0}{0,0882}\right) \exp\left(\frac{-\beta_1^2 + 0,214\beta_1}{0,78125}\right) \exp\left(\frac{-\beta_2^2 - 0,34\beta_2}{10658}\right) \\
& \times \exp\left(\frac{-\beta_3^2 - 0,16\beta_3}{0,03125}\right) \exp\left(\frac{-\beta_4^2 - 0,4\beta_4}{98}\right) \exp\left(\frac{-\beta_5^2 - 0,06\beta_5}{8}\right) \\
& \times \exp\left(\frac{-\beta_6^2 + 0,28\beta_6}{450}\right) \exp\left(\frac{-\beta_7^2 + 0,24\beta_7}{128}\right) \exp\left(\frac{-\gamma_0^2 + 4\gamma_0}{2}\right) \\
& \times \exp\left(\frac{-\gamma_1^2 - 108\gamma_1}{1682}\right) \exp\left(\frac{-\gamma_2^2 - 0,36\gamma_2}{2}\right).
\end{aligned}$$

Como essa distribuição a *posteriori* é, também, intratável analiticamente, novamente será proposto o procedimento de *Gibbs Sampling* para retirar amostras da distribuição a *posteriori*. Para isso, é importante obter as distribuições a *posteriori* condicionais de

cada parâmetro.

i) Distribuição a *posteriori* condicional de β_0 dado $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_0 é

$$\begin{aligned}
& p(\beta_0 | \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
\propto & \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \times \exp \left(\frac{-\beta_0^2 + 5,54\beta_0}{0,0882} \right).
\end{aligned}$$

ii) Distribuição a *posteriori* condicional de β_1 dado $\beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_1 é

$$\begin{aligned}
& p(\beta_1 | \beta_0, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
\propto & \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \times \exp \left(\frac{-\beta_1^2 + 0,214\beta_1}{0,78125} \right).
\end{aligned}$$

iii) Distribuição a *posteriori* condicional de β_2 dado $\beta_0, \beta_1, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_2 é

$$\begin{aligned}
& p(\beta_2 | \beta_0, \beta_1, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_2^2 - 0,34\beta_2}{10658} \right).
\end{aligned}$$

iv) Distribuição a *posteriori* condicional de β_3 dado $\beta_0, \beta_1, \beta_2, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_2, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_3 é

$$\begin{aligned}
& p(\beta_3 | \beta_0, \beta_1, \beta_2, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\beta_3^2 - 0,16\beta_3}{0,03125} \right).
\end{aligned}$$

v) Distribuição a *posteriori* condicional de β_4 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_4 é

$$\begin{aligned}
& p(\beta_4 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \times \exp \left(\frac{-\beta_4^2 - 0,4\beta_4}{98} \right).
\end{aligned}$$

vi) Distribuição a *posteriori* condicional de β_5 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_5 é

$$\begin{aligned}
& p(\beta_5 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \times \exp \left(\frac{-\beta_5^2 - 0,06\beta_5}{8} \right).
\end{aligned}$$

vii) Distribuição a *posteriori* condicional de β_6 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_7, \gamma_0, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_7, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_6 é

$$\begin{aligned}
& p(\beta_6 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_7, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \times \exp \left(\frac{-\beta_6^2 + 0, 28\beta_6}{450} \right).
\end{aligned}$$

viii) Distribuição a *posteriori* condicional de β_7 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \gamma_0, \gamma_1, \gamma_2$.

E ainda, dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \gamma_0, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de β_7 é

$$\begin{aligned}
& p(\beta_7 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \gamma_0, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \times \exp \left(\frac{-\beta_7^2 + 0, 24\beta_7}{128} \right).
\end{aligned}$$

ix) Distribuição a *posteriori* condicional de γ_0 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_1, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_1, \gamma_2$, a distribuição a *posteriori* condicional de γ_0 é

$$\begin{aligned}
& p(\gamma_0 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_1, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_2 z_{k2}))]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\gamma_0^2 + 4\gamma_0}{2} \right).
\end{aligned}$$

x) Distribuição a *posteriori* condicional de γ_1 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_2$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_2$, a distribuição a *posteriori* condicional de γ_1 é

$$\begin{aligned}
& p(\gamma_1 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_2, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_2 z_{k2}))]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\gamma_1^2 - 108\gamma_1}{1682} \right).
\end{aligned}$$

xi) Distribuição a *posteriori* condicional de γ_2 dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1$.

Dado $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1$, a distribuição a *posteriori* condicional de γ_2 é

$$\begin{aligned}
& p(\gamma_2 | \beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \\
& \propto \prod_{k=1}^n [\Gamma(\exp(\gamma_0 + \dots + \gamma_2 z_{k2}))]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n \left[\Gamma \left(\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right) \right]^{-1} \\
& \quad \times \prod_{k=1}^n y_k \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2} + \beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right] \\
& \quad \times \prod_{k=1}^n (1 - y_k) \left[\frac{\exp(\gamma_0 + \dots + \gamma_2 z_{k2})}{1 + \exp(\beta_0 + \beta_1 x_{k1} + \dots + \beta_7 x_{k7})} \right]^{-1} \\
& \quad \times \exp \left(\frac{-\gamma_2^2 - 0.36\gamma_2}{2} \right).
\end{aligned}$$

Gerar uma amostra de $\beta_0 | \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, y, X$ não é fácil, pois como a distribuição de $\beta_0 | \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \gamma_0, \gamma_1, \gamma_2, y, X$ não pertence a uma família de distribuição conhecida, assim como as demais condicionais, com isso, é utilizado um procedimento Monte Carlo em Cadeias de Markov-MCMC (Carlin e Gelfand, 1991), mais precisamente *Metropolis Hastings*, para gerar uma amostra aproximada desta distribuição a *posteriori*.

As estimativas são dadas pela média das cadeias geradas a partir da distribuição a *posteriori* de cada parâmetro do modelo, obtidas via Método de Monte Carlo em Cadeias de Markov-MCMC (Carlin e Gelfand, 1991).

O processo de estimação para o modelo de regressão beta no Software WinBugs é extremamente suscetível aos valores iniciais e, os valores verdadeiros dos parâmetros estão entre zero e um, então propomos valores iniciais para a geração das cadeias nesse mesmo intervalo. Portanto, serão geradas duas cadeias com os seguintes valores iniciais:

Cadeia 1: $(\beta_0 = 0, \beta_1 = 0, \beta_2 = 0, \beta_3 = 0, \beta_4 = 0, \beta_5 = 0, \beta_6 = 0, \beta_7 = 0, \gamma_0 = 0, \gamma_1 = 0, \gamma_2 = 0)$,

Cadeia 2: $(\beta_0 = 0, 5; \beta_1 = 0, 5; \beta_2 = 0, 5; \beta_3 = 0, 5; \beta_4 = 0, 5; \beta_5 = 0, 5; \beta_6 = 0, 5; \beta_7 = 0, 5; \gamma_0 = 0, 5; \gamma_1 = 0, 5; \gamma_2 = 0, 5)$.

Abaixo, na Tabela 7.11, temos o sumário a *posteriori* para os parâmetros do Modelo 8 e as estatísticas DIC e BIC.

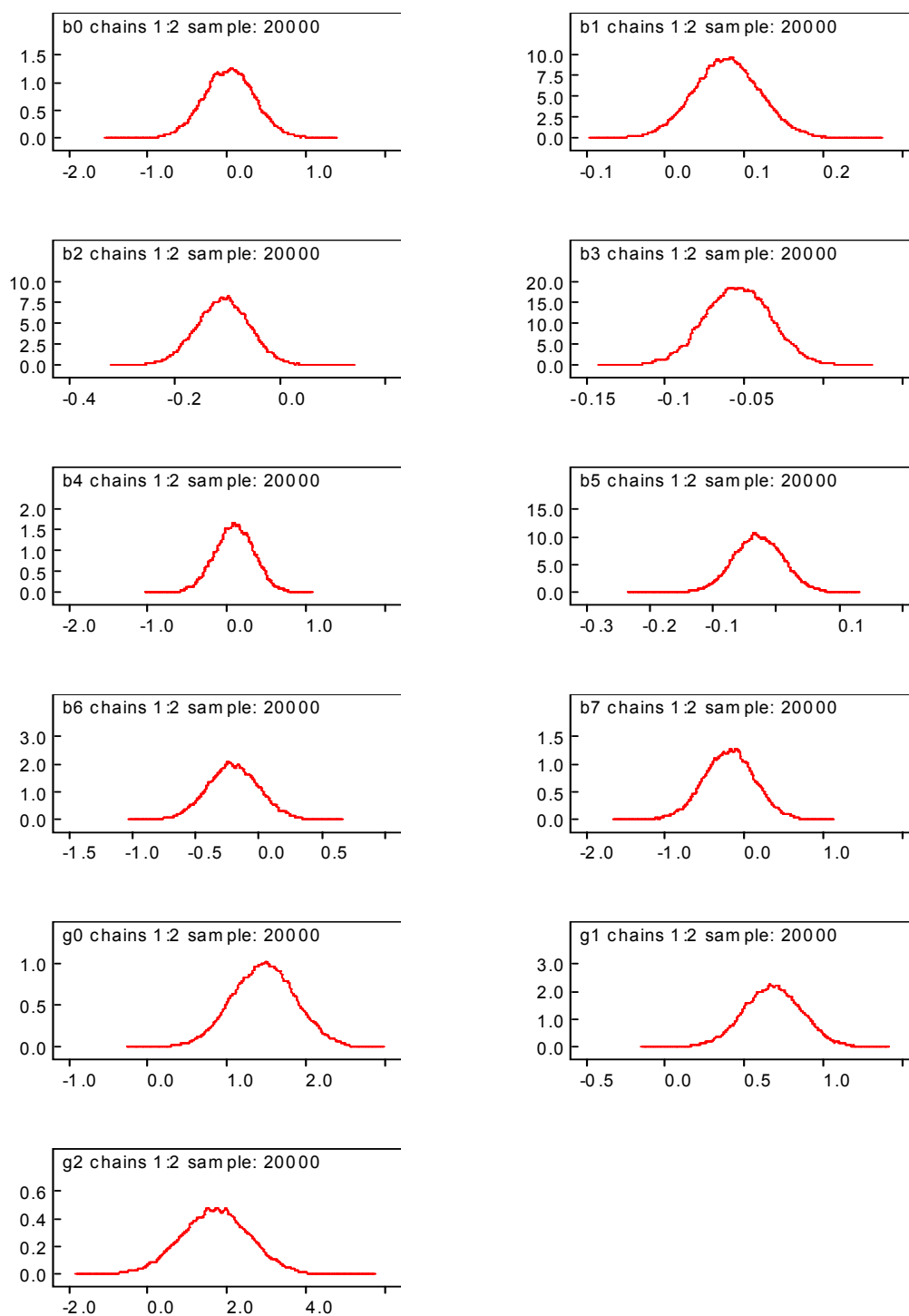
Tabela 7.11

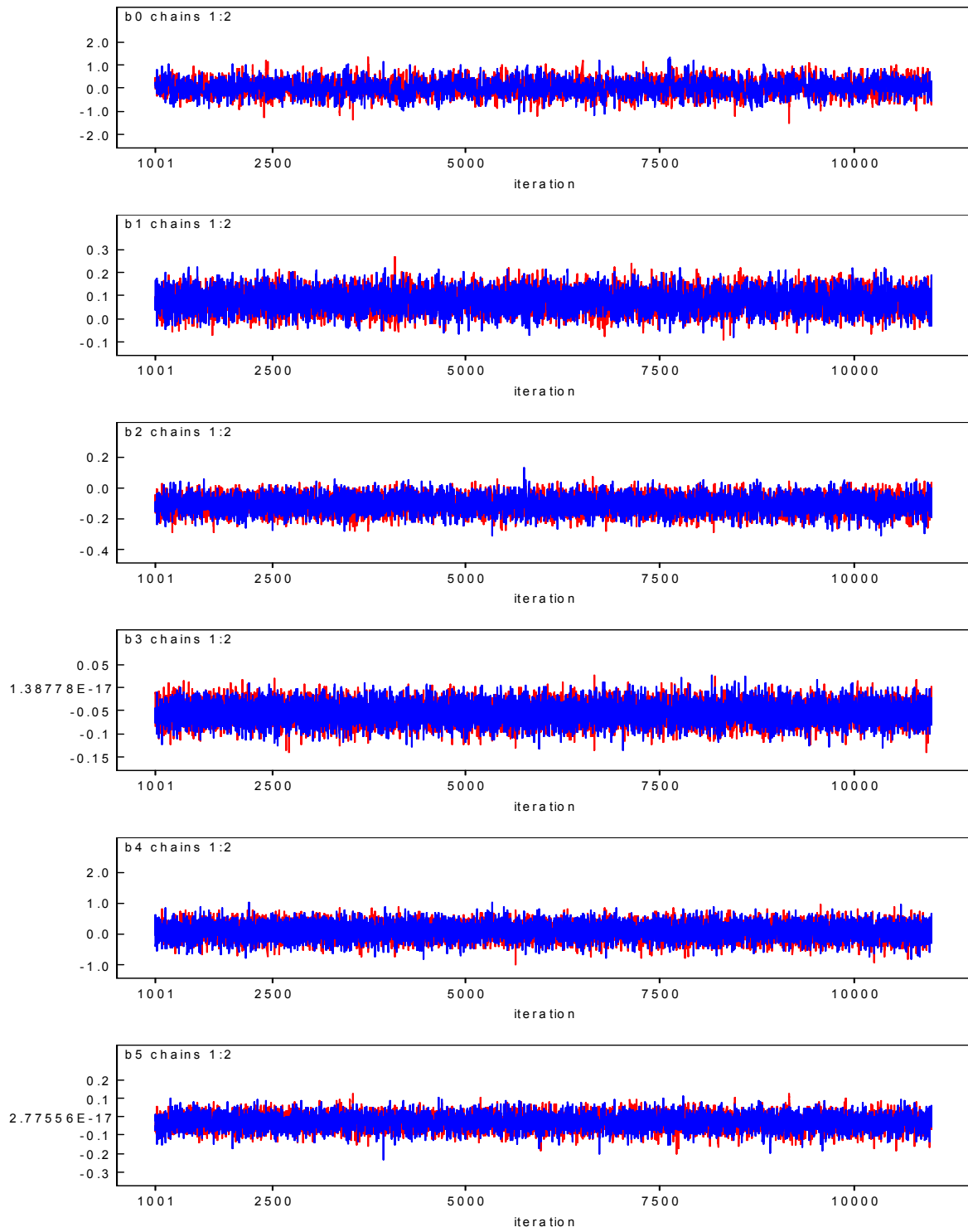
Sumário a *posteriori* para os parâmetros do **Modelo 8.**

| Parâmetros | Média | D.P. | 2,5% | mediana | 97,5% |
|------------|---------|--------|---------|---------|---------|
| β_0 | 0,0246 | 0,3209 | -0,6038 | 0,0271 | 0,654 |
| β_1 | 0,0778 | 0,0422 | -0,0046 | 0,0777 | 0,1618 |
| β_2 | -0,1083 | 0,0501 | -0,2072 | -0,1077 | -0,0110 |
| β_3 | -0,0543 | 0,0209 | -0,0955 | -0,0541 | -0,0140 |
| β_4 | 0,0889 | 0,246 | -0,3967 | 0,0916 | 0,5586 |
| β_5 | -0,0285 | 0,0389 | -0,1056 | -0,0287 | 0,0471 |
| β_6 | -0,203 | 0,2018 | -0,595 | -0,2055 | 0,1985 |
| β_7 | -0,1935 | 0,3203 | -0,83 | -0,1898 | 0,4347 |
| γ_0 | 1,464 | 0,3982 | 0,6842 | 1,467 | 2,246 |
| γ_1 | 0,6771 | 0,1825 | 0,3152 | 0,6773 | 1,03 |
| γ_2 | 1,706 | 0,8563 | 0,0234 | 1,713 | 3,385 |

DIC = -56,263 e **BIC** = -39,80741.

Gráficos das distribuições a *posteriori* marginais para os parâmetros do modelo 8 são mostrados na Figura 7.14. Na Figura 7.15 temos uma indicação de convergência para todos os parâmetros através do gráfico de convergência da cadeias do modelo 8. Através da Figura 7.16 as estatísticas de Gelman-Rubin nos mostram que à medida que aumentamos o número de iterações, temos um indicativo de convergência para todos os parâmetros do modelo 8.

Figura 7.14: Densidades *a posteriori marginais*.



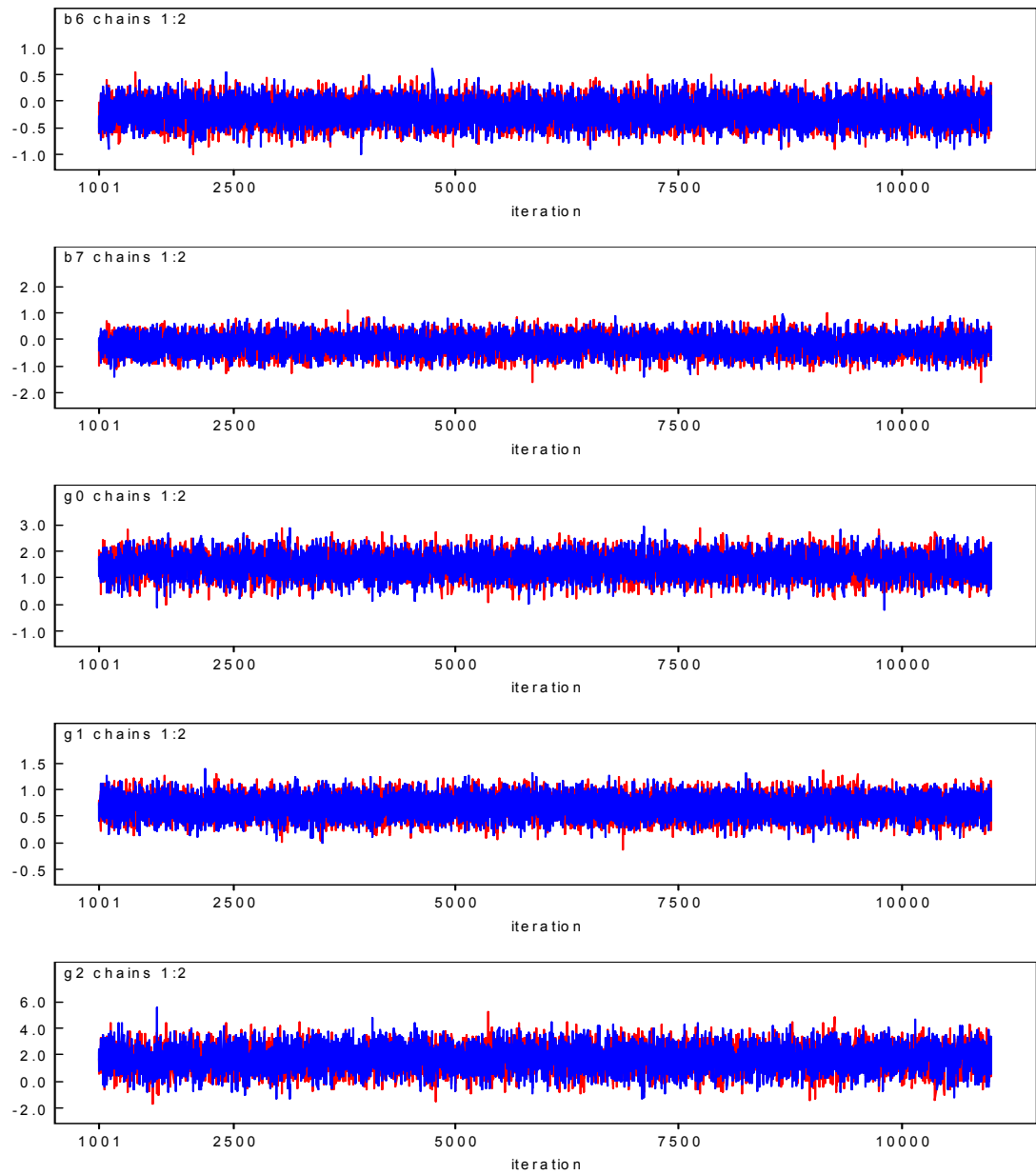


Figura 7.15: Gráficos de convergência das cadeias para o modelo 8.

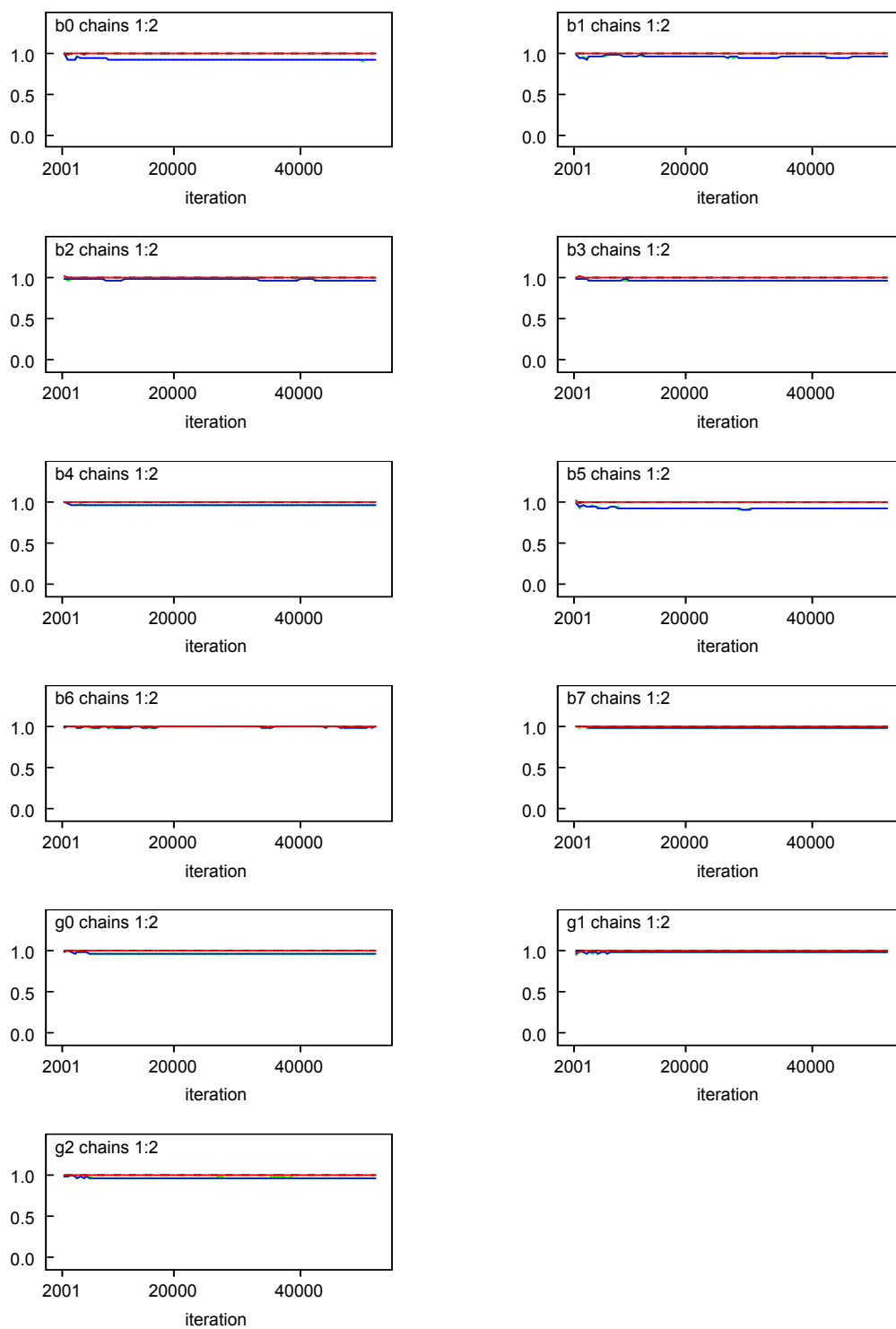


Figura 7.16: Gráficos dos índices R de Gelman e Rubin.

Comparação dos Modelos Bayesianos

Nesta seção comparamos os modelos Bayesianos levando em conta o conjunto de dados apresentado na aplicação. A comparação entre o **Modelo 6** e o **Modelo 7**. Para cada modelo foram geradas 52000 interações e foram descartadas as primeiras 2000 interações. Para evitar problemas de correlação nas cadeias foi dado um salto de tamanho 10 em cada simulação. Nos dois casos os históricos das cadeias de Markov para os parâmetros indicam que não existem problemas de convergência nas cadeias geradas e, visualmente houve convergência das estatísticas de Gelman-Rubin. Os valores dos critérios de informação BIC e DIC favorecem o **Modelo 6** (veja tabela 7.12) e justifica a utilização de um modelo de regressão beta sem o sub-modelo dispersão.

Tabela 7.12: Comparação entre os modelos 6 e 7.

| Modelos | DIC | BIC |
|----------|---------|-----------|
| Modelo 6 | -61.883 | -30.21453 |
| Modelo 7 | -78.407 | -38.03436 |

Apresentamos, agora, a comparação entre o **Modelo 6**, o **Modelo 7** e o **Modelo 8**, os dois primeiros foram ajustados considerando prioris não informativas, enquanto o último foi considerando prioris informativas. Para cada modelo foram geradas 21000 interações e foram descartadas as primeiras 1000 interações para fazer inferências com as distribuições *a posteriori*. Para evitar problemas de correlação nas cadeias foi dado um salto de tamanho 10 em cada simulação. Nos dois casos os históricos das cadeias de Markov para os parâmetros indicam que não existem problemas de convergência nas cadeias geradas e, visualmente houve convergência das estatísticas de Gelman-Rubin. Os valores dos critérios de informação BIC e DIC favorecem o **Modelo 8** (veja tabela 7.13) e justifica a utilização de um modelo de regressão beta com prioris informativas e com o sub-modelo dispersão.

Tabela 7.13: Comparação entre os modelos 6, 7 e 8.

| Modelos | DIC | BIC |
|----------|---------|-----------|
| Modelo 6 | -61.883 | -30.21453 |
| Modelo 7 | -78.407 | -38.03436 |
| Modelo 8 | -56.263 | -39.80741 |

Capítulo 8

Considerações finais

A metodologia de Regressão Beta é recomendável para situações onde se deseja modelar uma variável que seja representada por taxas ou proporções e, ainda, essa variável seja explicável através de outras covariáveis. No ajuste de um modelo de regressão beta, são inúmeras as dificuldades computacionais, por ser uma abordagem bastante recente na literatura e os métodos não estarem completamente implementados e validados, portanto, ao implementar um modelo de regressão beta em um dado pacote estatístico, devemos ser bem cautelosos. No WinBugs, por exemplo, todo o trabalho fica bastante sensível aos valores iniciais e às escalas das covariáveis.

No tratamento de dados simulados, em geral, verificou-se que os dados gerados estão bem próximas aos dados observados, o que evidencia um bom ajuste. Avaliamos, ainda, os resultados assintóticos em amostras finitas, utilizando simulação. Por fim, percebemos que nesse modelo de regressão beta, em que a resposta média é modelada através de uma estrutura de regressão, adotando a função logito como função de ligação, os estimadores de máxima verossimilhança para os parâmetros de regressão apresentam boas propriedades em amostras finitas, pois mostraram-se quase não-viciados e com uma distribuição próxima da distribuição normal de referência.

Sob o enfoque clássico, tanto no ajuste do modelo, utilizando a função de ligação logito, quanto a probito, as covariáveis: *Representantes Financeiros*, x_1 , *A variedade de Charter Schools*, x_2 , *Tempo*, x_5 , *Dimensões da População*, x_6 , e *Medidas Financeiras*, x_7 , não foram significativas ao nível de 5% . Mas, como visto na Tabela 7.5, a função de ligação logito gerou menor soma de quadrados dos resíduos.

Sob a metodologia clássica, fizemos uma comparação entre os modelos de regressão beta ajustados com as funções de ligações logito e probito, respectivamente e, notamos que a função de ligação logito se ajusta melhor aos dados, fizemos também uma comparação em um ajuste através de regressão beta e regressão linear. E verificamos que a regressão beta, em geral, apresentou vantagens quanto as médias, erros quadráticos médios e intervalos de confiança.

Sob a metodologia bayesiana, fizemos uma comparação entre os modelos com prioris vagas sem o sub-modelo de dispersão e com o sub-modelo de dispersão, respectivamente e, notamos que o modelo com priori vaga sem o sub-modelo de dispersão se ajusta melhor aos dados, mas devemos lembrar que ao utilizar este modelo, deve se ter cuidado, pois deve-se conhecer previamente o parâmetro de dispersão. Caso contrário, é mais coerente trabalhar com o modelo admitindo um sub-modelo de dispersão Fizemos também uma comparação entre um ajuste através de regressão beta com prioris não informativas e informativas. E verificamos que a regressão beta com prioris informativas, apresentou vantagens quanto as médias, desvios padrões e intervalos de credibilidade.

.1 Referências

- [1] ATKINSON, A. C. (1985). Plots, Transformations and Regression, An Introduction to Graphical Methods of Diagnostic Regression Analysis. UK: Oxford Science Publications.
- [2] BUCKLEY, J. (2000a). Does Charter Scholl Competition Improve Tradicional Public Schools? Civic Report.
- [3] BUCKLEY, J. (2000b). Estimation of Models with Beta-Distributed Dependent Variables: A Replication and Extension of Paolino (2001). *Political Analysis* 11(1): 314-338.
- [4] BUCKLEY, JACK E KÚSCOVÁ, SIMONA (2003). The Effect of Institutional Variation on Policy Outcomes – The Case of Charter Schools in the States, Boston College and Stony Brook University.
- [5] CARLIN, B.;GELFAND, A. E. (1991). An Iterative Monte Carlo Method for Nonconjugate Bayesian Analysis. *Statistics and Computation*, 1, pp. 119-128.
- [6] COOK, R. D. (1977). Detection of Influential Observations in Linear Regressions. *Technometrics* 19, 15 – 18.
- [7] COOK, R. D.;WEISBERG,S. (1982). Residuals and Influence in Regression. Chapman and Hall, New York.
- [8] DEMÉTRIO, Clarice G. B. Modelos Lineares Generalizados em Experimentação Agronômica. ESALQ/USP, 2002.
- [9] ESPINHEIRA, P. L.; FERRARI, S. L. P.;NETO, F. C. (2006). On Beta Residuals. SINAPE – 2006.
- [10] FERRARI, SILVIA L. P.; NETO, FRANCISCO CRIBARI J. (2004), Beta Regression For Rates and Proportions, *Journal of Applied Statistics*, v. 31, n. 7, p. 799-815.
- [11] FOX, J. (2002). An R and S-Plus Companion to Applied Regression. London: Sage Publications.
- [12] GELMAN, A.; CARLIN,J. B.; STERN, H. S. e RUBIN, D. B.. (1995). Bayesian Data Analysis. London: Chapman and Hall.
- [13] GILKS, W. R.; RICHARDSON, S. e Spiegelhalter, D. J. (eds) (1996). Markov

Chain Monte Carlo in Practice. London: Chapman and Hall.

[14] JACKMAN, SIMON. 2000. Estimation and inference via Bayesian Simulation. *American Journal of Political Science* 44(2): 375-404.

[15] JOHNSON, N. L.; KOTZ, S.; BALAKRISHNAN, N. (1995). *Continuous Univariate Distributions*, vol. 2. 2^o edition. New York: Wiley.

[16] KEISER, L. R., SOSS, J. (1998). With Good Cause: Bureaucratic Discretion and the Politics of Child Support Enforcement. *American Journal of Political Science* 42 (4): 1133-1156.

[17] KLAUS, L.P.; CRIBARI – NETO, F. Improved Maximum Likelihood Estimation in a New Class of Beta Regression Models. *Brazilian Journal of Probability and Statistics* (2005), 19, pp13-31.

[18] KRAINSKI, E.T. (2001), Modelo de Regressão Beta para a Posição Relativa da Espiga de Milho, UFPR.

[19] MARTIN, A. D. E QUINN, K.M. (2006). The MCMCpack Package. URL <http://mcmcpack.wustl.edu>.

[20] MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. 2nd ed., Chapman and Hall, Londres, 1989.

[21] NELDER, J. A.; WEDDERBURN, R. W. N. (1972). *Generalized Linear Models*. *Journal of the Royal Statistical Society A* 135, 370 - 384.

[22] PAOLINO, PHILIP. (2001). Maximum Likelihood Estimation of Models with Beta Distributed Dependent Variables. *Political Analysis* 9(4): 325-346.

[23] PAULA, G.A. (2004) *Modelos de Regressão com Apoio Computacional*. São Paulo. Edusp.

[24] PAULINO, C. D.; TURKMAN, M. A. A.; MURTEIRA, B. (2003) *Estatística Bayesiana*. Lisboa: Fundação Calouste Gulbenkian.

[25] RIPLEY, B.; VENABLES, W. (2000). *S Programming*. New York: Springer Verlag.

[26] SAS INSTITUTE (2005). *The GLIMMIX Procedure*. Cary, NC: SAS Institute.

[27] SIMAS, A.B. (2004), Beta regression for modeling rates and proportions, The betareg Package.

[28] SMITH, B. J. (2004). *Bayesian Output Analysis Program (BOA) for MCMC*. R

package version 1.1.2-1, URL <http://www.public-health.uiowa.edu/boa>.

[29] SMITHSON, M.; VERKUILEN, J. (2005). A Better lemon – Squeezer? Maximum Likelihood Regression with Beta – Distributed Dependent Variables. URL http://psychology.anu.edu.au/people/smithson/details/betareg/Smithson_Verkuilen06.pdf.

[30] SPIEGELHALTER, D. J.; THOMAS, A.; BEST, N. G.; LUNN, D. (2003). WinBugs Version 1.4 Users Manual. MRC Biostatistics Unit, Cambridge. URL <http://www.mrc-bsu.cam.ac.uk/bugs/>.

[31] TAMHANE, A. C., ANKENMAM, B. E. and YANG, Y. (2002). The bata Distribution as a Latent Response Model for Ordinal Data (I): Estimation of Location and Dispersion Parameters. Unpublished manuscript, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL.

[32] THOMAS, A.; O'HARA, R. B. (2004). OpenBugs. URL <http://mathstat.helsinki.fi/openbugs/>.

[33] VENABLES, W. N.; RIPLEY, B. D. (1999). Modern applied statistic with S-PLUS. New York: Wiley.

[34] WEI, B. C.; HU, Y. Q.; FUNG, W. K. (1998). Generalized Leverage and its Applications. *Scandinavian Journal of Statistics* 25, 25 – 37.

.2 Apêndices

.2.1 Apêndice A: Programa em R para avaliação das qualidades assintóticas dos estimadores

```
### Preparação #####  
n=500;  
library(betareg);  
set.seed(10);  
y=rbeta(n,5,3);  
x1=runif(n,0,1);  
x2=rnorm(n,0,1);  
x3=rt(n,4);  
x4=rexp(n,2);  
xdata=cbind(x1,x2,x3,x4);  
y=cbind(y);  
##### Primeiro ajuste #####  
mua=numeric();  
a=numeric();  
p=numeric();  
q=numeric();  
estimativas=matrix(NA,6,1000)  
y1=matrix(NA,n1,1000);  
x1a=numeric();  
x2a=numeric();  
x3a=numeric();  
x4a=numeric();  
trv20=numeric();  
cont1a=0;  
cont2a=0;  
cont3a=0;
```

```

cont4a=0;
for(t in 1:1000){
for(i in 1:n1){
x1a[i]=x1[i]
x2a[i]=x2[i]
x3a[i]=x3[i]
x4a[i]=x4[i]
a[i]=exp(b0 + b1*x1a[i] + b2*x2a[i] + b3*x3a[i] + b4*x4a[i])
mua[i]=(a[i])/(1+a[i])
p[i]=mua[i]*phi
q[i]=(1-mua[i])*phi
y1[i,t]=rbeta(1,p[i],q[i])}}
XA=cbind(x1a,x2a,x3a,x4a)
XAa=cbind(x1a,x2a)
for(k in 1:1000){
estimativas[1,k]=summary(betareg(y1[,k]~XA,link="logit"))$est[1]
estimativas[2,k]=summary(betareg(y1[,k]~XA,link="logit"))$est[2]
estimativas[3,k]=summary(betareg(y1[,k]~XA,link="logit"))$est[3]
estimativas[4,k]=summary(betareg(y1[,k]~XA,link="logit"))$est[4]
estimativas[5,k]=summary(betareg(y1[,k]~XA,link="logit"))$est[5]
estimativas[6,k]=summary(betareg(y1[,k]~XA,link="logit"))$est[6]
trv20[k]=2*(logLik(betareg(y1[,k]~XA, link="logit")) -
logLik(betareg(y1[,k]~XAa, link="logit")))}
for(k in 1:1000){
  if (trv20[k]>pchisq(0.90,2)) cont1a=cont1a+1 else cont1a=cont1a
  if (trv20[k]>pchisq(0.95,2)) cont2a=cont2a+1 else cont2a=cont2a
  if (trv20[k]>pchisq(0.99,2)) cont3a=cont3a+1 else cont3a=cont3a
  if (trv20[k]>pchisq(0.995,2)) cont4a=cont4a+1 else cont4a=cont4a
}
porcen_de_rejei1a=cont1a/1000;
porcen_de_rejei1a;

```

```
qqnorm(trv20,main="TRV n=20")
abline(0,1)
porcen_de_rejei2a=cont2a/1000;
porcen_de_rejei2a;
porcen_de_rejei3a=cont3a/1000;
porcen_de_rejei3a;
porcen_de_rejei4a=cont4a/1000;
porcen_de_rejei4a;
beta_zero=estimativas[1,]
beta_um=estimativas[2,]
beta_dois=estimativas[3,]
beta_tres=estimativas[4,]
beta_quatro=estimativas[5,]
phi=estimativas[6,]
beta_zero=as.numeric(beta_zero)
beta_um=as.numeric(beta_um)
beta_dois=as.numeric(beta_dois)
beta_tres=as.numeric(beta_tres)
beta_quatro=as.numeric(beta_quatro)
phi=as.numeric(phi)
m0=mean(beta_zero)
m1=mean(beta_um)
m2=mean(beta_dois)
m3=mean(beta_tres)
m4=mean(beta_quatro)
m5=mean(phi)
v0=var(beta_zero)
v1=var(beta_um)
v2=var(beta_dois)
v3=var(beta_tres)
v4=var(beta_quatro)
```



```
v5=var(phi)
medias_var=matrix(NA,6,2)
medias_var[,1]=cbind(m0,m1,m2,m3,m4,m5)
medias_var[,2]=cbind(v0,v1,v2,v3,v4,v5)
medias_var
#Intervalos de Confiança para os betas
#obtendo os dp's
fit1=betareg(y1[,1]~XA,link="logit")
dpb0<-fit1$stder[1]
dpb1<-fit1$stder[2]
dpb2<-fit1$stder[3]
dpb3<-fit1$stder[4]
dpb4<-fit1$stder[5]
dpphi<-fit1$stder[6]
#IC para beta0
icb0<-c(coef(fit1)[1]-qt(0.975,15)*dpb0,coef(fit1)[1]+qt(0.975,15)*dpb0)
#IC para beta1
icb1<-c(coef(fit1)[2]-qt(0.975,15)*dpb1,coef(fit1)[2]+qt(0.975,15)*dpb1)
#IC para beta2
icb2<-c(coef(fit1)[3]-qt(0.975,15)*dpb2,coef(fit1)[3]+qt(0.975,15)*dpb2)
#IC para beta3
icb3<-c(coef(fit1)[4]-qt(0.975,15)*dpb3,coef(fit1)[4]+qt(0.975,15)*dpb3)
#IC para beta4
icb4<-c(coef(fit1)[5]-qt(0.975,15)*dpb4,coef(fit1)[5]+qt(0.975,15)*dpb4)
#IC para phi
icphi<-c(coef(fit1)[6]-qt(0.975,15)*dpphi,coef(fit1)[6]+qt(0.975,15)*dpphi)
icb0
icb1
icb2
icb3
icb4
```

```
icphi
#### segundo ajuste ####
mub=numeric();
b=numeric();
p=numeric();
q=numeric();
estimativas=matrix(NA,6,1000)
y2=matrix(NA,n2,1000);
x1b=numeric();
x2b=numeric();
x3b=numeric();
x4b=numeric();
for(t in 1:1000){
for(i in 1:n2){
x1b[i]=x1[i]
x2b[i]=x2[i]
x3b[i]=x3[i]
x4b[i]=x4[i]
trv50=numeric();
cont1b=0;
cont2b=0;
cont3b=0;
cont4b=0;
b[i]=exp(b0 + b1*x1b[i] + b2*x2b[i] + b3*x3b[i] + b4*x4b[i])
mub[i]=(b[i])/(1+b[i])
p[i]=mub[i]*phi
q[i]=(1-mub[i])*phi
y2[i,t]=rbeta(1,p[i],q[i])}}
XB=cbind(x1b,x2b,x3b,x4b)
XBb=cbind(x1b,x2b)
for(k in 1:1000){
```

```

estimativas[1,k]=summary(betareg(y2[,k]~XB,link="logit"))$est[1]
estimativas[2,k]=summary(betareg(y2[,k]~XB,link="logit"))$est[2]
estimativas[3,k]=summary(betareg(y2[,k]~XB,link="logit"))$est[3]
estimativas[4,k]=summary(betareg(y2[,k]~XB,link="logit"))$est[4]
estimativas[5,k]=summary(betareg(y2[,k]~XB,link="logit"))$est[5]
estimativas[6,k]=summary(betareg(y2[,k]~XB,link="logit"))$est[6]
trv50[k]=2*(logLik(betareg(y2[,k]~XB, link="logit")) -
logLik(betareg(y2[,k]~XBb, link="logit"))))}
for(k in 1:1000){
  if (trv50[k]>pchisq(0.90,2)) cont1b=cont1b+1 else cont1b=cont1b
  if (trv50[k]>pchisq(0.95,2)) cont2b=cont2b+1 else cont2b=cont2b
  if (trv50[k]>pchisq(0.99,2)) cont3b=cont3b+1 else cont3b=cont3b
  if (trv50[k]>pchisq(0.995,2)) cont4b=cont4b+1 else cont4b=cont4b
}
porcen_de_rejei1b=cont1b/1000;
porcen_de_rejei1b;
qqnorm(trv50,main="TRV n=50")
abline(0,1)
porcen_de_rejei2b=cont2b/1000;
porcen_de_rejei2b;
porcen_de_rejei3b=cont3b/1000;
porcen_de_rejei3b;
porcen_de_rejei4b=cont4b/1000;
porcen_de_rejei4b;
beta_zero=estimativas[1,]
beta_um=estimativas[2,]
beta_dois=estimativas[3,]
beta_tres=estimativas[4,]
beta_quatro=estimativas[5,]
phi=estimativas[6,]
beta_zero=as.numeric(beta_zero)

```

```
beta_um=as.numeric(beta_um)
beta_dois=as.numeric(beta_dois)
beta_tres=as.numeric(beta_tres)
beta_quatro=as.numeric(beta_quatro)
phi=as.numeric(phi)
m0=mean(beta_zero)
m1=mean(beta_um)
m2=mean(beta_dois)
m3=mean(beta_tres)
m4=mean(beta_quatro)
m5=mean(phi)
v0=var(beta_zero)
v1=var(beta_um)
v2=var(beta_dois)
v3=var(beta_tres)
v4=var(beta_quatro)
v5=var(phi)
medias_var=matrix(NA,6,2)
medias_var[,1]=cbind(m0,m1,m2,m3,m4,m5)
medias_var[,2]=cbind(v0,v1,v2,v3,v4,v5)
medias_var
#Intervalos de Confiança para os betas
#obtendo os dp's
fit2=betareg(y2[,1]~XB,link="logit")
dpb0<-fit2$stder[1]
dpb1<-fit2$stder[2]
dpb2<-fit2$stder[3]
dpb3<-fit2$stder[4]
dpb4<-fit2$stder[5]
dpphi<-fit2$stder[6]
#IC para beta0
```

```
icb0<-c(coef(fit2)[1]-qt(0.975,45)*dpb0,coef(fit2)[1]+qt(0.975,45)*dpb0)
#IC para beta1
icb1<-c(coef(fit2)[2]-qt(0.975,45)*dpb1,coef(fit2)[2]+qt(0.975,45)*dpb1)
#IC para beta2
icb2<-c(coef(fit2)[3]-qt(0.975,45)*dpb2,coef(fit2)[3]+qt(0.975,45)*dpb2)
#IC para beta3
icb3<-c(coef(fit2)[4]-qt(0.975,45)*dpb3,coef(fit2)[4]+qt(0.975,45)*dpb3)
#IC para beta4
icb4<-c(coef(fit2)[5]-qt(0.975,45)*dpb4,coef(fit2)[5]+qt(0.975,45)*dpb4)
#IC para phi
icphi<-c(coef(fit2)[6]-qt(0.975,45)*dpphi,coef(fit2)[6]+qt(0.975,45)*dpphi)
icb0
icb1
icb2
icb3
icb4
icphi

#### segundo ajuste ###
muc=numeric();
c=numeric();
p=numeric();
q=numeric();
estimativas=matrix(NA,6,100)
y3=matrix(NA,n3,100);
x1c=numeric();
x2c=numeric();
x3c=numeric();
x4c=numeric();
trv100=numeric();
cont1c=0;
cont2c=0;
```

```

cont3c=0;
cont4c=0;
for(t in 1:100){
for(i in 1:n3){
x1c[i]=x1[i]
x2c[i]=x2[i]
x3c[i]=x3[i]
x4c[i]=x4[i]
c[i]=exp(b0 + b1*x1c[i] + b2*x2c[i] + b3*x3c[i] + b4*x4c[i])
muc[i]=(c[i])/(1+c[i])
p[i]=muc[i]*phi
q[i]=(1-muc[i])*phi
y3[i,t]=rbeta(1,p[i],q[i])}}
XC=cbind(x1c,x2c,x3c,x4c)
XCc=cbind(x1c,x2c)
for(k in 1:100){
estimativas[1,k]=summary(betareg(y3[,k]~XC,link="logit"))$est[1]
estimativas[2,k]=summary(betareg(y3[,k]~XC,link="logit"))$est[2]
estimativas[3,k]=summary(betareg(y3[,k]~XC,link="logit"))$est[3]
estimativas[4,k]=summary(betareg(y3[,k]~XC,link="logit"))$est[4]
estimativas[5,k]=summary(betareg(y3[,k]~XC,link="logit"))$est[5]
estimativas[6,k]=summary(betareg(y3[,k]~XC,link="logit"))$est[6]
trv100[k]=2*(logLik(betareg(y3[,k]~XC, link="logit")) -
logLik(betareg(y3[,k]~XCc, link="logit")))}
for(k in 1:1000){
  if (trv100[k]>pchisq(0.90,2)) cont1c=cont1c+1 else cont1c=cont1c
  if (trv100[k]>pchisq(0.95,2)) cont2c=cont2c+1 else cont2c=cont2c
  if (trv100[k]>pchisq(0.99,2)) cont3c=cont3c+1 else cont3c=cont3c
  if (trv100[k]>pchisq(0.995,2)) cont4c=cont4c+1 else cont4c=cont4c
}
porcen_de_rejeic=cont1c/1000;

```

```
porcen_de_rejei1c;
qqnorm(trv100,main="TRV n=100")
abline(0,1)
porcen_de_rejei2c=cont2c/1000;
porcen_de_rejei2c;
porcen_de_rejei3c=cont3c/1000;
porcen_de_rejei3c;
porcen_de_rejei4c=cont4c/1000;
porcen_de_rejei4c;
beta_zero=estimativas[1,]
beta_um=estimativas[2,]
beta_dois=estimativas[3,]
beta_tres=estimativas[4,]
beta_quatro=estimativas[5,]
phi=estimativas[6,]
beta_zero=as.numeric(beta_zero)
beta_um=as.numeric(beta_um)
beta_dois=as.numeric(beta_dois)
beta_tres=as.numeric(beta_tres)
beta_quatro=as.numeric(beta_quatro)
phi=as.numeric(phi)
m0=mean(beta_zero)
m1=mean(beta_um)
m2=mean(beta_dois)
m3=mean(beta_tres)
m4=mean(beta_quatro)
m5=mean(phi)
v0=var(beta_zero)
v1=var(beta_um)
v2=var(beta_dois)
v3=var(beta_tres)
```

```
v4=var(beta_quatro)
v5=var(phi)
medias_var=matrix(NA,6,2)
medias_var[,1]=cbind(m0,m1,m2,m3,m4,m5)
medias_var[,2]=cbind(v0,v1,v2,v3,v4,v5)
medias_var
#Intervalos de Confiança para os betas
#obtendo os dp's
fit3=betareg(y3[,1]~XC,link="logit")
dpb0<-fit3$std[1]
dpb1<-fit3$std[2]
dpb2<-fit3$std[3]
dpb3<-fit3$std[4]
dpb4<-fit3$std[5]
dpphi<-fit3$std[6]
#IC para beta0
icb0<-c(coef(fit3)[1]-qt(0.975,95)*dpb0,coef(fit3)[1]+qt(0.975,95)*dpb0)
#IC para beta1
icb1<-c(coef(fit3)[2]-qt(0.975,95)*dpb1,coef(fit3)[2]+qt(0.975,95)*dpb1)
#IC para beta2
icb2<-c(coef(fit3)[3]-qt(0.975,95)*dpb2,coef(fit3)[3]+qt(0.975,95)*dpb2)
#IC para beta3
icb3<-c(coef(fit3)[4]-qt(0.975,95)*dpb3,coef(fit3)[4]+qt(0.975,95)*dpb3)
#IC para beta4
icb4<-c(coef(fit3)[5]-qt(0.975,95)*dpb4,coef(fit3)[5]+qt(0.975,95)*dpb4)
#IC para phi
icphi<-c(coef(fit3)[6]-qt(0.975,95)*dpphi,coef(fit3)[6]+qt(0.975,95)*dpphi)
icb0
icb1
icb2
icb3
```



```
icb4
icphi

#### segundo ajuste ####
mud=numeric();
d=numeric();
p=numeric();
q=numeric();
estimativas=matrix(NA,6,100)
y4=matrix(NA,n4,100);
x1d=numeric();
x2d=numeric();
x3d=numeric();
x4d=numeric();
for(t in 1:100){
for(i in 1:n4){
x1d[i]=x1[i]
x2d[i]=x2[i]
x3d[i]=x3[i]
x4d[i]=x4[i]
d[i]=exp(b0 + b1*x1d[i] + b2*x2d[i] + b3*x3d[i] + b4*x4d[i])
mud[i]=(d[i])/(1+d[i])
p[i]=mud[i]*phi
q[i]=(1-mud[i])*phi
y4[i,t]=rbeta(1,p[i],q[i])}}
XD=cbind(x1d,x2d,x3d,x4d)
for(k in 1:100){
estimativas[1,k]=summary(betareg(y4[,k]~XD,link="logit"))$est[1]
estimativas[2,k]=summary(betareg(y4[,k]~XD,link="logit"))$est[2]
estimativas[3,k]=summary(betareg(y4[,k]~XD,link="logit"))$est[3]
estimativas[4,k]=summary(betareg(y4[,k]~XD,link="logit"))$est[4]
estimativas[5,k]=summary(betareg(y4[,k]~XD,link="logit"))$est[5]
```

```
estimativas[6,k]=summary(betareg(y4[,k]~XD,link="logit"))$est[6]
}
beta_zero=estimativas[1,]
beta_um=estimativas[2,]
beta_dois=estimativas[3,]
beta_tres=estimativas[4,]
beta_quatro=estimativas[5,]
phi=estimativas[6,]
beta_zero=as.numeric(beta_zero)
beta_um=as.numeric(beta_um)
beta_dois=as.numeric(beta_dois)
beta_tres=as.numeric(beta_tres)
beta_quatro=as.numeric(beta_quatro)
phi=as.numeric(phi)
m0=mean(beta_zero)
m1=mean(beta_um)
m2=mean(beta_dois)
m3=mean(beta_tres)
m4=mean(beta_quatro)
m5=mean(phi)
v0=var(beta_zero)
v1=var(beta_um)
v2=var(beta_dois)
v3=var(beta_tres)
v4=var(beta_quatro)
v5=var(phi)
medias_var=matrix(NA,6,2)
medias_var[,1]=cbind(m0,m1,m2,m3,m4,m5)
medias_var[,2]=cbind(v0,v1,v2,v3,v4,v5)
medias_var
#Intervalos de Confiança para os betas
```

```

#obtendo os dp's
fit4=betareg(y4[,1]~XD,link="logit")
dpb0<-fit4$stdr[1]
dpb1<-fit4$stdr[2]
dpb2<-fit4$stdr[3]
dpb3<-fit4$stdr[4]
dpb4<-fit4$stdr[5]
dpphi<-fit4$stdr[6]
#IC para beta0
icb0<-c(coef(fit4)[1]-qt(0.975,495)*dpb0,coef(fit4)[1]+qt(0.975,495)*dpb0)
#IC para beta1
icb1<-c(coef(fit4)[2]-qt(0.975,495)*dpb1,coef(fit4)[2]+qt(0.975,495)*dpb1)
#IC para beta2
icb2<-c(coef(fit4)[3]-qt(0.975,495)*dpb2,coef(fit4)[3]+qt(0.975,495)*dpb2)
#IC para beta3
icb3<-c(coef(fit4)[4]-qt(0.975,495)*dpb3,coef(fit4)[4]+qt(0.975,495)*dpb3)
#IC para beta4
icb4<-c(coef(fit4)[5]-qt(0.975,495)*dpb4,coef(fit4)[5]+qt(0.975,495)*dpb4)
#IC para phi
icphi<-c(coef(fit4)[6]-qt(0.975,495)*dpphi,coef(fit4)[6]+qt(0.975,495)*dpphi)
icb0
icb1
icb2
icb3
icb4
icphi

```

.2.2 Apêndice B: Programa em R para inferência Clássica usando as funções de ligação Logito e Probit

```
##### Programa 1 #####
```

```
library(betareg)
## Dados ##
## Os dados foram omitidos #####
y=cbind(yt)
xdata=cbind(x1,x2,x3,x4,x5,x6,x7)
## -----
##### ANÁLISE DESCRITIVA DOS DADOS
## -----
##### SUMÁRIOS
sumX1 = summary(x1)
sumX1 = c(sumX1, var(x1))
names(sumX1)[7]="Var"
sumX2 = summary(x2)
sumX2 = c(sumX2, var(x2))
names(sumX2)[7]="Var"
sumX3 = summary(x3)
sumX3 = c(sumX3, var(x3))
names(sumX3)[7]="Var"
sumX4 = summary(x4)
sumX4 = c(sumX4, var(x4))
names(sumX4)[7]="Var"
sumX5 = summary(x5)
sumX5 = c(sumX5, var(x5))
names(sumX5)[7]="Var"
sumX6 = summary(x6)
sumX6 = c(sumX6, var(x6))
names(sumX6)[7]="Var"
sumX7 = summary(x7)
sumX7 = c(sumX7, var(x7))
names(sumX7)[7]="Var"
sumY = summary(y)
```

```
sumY = c(sumY, var(y))
names(sumY)[7]="Var"
cat("\n\n");
cat("***** Analise Descritiva dos Dados *****", "\n\n");
cat("=== Sumário: variável Y ===", "\n\n");
sumY;
cat("\n\n");
cat("=== Sumário: variável X1 ===", "\n\n");
sumX1;
cat("=== Sumário: variável X2 ===", "\n\n");
sumX2;
cat("=== Sumário: variável X3 ===", "\n\n");
sumX3;
cat("=== Sumário: variável X4 ===", "\n\n");
sumX4
### ANÁLISE GRÁFICA INICIAL
hist(y)
boxplot(y)
identify(rep(1,length(y)),y)
plot(y)
qqnorm(y)
stem(y) ## gráfico ramo e folhas
par(mfrow=c(2,2))
hist(x1)
boxplot(x1)
plot(x1)
qqnorm(x1)
stem(x1) ## gráfico ramo e folhas
par(mfrow=c(2,2))
hist(x2)
boxplot(x2)
```

```
plot(x2)
qqnorm(x2)
stem(x2) ### gráfico ramo e folhas
par(mfrow=c(2,2))
hist(x3)
boxplot(x3)
plot(x3)
qqnorm(x3)
stem(x3) ### gráfico ramo e folhas
par(mfrow=c(2,2))
hist(x4)
boxplot(x4)
plot(x4)
qqnorm(x4)
stem(x4) ### gráfico ramo e folhas
par(mfrow=c(2,2))
hist(x5)
boxplot(x5)
plot(x5)
qqnorm(x5)
stem(x5) ### gráfico ramo e folhas
par(mfrow=c(2,2))
hist(x6)
boxplot(x6)
plot(x6)
qqnorm(x6)
stem(x6) ### gráfico ramo e folhas
par(mfrow=c(2,2))
hist(x7)
boxplot(x7)
plot(x7)
```

```
qqnorm(x7)
stem(x7) ## gráfico ramo e folhas
##### modelo1 - usando a ligação logito #####
fit1=betareg(y ~ xdata,link="logit")
summary.betareg(fit1)
logLik.betareg(fit1)
estfun.betareg(fit1)
aov(fit1)
deviance(fit1)
## DIAGNÓSTICOS DE COLINEARIDADE-verificando se a matriz de correlações
das covariáveis apresenta COLUNAS ALTAMENTE
## CORRELACIONADAS
x = model.matrix(fit1)
x = x[,-1]
round(cor(x),2)
#Intervalos de Confiança para os betas
#obtendo os dp's
dpb0<-fit1$stder[1]
dpb1<-fit1$stder[2]
dpb2<-fit1$stder[3]
dpb3<-fit1$stder[4]
dpb4<-fit1$stder[5]
dpb5<-fit1$stder[6]
dpb6<-fit1$stder[7]
dpb7<-fit1$stder[8]
dpphi<-fit1$stder[9]
#IC para beta0
icb0<-c(coef(fit1)[1]-qt(0.975,43)*dpb0,coef(fit1)[1]+qt(0.975,43)*dpb0)
#IC para beta1
icb1<-c(coef(fit1)[2]-qt(0.975,43)*dpb1,coef(fit1)[2]+qt(0.975,43)*dpb1)
#IC para beta2
```

```

icb2<-c(coef(fit1)[3]-qt(0.975,43)*dpb2,coef(fit1)[3]+qt(0.975,43)*dpb2)
#IC para beta3
icb3<-c(coef(fit1)[4]-qt(0.975,43)*dpb3,coef(fit1)[4]+qt(0.975,43)*dpb3)
#IC para beta4
icb4<-c(coef(fit1)[5]-qt(0.975,43)*dpb4,coef(fit1)[5]+qt(0.975,43)*dpb4)
#IC para beta5
icb5<-c(coef(fit1)[6]-qt(0.975,43)*dpb5,coef(fit1)[6]+qt(0.975,43)*dpb5)
#IC para beta6
icb6<-c(coef(fit1)[7]-qt(0.975,43)*dpb6,coef(fit1)[7]+qt(0.975,43)*dpb6)
#IC para beta7
icb7<-c(coef(fit1)[8]-qt(0.975,43)*dpb7,coef(fit1)[8]+qt(0.975,43)*dpb7)
#IC para phi
icphi<-c(coef(fit1)[9]-qt(0.975,43)*dpphi,coef(fit1)[9]+qt(0.975,43)*dpphi)
icb0
icb1
icb2
icb3
icb4
icb5
icb6
icb7
icphi
#####
## Testes ###
XAa=cbind(x3,x4)
fit1a=betareg(y~XAa,link="logit")
trv=2*(logLik(fit1)-logLik(fit1a))
trv
trv>pchisq(0.95,2)
####Diagnósticos #####
cooks.distance.betareg(fit1)

```



```
hatvalues.betareg(fit1)
vcov.betareg(fit1)
par(mfrow=c(1,1))
plot.betareg(fit1) ##gráficos p/ diagnósticos
par(mfrow=c(1,1))
envelope.beta(fit1)
preditos1=predict.betareg(fit1)
res_dev1=fit1$resd
res_pad1=fit1$resstd
res_ord1=fit1$residuals
preditos_residuos1=cbind(preditos1,res_dev1,res_pad1,res_ord1)
round(preditos_residuos1,4)
plot(preditos1,res_ord1,main="resíduos vs preditos")
identify(preditos1,res_ord1);
H1=fit1$h; ##### leverage
eta_chapeu1=fit1$etahat #####preditor linear ajustado
plot(eta_chapeu1,res_ord1,main="Verificando a função de ligação")
identify(eta_chapeu1,res_ord1);
gen.lev.betareg(fit1) #####leverage genralizado
leverage_gene1=gen.lev.betareg(fit1) #####leverage genralizado
diag(leverage_gene1)>=(16/51)
16/51; ###2k/n ###
leverage_gene1[4]
leverage_gene1[30]
###Análise sequencial #####
f1a=betareg(y ~1,link="logit")
f2a=betareg(y ~x3+x4,link="logit")
f3a=betareg(y ~x3+x4,link="logit")
f4a=betareg(y ~x3+x4+x2,link="logit")
f5a=betareg(y ~x3+x4+x1,link="logit")
f6a=betareg(y ~x3+x4+x5,link="logit")
```

```
f7a=betareg(y ~x3+x4+x6,link="logit")
f8a=betareg(y ~x3+x4+x2+x7,link="logit")
anova.betareg(f1a,f2a)
anova.betareg(f2a,f3a)
anova.betareg(f3a,f4a)
anova.betareg(f4a,f5a)
anova.betareg(f5a,f6a)
anova.betareg(f6a,f7a)
anova.betareg(f7a,f8a)
##### modelo1a - Usando a ligação logito - sem o ponto 4#####
y1a=y[-4]
xdata1=xdata[-4,]
fit1a=betareg(y1a ~xdata1,link="logit")
summary.betareg(fit1a)
logLik.betareg(fit1a)
estfun.betareg(fit1a)
aov(fit1a)
deviance(fit1a)
#Intervalos de Confiança para os betas
#obtendo os dp's
dpb0<-fit1a$stder[1]
dpb1<-fit1a$stder[2]
dpb2<-fit1a$stder[3]
dpb3<-fit1a$stder[4]
dpb4<-fit1a$stder[5]
dpb5<-fit1a$stder[6]
dpb6<-fit1a$stder[7]
dpb7<-fit1a$stder[8]
dpphi<-fit1a$stder[9]
#IC para beta0
icb0<-c(coef(fit1a)[1]-qt(0.975,42)*dpb0,coef(fit1a)[1]+qt(0.975,42)*dpb0)
```

```

#IC para beta1
icb1<-c(coef(fit1a)[2]-qt(0.975,42)*dpb1,coef(fit1a)[2]+qt(0.975,42)*dpb1)
#IC para beta2
icb2<-c(coef(fit1a)[3]-qt(0.975,42)*dpb2,coef(fit1a)[3]+qt(0.975,42)*dpb2)
#IC para beta3
icb3<-c(coef(fit1a)[4]-qt(0.975,42)*dpb3,coef(fit1a)[4]+qt(0.975,42)*dpb3)
#IC para beta4
icb4<-c(coef(fit1a)[5]-qt(0.975,42)*dpb4,coef(fit1a)[5]+qt(0.975,42)*dpb4)
#IC para beta5
icb5<-c(coef(fit1a)[6]-qt(0.975,42)*dpb5,coef(fit1a)[6]+qt(0.975,42)*dpb5)
#IC para beta6
icb6<-c(coef(fit1a)[7]-qt(0.975,42)*dpb6,coef(fit1a)[7]+qt(0.975,42)*dpb6)
#IC para beta7
icb7<-c(coef(fit1a)[8]-qt(0.975,42)*dpb7,coef(fit1a)[8]+qt(0.975,42)*dpb7)
#IC para phi
icphi<-c(coef(fit1a)[9]-qt(0.975,42)*dpphi,coef(fit1a)[9]+qt(0.975,42)*dpphi)
icb0
icb1
icb2
icb3
icb4
icb5
icb6
icb7
icphi
#####
####Diagnósticos #####
cooks.distance.betareg(fit1a)
hatvalues.betareg(fit1a)
vcov.betareg(fit1a)
par(mfrow=c(1,1))

```

```

plot.betareg(fit1a) ##gráficos p/ diagnósticos
par(mfrow=c(1,1))
envelope.beta(fit1a)
preditos1a=predict.betareg(fit1a)
res_dev1a=fit1a$resd
res_pad1a=fit1a$resstd
res_ord1a=fit1a$residuals
preditos_residuos1a=cbind(preditos1a,res_dev1a,res_pad1a,res_ord1a)
round(preditos_residuos1a,4)
plot(preditos1a,res_ord1a,main="resíduos vs preditos")
identify(preditos1a,res_ord1a);
H1=fit1a$h; ##### leverage
eta_chapeula=fit1a$etahat #####preditor linear ajustado
plot(eta_chapeula,res_ord1a,main="Verificando a função de ligação")
identify(eta_chapeula,res_ord1a);
gen.lev.betareg(fit1a) #####leverage genralizado
leverage_gene1a=gen.lev.betareg(fit1a) #####leverage genralizado
diag(leverage_gene1a)>=(16/50)
16/50; ###2k/n ###
leverage_gene1a[35]
leverage_gene1a[44]
##### modelo2 - Usando a ligação probito#####
fit2=betareg(y1a ~ xdata1,link="probit")
summary.betareg(fit2)
logLik.betareg(fit2)
estfun.betareg(fit2)
aov(fit2)
deviance(fit2)
#Intervalos de Confiança para os betas
#obtendo os dp's
dpb0<-fit2$stder[1]

```

```
dpb1<-fit2$stder[2]
dpb2<-fit2$stder[3]
dpb3<-fit2$stder[4]
dpb4<-fit2$stder[5]
dpb5<-fit2$stder[6]
dpb6<-fit2$stder[7]
dpb7<-fit2$stder[8]
dpphi<-fit2$stder[9]
#IC para beta0
icb0<-c(coef(fit2)[1]-qt(0.975,42)*dpb0,coef(fit2)[1]+qt(0.975,42)*dpb0)
#IC para beta1
icb1<-c(coef(fit2)[2]-qt(0.975,42)*dpb1,coef(fit2)[2]+qt(0.975,42)*dpb1)
#IC para beta2
icb2<-c(coef(fit2)[3]-qt(0.975,42)*dpb2,coef(fit2)[3]+qt(0.975,42)*dpb2)
#IC para beta3
icb3<-c(coef(fit2)[4]-qt(0.975,42)*dpb3,coef(fit2)[4]+qt(0.975,42)*dpb3)
#IC para beta4
icb4<-c(coef(fit2)[5]-qt(0.975,42)*dpb4,coef(fit2)[5]+qt(0.975,42)*dpb4)
#IC para beta5
icb5<-c(coef(fit2)[6]-qt(0.975,42)*dpb5,coef(fit2)[6]+qt(0.975,42)*dpb5)
#IC para beta6
icb6<-c(coef(fit2)[7]-qt(0.975,42)*dpb6,coef(fit2)[7]+qt(0.975,42)*dpb6)
#IC para beta7
icb7<-c(coef(fit2)[8]-qt(0.975,42)*dpb7,coef(fit2)[8]+qt(0.975,42)*dpb7)
#IC para phi
icphi<-c(coef(fit2)[9]-qt(0.975,42)*dpphi,coef(fit2)[9]+qt(0.975,42)*dpphi)
icb0
icb1
icb2
icb3
icb4
```

```

icb5
icb6
icb7
icphi
#####
## Testes ###
XAb=cbind(x3,x4)
fit2a=betareg(y~XAb,link="probit")
trv=2*(logLik(fit1)-logLik(fit2a))
trv
trv>pchisq(0.95,2)
####Diagnósticos #####
cooks.distance.betareg(fit2)
hatvalues.betareg(fit2)
vcov.betareg(fit2)
plot.betareg(fit2) ##gráficos p/ diagnósticos
par(mfrow=c(1,1))
envelope.beta(fit2)
preditos2=predict.betareg(fit2)
res_dev2=fit2$resd
res_pad2=fit2$resstd
res_ord2=fit2$residuals
preditos_residuos2=cbind(preditos2,res_dev2,res_pad2,res_ord2)
round(preditos_residuos2,4)
plot(preditos2,res_ord2,main="resíduos vs preditos")
identify(preditos2,res_ord2);
H2=fit2$h; ##### leverage
eta_chapeu2=fit2$etahat #####preditor linear ajustado
plot(eta_chapeu2,res_ord2,main="Verificando a função de ligação")
identify(eta_chapeu2,res_ord2);
gen.lev.betareg(fit2) #####levegare genralizado

```

```

leverage_gene2=gen.lev.betareg(fit2) #####leverage genralizado
diag(leverage_gene2)>=(16/50)
16/50; #####2k/n ###
leverage_gene2[35]
leverage_gene2[44]
###Análise sequencial #####
f1b=betareg(y ~1,link="probit")
f2b=betareg(y ~x3+x4,link="probit")
f3b=betareg(y ~x3+x4,link="probit")
f4b=betareg(y ~x3+x4+x2,link="probit")
f5b=betareg(y ~x3+x4+x1,link="probit")
f6b=betareg(y ~x3+x4+x5,link="probit")
f7b=betareg(y ~x3+x4+x6,link="probit")
f8b=betareg(y ~x3+x4+x2+x7,link="probit")
anova.betareg(f1a,f2a,f3a,f4a,f5a,f6a,f7a,f8a)
##### modelo3(reajustando o modelol sem as covariável x1,x2,x5, x6 e
x7 - usando a ligação logito #####
xdata2=cbind(x2,x3,x4,x6)
xdata2=xdata2[-4,]
fit3=betareg(y1a~xdata2,link="logit")
summary.betareg(fit3)
logLik.betareg(fit3)
estfun.betareg(fit3)
aov(fit3)
deviance(fit3)
#Intervalos de Confiança para os betas
#obtendo os dp's
dpb0<-fit3$stder[1]
dpb2<-fit3$stder[2]
dpb3<-fit3$stder[3]
dpb4<-fit3$stder[4]

```

```
dpb6<-fit3$stder[5]
dpphi<-fit3$stder[6]
#IC para beta0
icb0<-c(coef(fit3)[1]-qt(0.975,45)*dpb0,coef(fit3)[1]+qt(0.975,45)*dpb0)
#IC para beta2
icb2<-c(coef(fit3)[2]-qt(0.975,45)*dpb2,coef(fit3)[2]+qt(0.975,45)*dpb2)
#IC para beta3
icb3<-c(coef(fit3)[3]-qt(0.975,45)*dpb3,coef(fit3)[3]+qt(0.975,45)*dpb3)
#IC para beta4
icb4<-c(coef(fit3)[4]-qt(0.975,45)*dpb4,coef(fit3)[4]+qt(0.975,45)*dpb4)
#IC para beta6
icb6<-c(coef(fit3)[5]-qt(0.975,45)*dpb6,coef(fit3)[5]+qt(0.975,45)*dpb6)
#IC para phi
icphi<-c(coef(fit3)[6]-qt(0.975,45)*dpphi,coef(fit3)[6]+qt(0.975,45)*dpphi)
icb0
icb2
icb3
icb4
icb6
icphi
####Diagnósticos #####
cooks.distance.betareg(fit3)
hatvalues.betareg(fit3)
vcov.betareg(fit3)
plot.betareg(fit3) ##gráficos p/ diagnósticos
par(mfrow=c(1,1))
envelope.beta(fit3)
preditos3=predict.betareg(fit3)
res_dev3=fit3$resd
res_pad3=fit3$resstd
res_ord3=fit3$residuals
```



```

preditos_residuos3=cbind(preditos3,res_dev3,res_pad3,res_ord3)
round(preditos_residuos3,4)
plot(preditos3,res_ord3,main="resíduos vs preditos")
identify(preditos3,res_ord3);
H3=fit3$h; ##### leverage
eta_chapeu3=fit3$etahat #####preditor linear ajustado
plot(eta_chapeu3,res_ord3,main="Verificando a função de ligação")
identify(eta_chapeu3,res_ord3);
gen.lev.betareg(fit3) #####levegare genralizado
leverage_gene3=gen.lev.betareg(fit3) #####levegare genralizado
diag(leverage_gene3)>=(10/50)
10/50; ###2k/n ###
leverage_gene3[6]
leverage_gene3[35]
leverage_gene3[44]
leverage_gene3[47]
##### modelo4(reajustando o modelo2 sem a covariáveis x1, x2, x5, x6 e
x7 - usando a ligação probito #####
fit4=betareg(y1a ~xdata2,link="probit")
summary.betareg(fit4)
logLik.betareg(fit4)
estfun.betareg(fit4)
aov(fit4)
deviance(fit4)
#Intervalos de Confiança para os betas
#obtendo os dp's
dpb0<-fit4$stder[1]
dpb2<-fit4$stder[2]
dpb3<-fit4$stder[3]
dpb4<-fit4$stder[4]
dpb6<-fit4$stder[5]

```

```

dpphi<-fit4$stdr[6]
#IC para beta0
icb0<-c(coef(fit4)[1]-qt(0.975,45)*dpb0,coef(fit4)[1]+qt(0.975,45)*dpb0)
#IC para beta2
icb2<-c(coef(fit4)[2]-qt(0.975,45)*dpb2,coef(fit4)[2]+qt(0.975,45)*dpb2)
#IC para beta3
icb3<-c(coef(fit4)[3]-qt(0.975,45)*dpb3,coef(fit4)[3]+qt(0.975,45)*dpb3)
#IC para beta4
icb4<-c(coef(fit4)[4]-qt(0.975,45)*dpb4,coef(fit4)[4]+qt(0.975,45)*dpb4)
#IC para beta6
icb6<-c(coef(fit4)[5]-qt(0.975,45)*dpb6,coef(fit4)[5]+qt(0.975,45)*dpb6)
#IC para phi
icphi<-c(coef(fit4)[6]-qt(0.975,45)*dpphi,coef(fit4)[6]+qt(0.975,45)*dpphi)
icb0
icb2
icb3
icb4
icb6
icphi
####Diagnósticos #####
cooks.distance.betareg(fit4)
hatvalues.betareg(fit4)
vcov.betareg(fit1)
plot.betareg(fit4) ##gráficos p/ diagnósticos
par(mfrow=c(1,1))
envelope.beta(fit4)
preditos4=predict.betareg(fit4)
res_dev4=fit4$resd
res_pad4=fit4$resstd
res_ord4=fit4$residuals
preditos_residuos4=cbind(preditos4,res_dev4,res_pad4,res_ord4)

```

```

round(preditos _residuos4,4)
plot(preditos4,res _ord4,main="resíduos vs preditos")
identify(preditos4,res _ord4);
H4=fit4$h; ##### leverage
eta _chapeu4=fit4$etahat #####preditor linear ajustado
plot(eta _chapeu4,res _ord4,main="Verificando a função de ligação")
identify(eta _chapeu4,res _ord4);
gen.lev.betareg(fit4) #####levegare genralizado
leverage _gene4=gen.lev.betareg(fit4) #####levegare genralizado
diag(leverage _gene4)>=(10/50)
10/50; ###2k/n ###
leverage _gene4[6]
leverage _gene4[35]
leverage _gene4[44]
leverage _gene4[47]
#####Comparação entre regressão beta via logito e regressão beta via probito#####
comparar=cbind(yt,preditos3,preditos4,res _ord3,res _ord4)
comparar
sqres _mod3=sum((res _ord3)^2)
sqres _mod3
sqres _mod4=sum((res _ord4)^2)
sqres _mod4
sqres _mod3<sqres _mod4
##### modelo _ com regressão linear( sem as covariáveis x1,x5 e x7) #####
linear=lm(y1a ~xdata2)
summary(linear)
aov(linear)
pred _lin=predict(linear)
res _ord _lin=linear$res
plot(linear)
#####Comparação entre regressão beta e reg.linear#####

```

```

comparar=cbind(yt,preditos3,pred_lin,res_ord3,res_ord_lin)
comparar
sqres_mod3=sum((res_ord3)^2)
sqres_mod3
sqres_lin=sum((res_ord_lin)^2)
sqres_lin
sqres_mod3<sqres_lin

```

Otimização

```

##### Programa 4 #####
library(MASS)
## Dados ##

## Os dados foram omitidos #####

ydata=cbind(yt)
const=rep(1,length(ydata))
start=c(log(mean(ydata)/(1-mean(ydata))),-log((mean(ydata)-
(mean(ydata))^2-var(ydata))/var(ydata)))
start
##A função betareg calcula a verossimilhança ##
betareg=function(h,y,x,z)
{
  hx=x%*%h[1:length(x[1,])]
  mu=exp(hx)/(1+exp(hx))
  gz=z%*%h[length(x[1,])+1:length(z[1,])]
  phi=exp(-gz)
  loglik=lgamma(phi)-lgamma(mu*phi)-lgamma(phi-mu*phi)+
mu*phi*log(y)+(phi-mu*phi)*log(1-y)-log(y) -log(1-y)
  -sum(loglik,na.rm=TRUE)
}

```

```

## A função grad para calcular o gradiente #####
grad=function(h,y,x,z)
{
  hx=x%%h[1:length(x[1,])]
  gz=z%%h[length(x[1,])+1:length(z[1,])]
  gd=cbind(x*rep(exp(-gz+hx)*(log(y/(1-y))+digamma(exp(
-gz)/(1+exp(hx)))-digamma(exp(-gz+hx)/(1+exp(hx))))/(1+exp(hx))^2,
length(x[1,]),-z*rep(exp(-gz)*(log(1-y)+exp(hx)*log(y)+(1+
exp(hx))*digamma(exp(-gz))-digamma(exp(-gz)/(1+exp(hx)))-
exp(hx)*digamma(exp(-gz+hx)/(1+exp(hx))))/(1+exp(hx)),length(z[1,])))
  colSums(gd,na.rm=TRUE)
}
xdata=cbind(const,x1,x2,x3,x4,x5,x6,x7)
wdata=cbind(const,x6,x7)
betafit=nlminb(start,betareg,x=xdata,y=ydata,z=wdata)
## Esta é a verossimilhança e mensagem de convergência, seguido pelo gradiente da
solução
c(betafit$objective,betafit$message)
## estimativas dos Parâmetros, erro padrão assintotico, estatística z e níveis de sig-
nificancia ###
estimativas=cbind(estim=betafit$parameters,serr=sqrt(diag(vcov.nlminb(betafit))),
zstat=estim/serr,prob=1-pnorm(abs(zstat)))
col.names(estimativas)=c("parametros","erropad","zstat","prob");
estimativas
## Checando se o gradiente é zero ##
grad(betafit$parameters,ydata,xdata,wdata)
### Otimização #####
betaopt=optim(start,betareg,hessian=T,x=xdata,y=ydata,
z=wdata,method="BFGS")
estimativas_otimas=cbind(estim=betaopt$par,serr=sqrt(diag(solve(
betaopt$hessian))),zstat=estim/serr,prob=1-pnorm(abs(zstat)))

```

```

col.names(estimativas_otimas)=c("parametros_otimos","erropad","zstat","prob")
estimativas_otimas
Verossimilhanca=betaopt$value
verossimilhanca
grad(betaopt$par,ydata,xdata,wdata)
betaopt$convergence

```

.2.3 Apêndice C: Análise Clássica em SAS

** Programa1 em SAS**

```

title 'Regressão Beta no Caso Charter Scholls';
data charter_school;
input y x1 x2 x3 x4 x5 x6 x7;
cards;
;
run;

proc nlmixed data=charter_school ALPHA=0.05 tech=trureg hess cov CORR itde-
tails;
parms b0=3 b1=0.0006 b2=0.0003 b3=-0.0005 b4=-0.002 b5=0.002 b6=0.00001 b7=-
0.000001 c0=1 c1=1 c2=1;
xb = b0 + b1*x1 + b2*x2 + b3*x3 + b4*x4 + b5*x5 + b6*x6 + b7*x7;
mu = exp(xb)/(1 + exp(xb));
zc = c0 + c1*x6 + c2*x7;
phi = exp(-1*zc);
p = mu*phi;
q = phi- mu*phi;
ll = lgamma(p + q) - lgamma(p) - lgamma(q) + (p-1)*log(y) + (q-1)*log(1-y);
model y ~ general(ll);
predict mu out=y_pred_loc;
run;

```

.2.4 Apêndice D: Programas em WinBugs para a Análise Bayesiana

Modelo com phi conhecido e priori vaga normal

```

model {
  for (i in 1:n){
    a[i]<-b0+b1*x1[i]+b2*x2[i]+b3*x3[i]+b4*x4[i]+b5*x5[i]+b6*x6[i]+b7*x7[i];
    mu[i]<-exp(a[i])/(1+exp(a[i]));
    phi[i]<-17.29;
    V[i]<-mu[i]*(1-mu[i])/(phi[i]+1);
    p[i]<-(pow(mu[i],2)*(1-mu[i]))/(V[i]-mu[i]);
    q[i]<-(mu[i]*pow(1-mu[i],2)/V[i])-(1-mu[i]);
    y[i] ~ dbeta(p[i],q[i]);
    e[i]<-pow(mu[i]-y[i],2) #para calcular MSE
  }

  ## Prioris
  b0 ~ dnorm(0,0.01)
  b1 ~ dnorm(0,0.01)
  b2 ~ dnorm(0,0.01)
  b3 ~ dnorm(0,0.01)
  b4 ~ dnorm(0,0.01)
  b5 ~ dnorm(0,0.01)
  b6 ~ dnorm(0,0.01)
  b7 ~ dnorm(0,0.01)

  ## MSE
  mse<-mean(e[ ])
}

### Dados ###
list( Os dados foram omitidos)

### Valores iniciais ###
list(b0=0,b1=0,b2=0,b3=0,b4=0,b5=0,b6=0,b7=0)
list(b0=0.5,b1=0.5,b2=0.5,b3=0.5,b4=0.5,b5=0.5,b6=0.5,b7=0.5)

```

Modelo com phi desconhecido e priori vaga normal

```

model {
  for (i in 1:n){
    a[i]<-b0+b1*x1[i]+b2*x2[i]+b3*x3[i]+b4*x4[i]+b5*x5[i]+b6*x6[i]+b7*x7[i];
    mu[i]<-exp(a[i])/(1+exp(a[i]));
    d[i]<-g0+g1*x6[i]+g2*x7[i];
    phi[i]<-exp(d[i]);
    V[i]<-mu[i]*(1-mu[i])/(phi[i]+1);
    p[i]<-(pow(mu[i],2)*(1-mu[i]))/(V[i]-mu[i]);
    q[i]<-((mu[i]*pow(1-mu[i],2)/V[i])-(1-mu[i]));
    y[i] ~ dbeta(p[i],q[i]);
    e[i]<-pow(mu[i]-y[i],2) #para calcular MSE
  }
  ## Prioris
  b0 ~ dnorm(0,0.01)
  b1 ~ dnorm(0,0.01)
  b2 ~ dnorm(0,0.01)
  b3 ~ dnorm(0,0.01)
  b4 ~ dnorm(0,0.01)
  b5 ~ dnorm(0,0.01)
  b6 ~ dnorm(0,0.01)
  b7 ~ dnorm(0,0.01)
  g0 ~ dnorm(0,0.01)
  g1 ~ dnorm(0,0.01)
  g2 ~ dnorm(0,0.01)
  ## MSE
  mse<-mean(e[ ])
}
### Dados #####
list(Os dados foram omitidos)
### Valores iniciais ###

```



```
list(b0=0,b1=0,b2=0,b3=0,b4=0,b5=0,b6=0,b7=0,g0=0,g1=0,g2=0)
```

```
list(b0=0.5,b1=0.5,b2=0.5,b3=0.5,b4=0.5,b5=0.5,b6=0.5,b7=0.5,g0=0.5,g1=0.5,g2=0.5)
```

Modelo com phi desconhecido e priori informativa normal

```
model {
  for (i in 1:n){
    a[i]<-b0+b1*x1[i]+b2*x2[i]+b3*x3[i]+b4*x4[i]+b5*x5[i]+b6*x6[i]+b7*x7[i];
    mu[i]<-exp(a[i])/(1+exp(a[i]));
    d[i]<-g0+g1*x6[i]+g2*x7[i];
    phi[i]<-exp(d[i]);
    V[i]<-mu[i]*(1-mu[i])/(phi[i]+1);
    p[i]<-(pow(mu[i],2)*(1-mu[i]))/(V[i]-mu[i]);
    q[i]<-((mu[i]*pow(1-mu[i],2)/V[i])-(1-mu[i]));
    y[i] ~ dbeta(p[i],q[i]);
    e[i]<-pow(mu[i]-y[i],2) #para calcular MSE
  }
  ## Prioris
  b0 ~ dnorm(2.77,0.21)
  b1 ~ dnorm(0.107,04.0E2)
  b2 ~ dnorm(-0.17,73)
  b3 ~ dnorm(-0.08,2.0E3)
  b4 ~ dnorm(-0.20,7)
  b5 ~ dnorm(0.03,2)
  b6 ~ dnorm(-0.14,15)
  b7 ~ dnorm(-0.12,8)
  g0 ~ dnorm(-2,1)
  g1 ~ dnorm(0.54,29)
  g2 ~ dnorm(0.18,1)
  ## MSE
  mse<-mean(e[ ])
```

```
}  
### Dados ####  
list(Os dados foram omitidos)  
### Valores iniciais ###  
list(b0=0,b1=0,b2=0,b3=0,b4=0,b5=0,b6=0,b7=0,g0=0,g1=0,g2=0)  
list(b0=0.5,b1=0.5,b2=0.5,b3=0.5,b4=0.5,b5=0.5,b6=0.5,b7=0.5,g0=0.5,g1=0.5,g2=0.5)
```

.2.5 Apêndice E: Programas em R para calcular o BIC

```
### Programa para calcular o BIC #####  
## Modelo 5 ###  
Lmodelo5<-read.table("C:\\\\coda_model5.txt", header=F);  
L5<-Lmodelo5[,2];  
BIC5=-2*log(max(L5))+(8*log(51));  
BIC5;  
## Modelo6 ###  
Lmodelo6<-read.table("C:\\\\coda_model6.txt", header=F);  
L6<-Lmodelo6[,2];  
BIC6=-2*log(max(L6))+(11*log(51));  
BIC6;  
## Modelo 7 ###  
Lmodelo7<-read.table("C:\\\\coda_model7.txt", header=F);  
L7<-Lmodelo7[,2];  
BIC7=-2*log(max(L7))+(11*log(51));  
BIC7;
```